

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА МОИ УКРАЇНИ

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА МОИ УКРАЇНИ

Кваліфікаційна наукова праця
на правах рукопису

КОРЧИНСЬКИЙ ВЛАДИСЛАВ ВІКТОРОВИЧ

УДК 519.866:336.713]:303.722.4(477)(043.5)

ДИСЕРТАЦІЯ

**КЛАСТЕРНЕ МОДЕЛЮВАННЯ ФІНАНСОВОГО СТАНУ КОМЕРЦІЙНИХ
БАНКІВ УКРАЇНИ**

051 «Економіка»

Подається на здобуття наукового ступеня доктора філософії

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

В. В. Корчинський В. В. Корчинський

Науковий керівник: Великоіваненко Галина Іванівна, кандидат фізико-математичних наук, професор

Київ – 2023

АНОТАЦІЯ

Корчинський В. В. Кластерне моделювання фінансового стану комерційних банків України. - Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата економічних наук (доктора філософії) за спеціальністю 051 «Економіка» – Київський національний економічний університет ім. Вадима Гетьмана, Київ, 2023.

Дисертаційна робота присвячена питанням розробки методологічних засад аналізу вітчизняних комерційних банків та реалізації системи моделей та алгоритмів для підтримки проведення такого аналізу з метою аналізу фінансового стану комерційних банків, їх надійності та ефективності. Об'єктом дослідження є фінансовий стан комерційних банків. Предметом дослідження виступає інструментарій економіко-математичного моделювання та машинного навчання, що застосовуються для аналізу фінансового стану комерційних банків.

Незважаючи на відносне поліпшення загального стану банківського сектору у 2019-2021 роки, все ще можна було спостерігати певні негативні тенденції у вітчизняній банківській сфері та у фінансовому секторі в цілому, що може бути наслідком фінансових криз попередніх років, або свідчити про ще недостатню зрілість вітчизняних фінансової та банківської систем. Так, з 2013 року з ринку вибуло 108 банківських установ, частка проблемних активів в портфелі банків хоча і має позитивну динаміку до зменшення, однак залишається на високому рівні, показники фінансової ефективності ROA та ROI хоча і вийшли на додатний рівень за останні роки, однак демонструють суттєві флуктуації. Крім іншого, спостерігається скорочення банками своїх фізичних відділень, що може свідчити про необхідність скорочення операційної діяльності та/або адміністративних витрат. Капітало-стійкість банків, відповідно до проведених НБУ стрес-тестувань, не є оптимальною. В цілому, у досліджуваному періоді, незважаючи на деякі позитивні тенденції, в банківському секторі України все ще є проблемні аспекти, які не дають можливості стверджувати про однозначно високий рівень фінансового стану, надійності та ефективності комерційних банків в системі. Саме тому питання

розробки системи моделей та алгоритмів для аналізу банківських установ, що ґрунтується на відповідній методології, є актуальним. Повномасштабне вторгнення, що розпочалося на території України 24 лютого 2022 року також справило суттєвий вплив на економіку держави, та на функціонування банківської системи зокрема, розпочався новий період нестабільності та розбалансування. Дослідження, присвячене періоду після-кризового відновлення 2015-2021 років, так само буде актуальне після закінчення війни з точки зору ретроспективного аналізу післявоєнного періоду відновлення, та буде перспективним з точки зору його адаптації до аналізу кризового воєнного періоду.

У результаті проведеного дослідження розглянуто предметну область та понятійний апарат, надано авторські визначення ключових компонентів понятійного апарату, висвітлено їх взаємозв'язки між собою. Так, розглянуте ключове поняття фінансового стану комерційного банку в контексті досліджень інших вчених, запропоновано під фінансовим станом банку розглядати його здатність забезпечувати надійність та ефективність своєї роботи: забезпечувати ефективний ризик-менеджмент, підтримувати високу якість активів, здатність залучати ресурси, забезпечувати адекватний рівень капіталу та ліквідності для захисту депозитів на випадок паніки вкладників та позик на випадок падіння їх якості, і в результаті генерувати прибуток та забезпечувати достатній рівень маржі.

Визначено ефективність банківської діяльності як багатогранне поняття, що не тільки характеризує здатність ефективно використовувати ресурси (власні чи залучені) з метою максимізації прибутку, а й здатність забезпечувати високу надійність та адекватний механізм управління ризиками з метою підвищення надійності та конкурентоспроможності на ринку. При цьому ефективність діяльності визначено як відповідність фінансового стану банківської установи такому рівню, що надавав би їй можливість вести безперебійну діяльність направлену на отримання прибутку через підтримання основних бізнес-ліній з проведення кредитно-інвестиційних, комісійно-розрахункових та торгових операцій, та збереження власної надійності та платоспроможності, серед іншого через підтримання рівня капіталу та ліквідності.

Такі категорії як прибутковість, конкурентоспроможність, адекватність капіталу та надійність, ліквідність та платоспроможність визначено як складові компоненти фінансової стійкості, які характеризуються набором показників, що описують фінансовий стан комерційного банку.

Ризик визначено як один з ключових елементів, що має безпосередній вплив на фінансову стійкість банку через її компоненти. Підкреслюється те, що в категоріальній системі предметної області дослідження категорія ризику є однією з найбільш значущих, так як реалізація ризику може суттєво вплинути на будь-який компонент фінансової стійкості банку. В банківському секторі різного роду ризики є тісно пов'язаними з результатами діяльності банківських установ, а їх реалізація може спричинити суттєві зміни фінансового стану. В деяких випадках реалізація ризику може навіть призвести до дефолту банківської установи.

Головний фокус дослідження - це проведення аналізу саме фінансового стану, що означає дослідження фінансової стійкості комерційних банків через моделювання набору показників, що її характеризують.

Проведено якісний аналіз банківських ризиків, розглянуто декілька їх класифікацій серед яких класифікації Базель II, Базель III та класифікація НБУ, запропоновано авторську класифікацію ризиків комерційних банків, в якій ключовими видами ризику виділено кредитний, ринковий та операційний ризики, як такі, що мають найвищі частоту реалізації та можливі збитки для комерційних банків.

Розглянуто класичні моделі та методи управління ефективністю банківської установи, оцінювання банківських ризиків, серед яких прості методи засновані на системі показників рентабельності, вартості, капіталу; методи оцінювання валютних, процентних ризиків, ризиків ліквідності, операційних, кредитних ризиків, засновані на розрахункових підходах Базелю, геп-аналізі, регресійних та авто-регресійних моделях, VAR-аналізі, класифікаційних рейтингових та прогнозних моделях, скорингових моделях, дискримінаційних моделях та моделях комплексного аналізу. Розглянуто також більш сучасні та інноваційні моделі, засновані на нечіткій логіці, аналізі виживання, штучних нейронних мережах.

Для досягнення цілей дисертаційної роботи, було вирішено застосувати підходи картографування через отримання двовимірних проекцій за допомогою алгоритмів зниження розмірності з подальшим аналізом, та кластеризації при алгоритмах зниження розмірності. Зважаючи на поставлені цілі дослідження і специфіку завдань класифікації та кластеризації, та розглянуті в роботі обмеження інших сучасних підходів, зроблено висновок що вказані вище підходи будуть найбільш ефективними для побудови моделі аналізу фінансового стану банків.

Вперше сформовано та представлено послідовний багатокроковий підхід до проведення аналізу фінансового стану вітчизняних комерційних банків, на основі якого спроектовано та реалізовано систему моделей та алгоритмів машинного навчання, а саме алгоритмів зниження розмірності t-SNE та UMAP та алгоритму ієрархічної агломеративної кластеризації методом Уорда, що забезпечили відносну простоту реалізації, високу якість візуалізації отримуваних результатів та роздільної здатності двовимірного проектування багатовимірних даних, а отже і адекватність подальшої економічної інтерпретації отриманих кінцевих результатів дослідження.

Запропоновано авторське бачення щодо формування набору ознак в модель, на основі аналітичного та експертного підходів. Такий спосіб забезпечує всебічний та більш повний аналіз функціонування комерційних банків і відповідно більш надійний аналіз їх фінансового стану. Відібраний набір показників характеризує активи, пасиви, капітал, ліквідність, фінансову ефективність, податкове навантаження комерційних банків та враховує системність (вагу активів в системі).

Розглянуто різні методи попередньої обробки даних, на основі проведених тестів обрано методи вінзорування, класичної та робустної стандартизації. Підтверджено придатність даних до кластеризації за допомогою тесту Хопкінса та показника тенденції.

Більш детально розглянуто обрані для аналітичної системи алгоритми зниження розмірності та кластеризації, визначено оптимальні гіпер-параметри обґрунтовано їх перспективність для досягнення мети дослідження.

Успішно відділено інвестиційно привабливі установи від потенційно проблемних. Виділено 15 кластерів, що умовно можна розділити на 4 класи: клас А (першокласні банки, висока фінансова стійкість, надійність, фінансова ефективність), складається з двох кластерів та містить 14 банків; клас Б (потенційно інвестиційно привабливі установи, фінансовий стан, надійність та стійкість на достатньо високому рівні, проте мають ознаки проблем в своїй діяльності, що заважають їм перейти до кластерів класу А) складається з 2 кластерів та містить 16 банків; клас «ризикований» (установи, що характеризуються суттєвими відхиленнями по ряду показників, що однак не дає однозначно охарактеризувати ці банки як проблемні, однак підвищує їх до різних видів ризиків, фінансова стійкість, ефективність та надійність є низькою), що містить 7 кластерів з 25 банків; та клас «проблемний» (установи, що характеризуються суттєвими відхиленнями по більшості своїх показників), що складається з 4 кластерів та 5 банків.

Підтверджено, що банки які знаходяться у багатовимірному просторі різнопланових ознак, що описують їх роботу, знаходяться поруч з подібними собі. Саме з цієї причини вдалося отримати кластери з адекватним розподілом банків між ними на основі спроектованої двовимірної карти.

Досліджено поведінку комерційних банків в динаміці. Доведено, що якщо за останні декілька років банк демонструє активні переміщення на двовимірній проекції – це можна розглядати як тривожну ознаку і негативну характеристику фінансового стану та надійності банку.

Візуально продемонстровано, що проблемні банки формують цілу окрему область на двовимірній карті і добре дискримінуються від фінансово ефективних, надійних банків, так як розміщуються на двовимірній проекції у діаметрально протилежній частині проекції.

Експериментально доведено, на основі порівняльного аналізу, що якість моделювання за використання алгоритму ієрархічної агломеративної кластеризації методом Уорда є високою а побудована система моделей та алгоритмів зниження розмірності та кластеризації є найбільш ефективним рішенням для досягнення мети дослідження: проведення аналізу фінансового стану вітчизняних комерційних

банків. Порівняльний аналіз проводився на базі метрик силуету та оцінок Калінські-Харабаза та Девіса-Болдуїна. Алгоритми зниження розмірності Kernel PCA, Spectral Embedding, MDS та ISOMAP сформували двовимірні проєкції непридатні до подальшої кластеризації, алгоритм t-SNE видає якісні результати для візуального аналізу двовимірної проєкції, а за результатами зваженого ранжування на основі зазначених метрик, найкращий результат кластеризації досягається за використання алгоритмів UMAP та ієрархічної агломеративної кластеризації методом Уорда.

Практичне значення отриманих результатів полягає у обґрунтуванні та демонстрації ефективності застосування методологічного підходу до аналізу фінансового стану комерційних банків та системи моделей та алгоритмів, що застосовуються для реалізації даного підходу.

Наукові розробки, теоретичні концепції та практичні рекомендації дослідження були використані в практичній діяльності, що підтверджено відповідними довідками про впровадження результатів дисертаційної роботи:

1. ТОВ «Інтеркабель Київ» (довідка № 161/1.1-07 від 10.03.2023)
2. ОУ «Wild Web Art» (довідка №11/08-23 від 08.03.2023).

Ключові слова: фінансовий стан, ризик, банківський ризик, ризик-менеджмент, комерційні банки, банківська система, машинне навчання, зниження розмірності, кластерний аналіз.

ABSTRACT

Korchynskiy V. V. Cluster modeling of the financial condition of commercial banks of Ukraine. – Qualification research work as a manuscript.

Thesis for the degree of candidate of economic sciences in the specialty 051 Economy. – Kyiv National Economic University named after Vadym Hetman, Kyiv, 2023.

The dissertation is devoted to the development of methodological principles for the analysis of domestic commercial banks and the implementation of a set of models and algorithms to support such analysis in order to assess the financial condition of commercial banks, their efficiency and reliability. The object of the study is the financial condition of commercial banks. The subject of the research is the toolkit of economic and mathematical modeling and machine learning, which are used to diagnose the financial condition of commercial banks.

Despite the relative improvement of the general condition of the banking sector in 2019-2021, it was still possible to observe certain negative trends in the domestic banking sphere and in the financial sector as a whole, which may be a consequence of the financial crises of previous years, or indicate the still insufficient maturity of the domestic financial and banking system. Thus, since 2013, 108 banking institutions have left the market, the share of problem assets in the portfolio of banks, although it has a positive trend to decrease, but remains at a high level, indicators of financial efficiency ROA and ROI, although they have reached a positive level in recent years, but show significant fluctuations. Among other things, banks are reducing their physical branches, which may indicate the need to reduce operational activity and/or administrative costs. According to the NBU stress tests, the capital stability of banks is not optimal. In general, in the studied period, there are still problematic aspects in the banking sector of Ukraine despite some positive trends, which do not give an opportunity to assert an unambiguously high level of financial condition, reliability and efficiency of commercial banks in the system. That is why the issue of developing a system of models and algorithms for diagnosing banking institutions, based on the appropriate methodology, is urgent.

The full-scale invasion that began on the territory of Ukraine on February 24, 2022 also had a significant impact on the state's economy, and on the functioning of the banking system in particular, a new period of instability and imbalance began. Research on the post-crisis recovery period 2015-2021 will also be relevant after the end of the war from the post-war recovery period retrospective analysis point of view, and it will be promising from the point of view of its adaptation to the analysis of the crisis war period.

As a result of the research, the subject area of research and the conceptual apparatus were analyzed, the author's definitions of the key components of the conceptual apparatus were provided, and their interrelationships were highlighted. Thus, the key concept of the financial condition of a commercial bank was considered in the context of the research of other scientists, it was proposed to consider the financial condition of the bank as its ability to ensure the reliability and efficiency of its work: to ensure effective risk management, maintain high quality of assets, the ability to attract resources, ensure an adequate level of capital and liquidity to protect deposits in case of panic of depositors and loans in case of a drop in their quality, and as a result generate income and ensure a sufficient level of margin.

The effectiveness of banking activity is defined as a multifaceted concept that not only characterizes the ability to effectively use resources (own or borrowed) in order to maximize profit, but also the ability to ensure high reliability and an adequate risk management mechanism in order to increase reliability and competitiveness in the market. At the same time, the effectiveness of the activity is defined as the compliance of the financial condition of the banking institution to such a level that would enable it to conduct uninterrupted activities aimed at obtaining profit through the maintenance of the main business lines of credit and investment, commission settlement and trade operations, and the preservation of its own reliability and solvency, among other things by maintaining the level of capital and liquidity.

Such categories as profitability, competitiveness, capital adequacy and reliability, liquidity and solvency are defined as components of financial stability, which are characterized by a set of indicators describing the financial condition of a commercial bank.

Risk is defined as one of the key elements that has a direct impact on the financial stability of the bank through its components. It is emphasized that in the categorical system of the subject area of the study, the risk category is one of the most significant, since the realization of the risk can significantly affect any component of the bank's financial stability. In the banking sector, various types of risks are closely related to the results of banking institutions' activities, and their implementation can cause significant changes in the financial state. In some cases, the realization of the risk may even lead to the default of the banking institution

The main focus of the research is the analysis of the financial state itself, which means the study of the financial stability of commercial banks through the modeling of a set of indicators that characterize it.

A qualitative analysis of banking risks was carried out, several of their classifications were considered, including Basel II, Basel III and the NBU classification, an author's classification of commercial bank risks was proposed, in which credit, market and operational risks were identified as key types of risk, as those with the highest frequency of realization and possible losses for commercial banks.

Classical models and methods of managing the efficiency of a banking institution, assessing banking risks, including simple methods based on a system of indicators of profitability, cost, capital, are considered; methods of assessing currency, interest rate risks, liquidity risks, operational, credit risks, based on Basel calculation approaches, gap analysis, regression and auto-regression models, VAR analysis, classification rating and forecasting models, scoring models, discrimination models and models of complex analysis. More modern and innovative models based on fuzzy logic, survival analysis, and artificial neural networks are also considered.

To achieve the goals of the dissertation work, it was decided to apply mapping approaches through obtaining two-dimensional projections using dimensionality reduction algorithms with further analysis, and clustering supported by dimensionality reduction algorithms. Taking into account the set goals of the research and the specifics of classification and clustering tasks, and the limitations of other modern approaches

considered in the work, it was concluded that the above approaches will be the most effective for building a model for diagnosing the financial condition of banks.

For the first time, a consistent three-step approach to the diagnosis of the financial condition of domestic commercial banks is presented, on the basis of which a system of machine learning models and algorithms was designed and implemented, namely the t-SNE and UMAP dimensionality reduction algorithms and the hierarchical agglomerative clustering algorithm using the Ward method, which provided a relative ease of implementation, high quality of visualization of the obtained results and resolution of two-dimensional design of multidimensional data, and therefore the adequacy of further economic interpretation of the obtained final research results.

An author's approach to forming a set of features into a model, based on analytical and expert approaches, is proposed. This approach provides a comprehensive and more complete analysis of the functioning of commercial banks and, accordingly, a more reliable diagnosis of their financial condition. The selected set of indicators characterizes the assets, liabilities, capital, liquidity, financial efficiency, tax burden of commercial banks and takes into account the systemic nature (weight of assets in the system).

Different approaches to data pre-processing were considered, and based on the conducted tests, winsoring, classical and robust standardization methods were chosen. The suitability of the data for clustering was confirmed using the Hopkins test and the tendency indicator.

The dimensionality reduction and clustering algorithms selected for the diagnostic system are considered in more detail, the optimal hyper-parameters are determined and their prospects for achieving the research goal are substantiated.

Successfully separated investment-attractive institutions from potentially problematic ones. 15 clusters have been identified, which can be conditionally divided into 4 classes: class A (first-class banks, high financial stability, reliability, financial efficiency), consists of two clusters which contains 14 banks; class B (potentially investment-attractive institutions, financial condition, reliability and stability at a sufficiently high level, but with signs of problems in their activities that prevent them from moving to class A clusters) consists of 2 clusters which contains 16 banks; the

"risky" class (institutions characterized by significant deviations in a number of indicators, which, however, does not make it possible to unequivocally characterize these banks as problematic, but makes them vulnerable to various types of risks, financial stability, efficiency and reliability are low), containing 7 clusters with 25 banks; and the "problematic" class (institutions characterized by significant deviations in most of their indicators), consisting of 4 clusters with 5 banks.

It has been confirmed that points-banks located in the multidimensional space of various features describing their work are close to their peers. That is what made it possible to obtain clusters with an adequate distribution of banks between them on the basis of the projected two-dimensional map.

The behavior of commercial banks in dynamics is studied. It has been proven that if the bank shows active movements on the two-dimensional projection over the last few years, this can be considered as an alarming sign and a negative characteristic of the bank's financial condition and reliability.

It is visually demonstrated that problematic banks form a whole separate area on the two-dimensional map and are well discriminated from financially efficient, reliable banks, as they are located in the diametrically opposite part of the two-dimensional projection.

It has been experimentally proven, on the basis of a comparative analysis, that the quality of modeling using the algorithm of hierarchical agglomerative clustering with the Ward method is high, and the selected set of models and algorithms for dimensionality reduction and clustering is the most effective solution for achieving the goal of the study: conducting diagnostics of the financial condition of domestic commercial banks. Comparative analysis was carried out on the basis of silhouette metrics and Kalinski-Kharabasz and Davies-Bouldien scores. The Kernel PCA, Spectral Embedding, MDS, and ISOMAP dimensionality reduction algorithms generated two-dimensional projections unsuitable for further clustering, the t-SNE algorithm produces qualitative results for visual analysis of the two-dimensional projection, and according to the results of weighted ranking based on the specified metrics, the best clustering result is achieved

using the algorithms system of UMAP and hierarchical agglomerative clustering with the Ward's method.

The practical significance of the obtained results lies in the substantiation and demonstration of the effectiveness of the methodological approach of diagnosing the financial condition of commercial banks and the set of models and algorithms used to implement this approach.

Scientific developments, theoretical concepts and practical recommendations of the study were used in practical activities, which is confirmed by relevant certificates on the implementation of the results of the dissertation work:

1. "Intercable Kyiv" LLC (reference No. 161/1.1-07 dated 03/10/2023)
2. "Wild Web Art" OÜ (reference No. 11/08-23 dated 03/08/2023).

Keywords: financial condition, risk, banking risk, risk management, commercial banks, banking system, machine learning, dimensionality reduction, cluster analysis.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

У наукових фахових виданнях України:

1. Корчинський В. В., Долінський Л. Б. Ідентифікація та кількісне оцінювання кредитного ризику комерційного банку. *Економічний аналіз: зб. наук. праць*. 2016. Т. 25, № 1. С. 180–189. (1,12 друк. арк., з них 1 друк. арк. належить особисто автору: побудова класифікації ризиків, побудова моделі дискримінантного аналізу позичальників для оцінки кредитного ризику банків, побудовано карту кредитного ризику);
2. Великоіваненко Г. І., Корчинський В. В., Чернишова В. В. Дослідження ефекту перенавчання нейронних мереж на прикладі задачі аплікаційного скорингу. *Нейро-нечіткі технології моделювання в економіці*. 2016. № 5. С. 3–23. (1,22 друк. арк., з них 1,15 друк. арк. належить особисто автору: проведене моделювання кредитного ризику позичальників-фізичних осіб та оцінка адекватності моделей дослідження);
3. Корчинський В. В., Чернишова В. В. Побудова скорингової моделі на основі нейронної мережі типу Персептрон. *Український журнал прикладної економіки*. 2016. Т. 1, № 4. С. 46–55. (1,25 друк. арк., з них 1,12 друк. арк. належить особисто автору: проведення аналізу та нормалізації статистичних даних, побудова нейронної мережі та дослідження ефективності та якості її прогнозних властивостей);
4. Корчинський В. В., Долінський Л. Б., Мірошніченко І. В. Кластерний аналіз надійності банківських установ в Україні. *Моделювання та інформаційні системи в економіці: зб. наук. праць*. 2017. № 93. С. 49–73. (1,2 друк. арк., з них 1,08 друк. арк. належить особисто автору: побудовано нейронну мережу типу карт самоорганізації Кохонена для кластерного аналізу надійності комерційних банків);
5. Лаговський В. В., Долінський Л. Б., Корчинський В. В. Особливості кредитно-інвестиційної діяльності на фінансовому ринку України. *Економічний вісник: науковий фаховий журнал*. 2020. Т. 25, № 4. С. 74–88. (0,58 друк. арк., з них 0,52 друк. арк. належить особисто автору: проведено детальний статистичний аналіз ряду показників, що характеризують динаміку економічного розвитку і

розвитку фінансового сектору, проведено кореляційних аналіз цих показників і визначено істотну кореляційну залежність між ними);

У зарубіжному періодичному науковому виданні, що зареєстроване у міжнародній наукометричній базі Scopus:

6. Korchynskyi V., Velykoivanenko H. Application of Clustering in the Dimensionality Reduction Algorithms for Separation of Financial Status of Commercial Banks in Ukraine. *Universal Journal of Accounting and Finance*. 2022. Vol. 10, no. 1. P. 148–60. URL: <https://doi.org/10.13189/ujaf.2022.100116>. (1,18 друк. арк., з них 1,06 друк. арк. належить особисто автору: побудовано систему алгоритмів машинного навчання для проведення кластерного аналізу фінансового стану комерційних банків України, проведено аналіз фінансового стану комерційних банків на основі результатів зниження роозмірності);

В інших виданнях:

7. Корчинський В. В. Побудова системи ризик-менеджменту в комерційному банку. *Інноваційний ресурс соціально-економічного прогресу України*: зб. доп. 83-ї наук. конф. студ. КНЕУ, м. Київ, 25–27 квіт. 2016 р. Київ, 2016. С. 290–293. (0,17 друк. арк., з них 0,15 друк. арк. належить особисто автору: запропоновано дискримінантну модель для оцінки надійності позичальника і кредитного ризику) (форма участі: очна);

8. Корчинський В. В., Чергишова В. В. Скорингова модель оцінки кредитоспроможності клієнтів банків на основі інструментарію нечіткої логіки. *Інноваційний ресурс соціально-економічного прогресу України*: зб. доп. 83-ї наук. конф. студ. КНЕУ, м. Київ, 25–27 квіт. 2016 р. Київ, 2016. С. 602–604. (0,12 друк. арк., з них 0,1 друк. арк. належить особисто автору: запропоновано скорингову модель для оцінки кредитоспроможності клієнтів банків) (форма участі: очна);

9. Долінський Л. Б., Корчинський В. В. Відбір кількісних показників для проведення кластеризації вітчизняних банків в аспекті фінансової стійкості. *Економіко-математичне моделювання*: зб. мат. Першої нац. наук.-матод. конф, м. Київ, 30 верес. – 1 жовт. 2016 р. Київ, 2016. С. 122–123. (0,12 друк.

арк., з них 0,1 друк. арк. належить особисто автору: запропоновано та обґрунтовано перелік показників для подальшого проведення кластеризації вітчизняних банків) (форма участі: очна);

10. Матвійчук А. В., Корчинський В. В., Чернишова В. В. Алгоритм побудови скорингової моделі на основі нейронної мережі персептронного типу. *КМОСС: Матеріали II Всеукр. науково-техн. конф.*, м. Дніпро, 1–3 листоп. 2016 р. Дніпро, 2016. С. 202–203. (0,12 друк. арк., з них 0,1 друк. арк. належить особисто автору: розглянуто алгоритм побудови скорингової моделі для оцінки кредитоспроможності позичальників) (форма участі: заочна);

11. Долінський Л. Б., Корчинський В. В. Кластерний аналіз надійності вітчизняних банківських установ. *Інноваційний ресурс соціально-економічного розвитку України: зб. доп. 84-ї наук. конф. студ. КНЕУ, м. Київ, 24 квіт. – 23 трав. 2017 р. Київ, 2017. С. 389–391.* (0,11 друк. арк., з них 0,1 друк. арк. належить особисто автору: запропоновано та розглянуто концептуальний підхід до проведення кластерного аналізу надійності вітчизняних банків) (форма участі: очна);

12. Долінський Л. Б., Корчинський В. В. Застосування сучасних інформаційних технологій для кластеризації українських комерційних банків в аспекті фінансової стійкості. *Математичні методи, моделі та інформаційні технології в економіці: Матеріали VI Міжнар. науково-метод. конф.*, м. Чернівці, 18–19 квіт. 2019 р. Чернівці, 2019. С. 88–92. (0,2 друк. арк., з них 0,18 друк. арк. належить особисто автору: обґрунтована доцільність застосування сучасних та інноваційних інформаційних технологій для проведення кластеризації комерційних банків) (форма участі: заочна);

13. Долінський Л. Б., Корчинський В. В. Аналіз особливостей кредитно-інвестиційної діяльності банків в Україні. *Економіка підприємства: сучасні проблеми теорії та практики: Матеріали дев'ятої міжнар. наук.-практ. конф.*, м. Одеса, 11–12 верес. 2020 р. Одеса, 2020. С. 105–106. (0,12 друк. арк., з них 0,1 друк. арк. належить особисто автору: проведений аналіз фінансового стану банківського сектору України в динаміці) (форма участі: заочна);

14. Великоіваненко Г. І., Корчинський В. В. Моделювання фінансового стану комерційних банків України на основі інструментів та методів машинного навчання. *Актуальні проблеми економіки, фінансів, обліку та права в умовах сучасних викликів*: зб. тез доп. міжнар. науково-практ. конф., м. Полтава, 25 серп. 2021 р. Полтава, 2021. С. 53–54. (0,2 друк. арк., з них 0,18 друк. арк. належить особисто автору: запропоновано та розглянуто підхід до аналізу фінансового стану комерційних банків за допомогою інструментів машинного навчання) (форма участі: заочна);

15. Великоіваненко Г. І., Корчинський В. В. Аналіз стабільності показників фінансового стану комерційних банків України на основі методів машинного навчання. *КМОСС: Матеріали VII Міжнар. науково-техн. конф.*, м. Дніпро, 3–5 листоп. 2021 р. Дніпро, 2021. С. 93–94. (0,12 друк. арк., з них 0,1 друк. арк. належить особисто автору: проведено аналіз стабільності показників комерційних банків в часі і розглянуто залежність між динамікою показників та фінансовим станом банків) (форма участі: заочна).

ЗМІСТ

ВСТУП.....	19
РОЗДІЛ 1	29
ТЕОРЕТИЧНІ ЗАСАДИ МОДЕЛЮВАННЯ ФІНАНСОВОГО СТАНУ КОМЕРЦІЙНИХ БАНКІВ.....	29
1.1. Аналіз поточного стану банківської системи та тенденцій її розвитку	29
1.2. Сутність, особливості та аналіз фінансового стану комерційних банків	35
1.3. Моделі та методи аналізу фінансового стану комерційних банків та ризиків банківської діяльності.....	64
Висновки до розділу 1	99
РОЗДІЛ 2	103
СИСТЕМА МОДЕЛЕЙ АНАЛІЗУ ФІНАНСОВОГО СТАНУ КОМЕРЦІЙНИХ БАНКІВ УКРАЇНИ	103
2.1. Відбір показників та ознак до моделі аналізу фінансового стану комерційних банків	103
2.2. Модель аналізу фінансового стану комерційних банків на підґрунті алгоритмів зниження розмірності	119
2.3. Модель аналізу фінансового стану комерційних банків на основі ієрархічного кластерного аналізу	134
Висновки до розділу 2	143
РОЗДІЛ 3	146
ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ФІНАНСОВОГО СТАНУ КОМЕРЦІЙНИХ БАНКІВ УКРАЇНИ НА ПІДҐРУНТІ СИСТЕМИ МОДЕЛЕЙ ЗНИЖЕННЯ РОЗМІРНОСТІ ТА КЛАСТЕРИЗАЦІЇ	146
3.1. Інформаційне забезпечення системи моделей аналізу фінансового стану комерційних банків	146
3.2. Експериментальне дослідження фінансового стану комерційних банків України на підґрунті алгоритмів зниження розмірності.....	156
3.3. Експериментальне дослідження фінансового стану комерційних банків України на підґрунті алгоритму кластеризації	168
Висновки до розділу 3	190
ВИСНОВКИ.....	193
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	197
ДОДАТКИ.....	219

ВСТУП

Актуальність теми. Банківську систему можна назвати однією з найважливіших складових національної економіки, оскільки вона пов'язана з усіма її секторами й саме вона забезпечує єдність економічного простору за рахунок підтримки руху фінансових потоків, грошового обігу, обігу капіталу, надання потрібного фінансування промисловим підприємствам, державному бюджету, приватним господарствам. Усе це сприяє здійсненню ефективної економічної діяльності суб'єктами господарювання, забезпечує стабільність економіки країни та сталий розвиток суспільства загалом. Вітчизняний фінансовий сектор загалом, та банківський сектор зокрема за роки незалежності пройшли через низку макроекономічних шоків, таких як кризи 2008 та 2014 років, після кожної з них відбувалися певні відновлювальні та трансформаційні процеси, в рамках яких банківські установи демонстрували високу волатильність своїх показників в динаміці, їх функціонування можна було охарактеризувати як нестабільне. В період з 2015 по 2021 можна говорити про певну поступову стабілізацію банківської системи: було проведено масштабну чистку комерційних банків, спостерігається поступове підвищення загальної дохідності, проводиться послідовна реструктуризація проблемної заборгованості банками, через що спостерігається зменшення частки проблемних активів в банківському кредитному портфелі, все менше банківських установ порушують нормативи регулятора, навіть незважаючи на введення нових нормативів ліквідності згідно європейської практики та їх поступове посилення. Однак, незважаючи на все це, не можна стверджувати про однозначну надійність, ефективність банківської системи, стійкий фінансовий стан окремих її елементів. Так, стрімке зменшення кількості комерційних банків за останні роки може свідчити про низький рівень зрілості банківської системи, і процес вибуття установ з ринку продовжується, але вже не через «чистку банків», а з природних причин: деякі банки не в змозі утриматись на ринку через проблеми з рентабельністю, капіталізацією або ліквідністю. Проблемні кредити хоча і поступово зменшуються, все ще складали суттєву частку в

банківських портфелях станом на 2020-2021 роки, і динаміку на наступні роки передбачити досить складно. Показники дохідності *ROA* та *ROI* хоча і знаходяться на додатному рівні з середини 2018 року, проте демонструють суттєві коливання в часі, «маятникове» зменшення показників, що зважаючи на інфляційні процеси може свідчити про фінансову нестійкість, нестабільність. Про скорочення операційної діяльності банків може свідчити зменшення структурних підрозділів на 32.5% станом на початок 2022 року. Все це може вказувати на певні проблемні процеси, які все ще притаманні вітчизняній банківській системі, і подолання яких може зайняти тривалий час. Курс на європейські стандарти банківської справи, з одного боку, стимулюють еволюційний розвиток української банківської системи, а з іншого, можуть призвести до нових «шоків». Починаючи з 24 лютого 2022 року в Україні відбувається повномасштабне вторгнення та впроваджено режим воєнного стану, держава переживає черговий період часткової дестабілізації та розбалансування економіки в цілому та банківського сектору зокрема. Але після завершення війни невідворотно настане період відновлення, і саме тому тематика дослідження фінансового стану комерційних банків та моделей їх в періоди після-кризового відновлення та після-кризової стабілізації є і буде залишатись актуальною, а розвиток ідей даного дослідження в бік аналізу банківського сектору та його поведінки під час періоду війни може дати перспективні результати для визначення реальної стійкості банківської системи до форс-мажорних, кризових явищ.

Проблеми функціонування та розвитку банківських систем досліджували такі вчені як: Винниченко А. В., Герасимович А. М., Дзюблюк О. В., Довгаль Ю. С., Єріс Л. М., Загородній А. Г., Кочетков В. М., Краснова І. В. Науменкова С. В., Примостка Л. О., Парасій-Варгуненко І. М., Рисін В. В., Чамар Р. О., Щербаков Г. В. та інші.

Значний вклад в дослідження та розвиток моделей та методів економіко-математичного моделювання призначених для управління економічними процесами та ризиками зроблено такими вченими як: Великоіваненко Г. І., Вітлінський В. В., Галіцин В. К., Геєць В. М., Гур'янова Л. С., Заруцька О. П.,

Камінський А. Б., Лавренюк В. В., Кишакевич Б. Ю., Клебанова Т. С., Клименюк М. М., Лук'яненко Г. І., Матвійчук А. В., Максишко Н. К., Меркулова Т. В., Негрей М. В., Скрипниченко М. І., Соловйов В. М., Суслов О. П., Черняк О. І., та інші.

Огляд наявних досліджень в області дослідження комерційних банків з точки зору їх фінансового стану показує, з одного боку, достатньо високий рівень зацікавленості в даній темі як вітчизняних так і іноземних дослідників, з іншого – наявність деяких прогалин в даній області. Так, застосування саме системи моделей не має широкого висвітлення, автори надають перевагу застосуванню переважно невеликої кількості показників, що описують діяльність банківських установ з одного боку. При цьому часто застосовуються невеликі набори даних, за невеликий проміжок часу, дослідження є статичними, тобто розглядають відібрані показники одномоментно. Частина досліджень швидко втрачає актуальність, так як запропоновані в них методи та моделі не є робустними, тобто не масштабуються на нові масиви даних, по новим періодам, або при використанні даних по інших країнах. На думку автора, доцільним є застосування системи моделей на основі алгоритмів машинного навчання, а саме зниження розмірності та кластеризації, що може бути легко адаптований до нових даних, із застосуванням ретельно відібраних показників, таких що найбільш повно та всебічно описували б функціонування банківської установи, а сам процес моделювання має підпорядковуватись чіткому та послідовному методологічному підходу в рамках якого передбачається відбір найкращого методу (методів) первинної обробки даних та попередній аналіз самих даних на предмет їх придатності до кластеризації. Такий комплексний підхід не має широкого застосування та висвітлення в світовій чи вітчизняній літературі, що зумовлює актуальність, мету і завдання даного дослідження.

Зв'язок з науковими програмами, планами, темами. Дослідження виконувалось відповідно до тематики науково-дослідних робіт Київського національного економічного університету імені Вадима Гетьмана та безпосередньо пов'язане із такими темами: ініціативна науково-дослідна тема «Розвиток

методології та інструментарію моделювання економічних систем у контексті підвищення економічної безпеки» (державний реєстраційний номер 0116U001428, 2016-2022 рр.), ініціативна науково-дослідна тема «Технології Дата Сайнс у дослідженні економічних процесів» (державний реєстраційний номер 0122U001994, з 2022 р.) та науково-дослідна робота за кошти державного бюджету «Моделювання ризиків та безпеки в цифровій економіці» (державний реєстраційний номер 0120U102151, 2020-2022 рр).

Мета і завдання дослідження. Метою наукового дослідження є вирішення задачі аналізу фінансового стану вітчизняних комерційних банків через реалізацію системи економіко-математичних моделей та алгоритмів машинного навчання на основі ґрунтовного методологічного підходу до аналізу фінансового стану банківських установ.

Для досягнення даної мети у дисертаційній роботі поставлено та вирішено наступні завдання:

- Розглянути стан банківського сектору на досліджуваний період з 2015 по 2021 роки а також поточні тенденції;
- Сформувати теоретичне підґрунтя на основі детального аналізу предметної області та понятійного апарату дослідження, серед іншого: систематизувати ключові поняття предметної області та їх взаємозв'язки;
- Провести якісний аналіз банківських ризиків, розглянути їх основні класифікації, дослідити взаємозв'язки між різними видами ризиків;
- Розглянути найбільш широко вживані вітчизняні та іноземні класичні моделі дослідження ризику комерційних банків та комплексні моделі діяльності підприємств, запропонувати методи та моделі для проведення дослідження та обґрунтувати їх вибір;
- Розробити та запропонувати загальний методологічний підхід до аналізу фінансового стану комерційних банків
- Сформувати початковий набір показників діяльності комерційних банків. Провести процедуру відбору ознак в модель на основі підходу, такого, що в результаті забезпечував би високу роздільну здатність кластеризації та

загальну якість моделі, адекватність інтерпретації результатів;

- Запропонувати моделі та алгоритми для вирішення задачі двовимірного проектування багатовимірних даних та задачі кластеризації. Розглянути їх математичний апарат та розкрити логіку та етапи роботи. Виділити та розглянути основні метрики для оцінки якості кластеризації.
- Спроектувати та реалізувати систему моделей, що відповідатиме за двовимірне проектування багатовимірних даних та кластеризацію двовимірних проекцій, провести оптимізацію всіх алгоритмів на основі експериментів з гіпер-параметрами, випробувати різні методи первинної обробки даних;
- Провести аналіз набору вхідних даних на предмет їх придатності до кластеризації;
- Провести аналіз адекватності запропонованого методологічного підходу та побудованої системи моделей через доведення наступних гіпотез:
 - Відібрані алгоритми здатні на достатньому рівні зберігати локальну та глобальну структуру багатовимірних даних, що уможливорює подальшу економічну інтерпретацію результатів та кластеризацію двовимірних проекцій і забезпечує їх адекватність;
 - Всі, або більшість банків мають схожі показники, за якими можливо охарактеризувати проблемність того чи іншого банку, і можна очікувати, що в результаті зниження розмірності даних, за умови збереження їх глобальної структури, проблемні банки будуть згруповані разом;
 - Різкі або постійні зміни (коливання) в показниках роботи банку можуть свідчити про наявність проблем в їх операційній діяльності, і такі зміни будуть супроводжуватись переміщенням банків по двовимірній карті-проекції;
- Провести комплексне експериментальне дослідження з метою аналізу фінансового стану вітчизняних комерційних банків. Провести економічну інтерпретацію отриманих результатів роботи побудованої системи

моделей аналізу фінансового стану, провести профілювання комерційних банків згідно отриманим результатам алгоритму кластеризації;

- Провести порівняльний аналіз ряду найбільш вживаних алгоритмів зниження розмірності та кластеризації на основі загальноприйнятих метрик оцінки якості кластеризації з метою визначення ефективності запропонованої системи алгоритмів та моделей.

Об'єктом дослідження є фінансовий стан комерційних банків України.

Предмет дослідження – інструментарій економіко-математичного моделювання та машинного навчання, що застосовуються для аналізу фінансового стану комерційних банків.

Методами дослідження є статистичні методи, методи синтезу та аналізу, системного аналізу, критичний метод та метод порівняльного аналізу, графічні методи, економіко-математичного моделювання, методи факторного аналізу та методи відбору ознак, алгоритми машинного навчання, а саме: алгоритми багатовимірної проекції (або зниження розмірності) призначені для роботи з багатовидами (t-SNE та UMAP), та алгоритми ієрархічної (деревовидної) кластеризації (агломеративна ієрархічна кластеризація методом Уорда).

Інформаційною базою дослідження є відкриті щоквартальні дані Національного банку України по зведеним балансовим залишкам комерційних банків за період з I кварталу 2015 року по IV квартал 2021 року.

Наукова новизна одержаних результатів полягає у вирішенні науково-практичної задачі аналізу фінансового стану комерційних банків України на підґрунті методів та засобів машинного навчання, розробці нового комплексного ґрунтового методичного підходу до аналізу фінансового стану банківських установ для підтримки управлінських рішень щодо вибору комерційного банку, зокрема як інвестиційного об'єкту, експериментальному дослідженні ефективності різних моделей та алгоритмів машинного навчання в контексті вирішення поставлених завдань, поглибленні теоретичних засад у питанні дослідження фінансового стану банків та банківських ризиків. У процесі дослідження отримані такі наукові результати:

вперше:

- запропоновано комплексний послідовний підхід до проведення аналізу фінансового стану вітчизняних комерційних банків, на основі якого спроектовано та реалізовано систему моделей та алгоритмів машинного навчання, а саме алгоритмів зниження розмірності та алгоритму кластеризації. Це забезпечує відносну простоту реалізації, високу якість візуалізації отримуваних результатів та роздільної здатності двовимірного проектування багатовимірних даних, а отже і адекватність подальшої економічної інтерпретації отриманих кінцевих результатів дослідження;
- запропоновано авторське бачення щодо формування набору ознак в модель, на основі аналітичного та експертного підходів. Це забезпечує всебічний та більш повний аналіз функціонування комерційних банків і відповідно більш надійний аналіз їх фінансового стану. Відібраний набір показників характеризує активи, пасиви, капітал, ліквідність, фінансову ефективність, податкове навантаження комерційних банків та враховує системність (вагу активів в системі);

удосконалено:

- та систематизовано понятійний апарат предметної області дослідження, надано авторські визначення деяких його ключових елементів, запропоновано схему їх взаємозв'язків між собою;
- та систематизовано наявні класифікації банківських ризиків, запропоновано авторську класифікацію банківських ризиків, що враховує види ризиків представлені в нормативній документації, працях та дослідженнях інших вчених;
- та обґрунтовано в рамках методологічного підходу побудову системи моделей та алгоритмів для підтримки аналізу фінансового стану комерційних банків, на підґрунті алгоритмів зниження розмірності t-SNE та UMAP та алгоритм агломеративної ієрархічної кластеризації методом Уорда. Така система моделей дозволяє отримувати двовимірні проекції, що в результаті оптимізації гіпер-параметрів, забезпечує високу якість

збереження як локальної так і глобальної структури даних, що в свою чергу надає можливість провести кластеризації двовимірних проекцій, отримавши при цьому адекватні з точки зору економічної інтерпретації результати;

дістали подальшого розвитку:

- загальнонаукові положення щодо застосування обраних для проведення експериментального дослідження алгоритмів зниження розмірності та кластеризації в контексті вирішення задачі проведення комплексного аналізу фінансового стану комерційних банків.

Практичне значення одержаних результатів. Запропонований автором методологічний підхід стосовно підтримки процесу аналізу фінансового стану комерційних банків на базі системи алгоритмів зниження розмірності та кластеризації можуть бути корисними в першу чергу вкладникам (як фізичним так і юридичним особам), українським та іноземним інвесторам, що розглядають банківські установи як об'єкти інвестування.

Основні положення та результати проведеного дослідження можуть представляти науковий та практичний інтерес для органів державного регулювання у фінансовій сфері в цілому, та у сфері банківської діяльності зокрема, для рейтингових та консалтингових агенцій.

Підхід і набір моделей, методів та алгоритмів, застосований в нашій роботі, може бути досить легко екстрапольований на інші країни, однак варто уважно віднестись до формування набору показників.

Результати дослідження набули практичного впровадження в діяльності виробничої компанії ТОВ «Інтеркабель Київ» (довідка № 161/1.1-07 від 10.03.2023) та ІТ компанії ОУ «Wild Web Art» (довідка №11/08-23 від 08.03.2023). Розроблені автором система економіко-математичних моделей та алгоритмів та методологічний підхід для їх застосування з метою аналізу фінансового стану комерційних банків використовуються цими організаціями для обґрунтування управлінських рішень щодо вибору банків-партнерів та об'єктів для інвестування.

Апробація результатів дисертації. Основні положення та результати

дисертації доповідались і обговорювались на міжнародних і всеукраїнських наукових, науково-методичних, науково-практичних та науково-технічних конференціях, основними з яких є: 83-я наукова студентська конференція «Інноваційний ресурс соціально-економічного прогресу України» (25-27 квітня 2016 р., м. Київ); Перша національна науково-методична конференція «Економіко-математичне моделювання» (30 вересня – 1 жовтня 2016 р., м. Київ); II Всеукраїнська науково-технічна конференція з міжнародною участю «Комп'ютерне моделювання та оптимізація складних систем» (1-3 листопада 2016 р., м. Дніпро); 84-а наукова студентська конференція «Інноваційний ресурс соціально-економічного розвитку країни» (24 квітня – 23 травня 2017 р., м. Київ); IV Міжнародна науково-методична конференція «Математичні методи, моделі та інформаційні технології в економіці» (18-19 квітня 2019 р., м. Чернівці); IX Міжнародна науково-практична конференція «Економіка підприємства: сучасні проблеми теорії та практики» (11-12 вересня 2020 р., м. Одеса); Міжнародна науково-практична конференція «Актуальні проблеми економіки, фінансів, обліку та права в умовах сучасних викликів» (25 серпня 2021 р., м. Полтава); VII Міжнародна науково-технічна конференція «Комп'ютерне моделювання та оптимізація складних систем» (3-5 листопада 2021 р., м. Дніпро).

Особистий внесок здобувача полягає в одноосібно виконаному науковому дослідженні, яке відображає авторський підхід до побудови системи моделей для проведення аналізу фінансового стану комерційних банків на базі алгоритмів зниження розмірності та кластеризації. Усі наукові результати, які викладено в дисертації, одержані автором самостійно, з наукових праць, виданих у співавторстві, використано лише ті матеріали, які є результатом особистої роботи здобувача.

Публікації за темою дисертації. Основні результати дисертаційного дослідження опубліковано у 15 наукових працях загальним обсягом 7,7 друк. арк. (з яких особисто автору належать 7,04 друк. арк.), з них: 5 у наукових фахових виданнях України, що входять до міжнародних наукометричних баз, 1 - у зарубіжному виданні, що входить до наукометричної бази, 9 - в інших наукових

виданнях.

Структура та обсяг дисертації. Дисертаційна робота складається з вступу, трьох розділів, висновків, списку використаних джерел, додатків. Загальний обсяг дисертації складає 215 сторінок, 29 рисунків на 1 сторінках, 7 таблиць на 5 сторінках, 7 додатків на 28 сторінках. Список використаних джерел складає 194 позицій.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ЗАСАДИ МОДЕЛЮВАННЯ ФІНАНСОВОГО СТАНУ КОМЕРЦІЙНИХ БАНКІВ

1.1. Аналіз поточного стану банківської системи та тенденцій її розвитку

Фінансовий сектор України в цілому, та її банківський сектор зокрема, можна назвати високо-динамічними системами, такими що постійно змінюються, іноді демонструють значні коливання показників, що ці системи характеризують. Закономірно, що пікові коливання приходяться на кризові періоди та періоди після-кризового відновлення. Зазначимо, що проведення аналізу є важливим навіть за стабільності банківського середовища, однак безперечно актуальність питання дослідження фінансового стану банківського сектору та комерційних банків як його складових є тим вищим, чим менш стабільним є банківська система. Саме тому має сенс розпочати дослідження саме з аналізу поточного стану банківського сектору, визначити основні тенденції що притаманні йому станом на початок року, в якому проводиться дослідження.

В першу чергу, розглянемо динаміку кількості активних банків, для цього звернемося до Рис. 1.1.

На Рис. 1.1 зображений графік, що показує зміну кількості активних банків в банківській системі України за 6 крайніх років до початку повномасштабної війни [1]. Під лінією графіку вказано кількість активних банків в системі на кожен період, над лінією графіку вказано дельту, тобто зміну в кількості активних банків по відношенню до попереднього періоду. Можна спостерігати суттєве зменшення кількості банків за розглядуваний період: з 2016 року було закрито 44 банки, тобто більше третини всіх активних установ.

Загалом, велика чистка банків почалася ще з 2013 року [2], за весь період очищення банківської системи до початку 2022 року, з неї вибуло 108 банків (сюди входять також банки, що пройшли через процедуру злиття). З них тільки 18 установ

закрилися з причин, не пов'язаних з порушеннями банківської діяльності. Це свідчить про суттєву незрілість системи станом на момент початку її очищення.

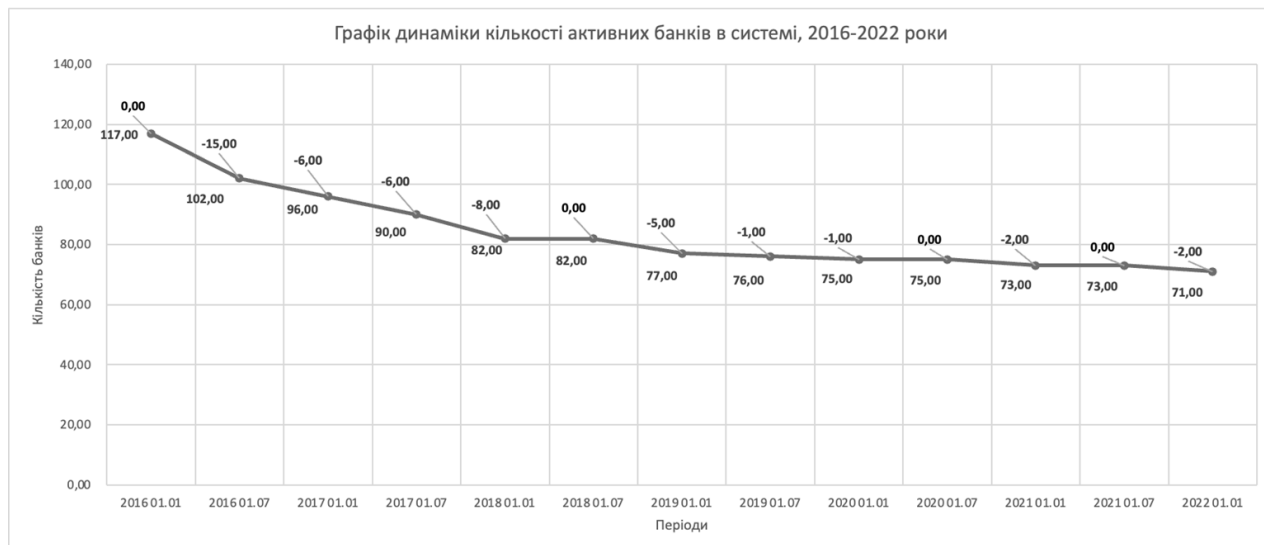


Рис. 1.1. Графік динаміки кількості активних банків в банківській системі України

Джерело: авторські розробки на основі [1]

З іншого боку, це також свідчить про успішне виконання «Комплексної програми розвитку фінансового сектору України до 2020 року» [3], що серед іншого передбачає проведення очистки фінансового сектору та виведення неплатоспроможних банків з ринку. Серед заявлених можливостей та позитивних змін після очищення зазначаються: підвищення надійності вкладів; підвищення активності кредитування малого та середнього бізнесу; збільшення обсягів кредитних ресурсів, доступних для бізнесу; зниження вартості кредитів; забезпечення прозорості власності банківських підприємств; підвищення стійкості банків до макроекономічних шоків, зміцнення показників капіталу та ліквідності; підвищення ефективності функціонування монетарної політики.

Наразі діє новий документ, Стратегія Національного банку України до 2025 року [4], що серед іншого передбачає виконання наступних цілей: збереження зваженої монетарної політики, режиму інфляційного таргетування, плаваючого курсоутворення на ринку; приведення нагляду за банківськими та небанківськими установами до стандартів ЄС; залучення населення до використання фінансових послуг; створення законодавства для платіжного ринку, такого що стимулював би

розвиток інновацій на ринку; підвищення операційної ефективності НБУ та початок цифрової трансформації.

В цілому, зменшення кількості банків в системі, так само як і деякі інші результати виконання стратегії НБУ до 2020 року, можна віднести до позитивних тенденцій. А дотримання нової стратегії має суттєво посприяти виведенню вітчизняного банківського сектору на новий рівень, наблизивши його до європейських стандартів банківської справи.

Розглянемо динаміку обсягів кредитних операцій та частки непрацюючих кредитів, для цього звернемося до Рис. 1.2.

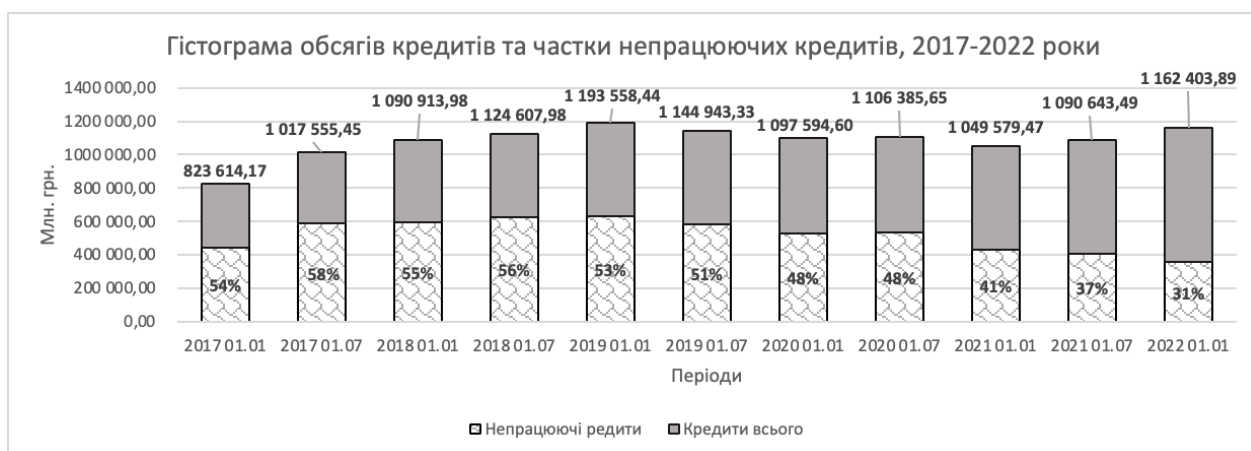


Рис. 1.2. Гістограма обсягів кредитів та частки непрацюючих кредитів, 2017-2022 роки

Джерело: авторські розробки на основі [1]

Як можна побачити з гістограми на Рис. 1.2, станом на початок 2022 року спостерігається найнижчий показник частки непрацюючих кредитів за весь розглядуваний період (період за який доступні статистичні дані НБУ по даному показнику). При цьому обсяги кредитування є досить високі, так в порівнянні з початком 2017 року вони виросли на 338 789 млн. грн., тобто на 41%. Обсяги активів в системі в цілому за цей період також суттєво зросли, збільшившись з 1 256 298 млн. грн. до 1 979 298 млн. грн., тобто аж на 723 000 млн. грн., або на 57,5%. Отже, в контексті показників обсягів та якості позик та активів також спостерігаємо в цілому позитивну тенденцію, проте варто відзначити, що ще у 2020 та 2021 році частку непрацюючих кредитів можна було охарактеризувати як досить високу. І навіть станом на початок 2022, незважаючи на відчутне зниження даного

показника, він все ще є вищим за той рівень, який необхідний для відповідності банківським стандартам ЄС. При цьому зниження частки неякісних кредитів може бути не таким суттєвим на фоні зростання загальних обсягів позик.

Інший аспект, якому варто приділити увагу це фінансова ефективність. Розглянемо його в контексті показників рентабельності активів (ROA) та рентабельності капіталу (ROE), динаміку яких зображено на Рис. 1.3.



Рис. 1.3. Графік динаміки показників ROA та ROE по банківському сектору за період з початку 2016 по початок 2022 роки

Джерело: авторські розробки на основі [1]

З Рис. 1.3 можемо зробити висновок, що фінансова ефективність банківської системи також зазнає росту за останній рік, спостерігається зростання ROA на 1,37% та ROE на 13,68%. Однак тут сказати про однозначно позитивну тенденції складно, так як очевидною є відносно висока флуктуація показників. Так, видно, що з початку 2019 року відбувається покращення ситуації, проте за 2019-2021 роки показники ROA та ROE періодично демонструються як зростання так і спад, з чого можна зробити припущення про нестабільність фінансової ефективності вітчизняних банківських установ.

Іншим показником який варто розглянути є загальна кількість структурних підрозділів банків. Так, з 2016 року кількість підрозділів по всім банкам зменшилася з 11 871 одиниці до 6 685 на початку 2022 [1]. При цьому кількість підрозділів, що закрились у зв'язку з вибуттям банків з ринку складає всього 1 318 одиниць, отже за розглядуваний період 3 868 структурних підрозділів були закриті,

що може свідчити про скорочення операційної діяльності та ділової активності деяких банків, що в свою чергу є негативною тенденцією. З іншого боку така тенденція може пояснюватись поступовим переходом клієнтів банків в онлайн, оптимізацією бізнес-процесів банків.

Ще однією позитивною тенденцією, яку можна виділити, це поступовий перехід до сучасних європейських стандартів банківського нагляду. На сьогоднішній день банківська система України все ще більшою мірою функціонує на засадах Базеля II, проте можемо спостерігати покрокове впровадження елементів наступного покоління стандарту, Базеля III, яким вже певний час керуються банківські системи країн ЄС. Так, очищення банківського сектору та цілі передбачені старою та новою стратегіями є елементами такого переходу, так само як і впровадження таких нових важливих для розвитку банківського сектору показників, як LCR та NSFR (останній з 1 квітня 2023 року встановлено на рівні 100%), більш детально розглянуті в роботах Ларіонової К. Л. [5] та Миненко Л. М. [6]. Також, починаючи з 2020 року було введено в дію механізм визначення мінімального розміру операційного ризику та відповідного розрахунку нормативів достатності капіталу під операційні ризики, у відповідності до Базель III.

З 24 лютого 2022 року в Україні розпочалося повномасштабне вторгнення та запроваджено воєнний стан, для економіки та зокрема банківського сектору розпочався новий кризовий період, що характеризується нестабільністю та дисбалансом. Аналізуючи стан банківського сектору у 2022 році, можна зробити висновок про відсутність сталої тенденції до поліпшення чи погіршення ситуації, спостерігається деяка волатильність. Так, можна відзначити стрімкий ріст банківських активів на 17,9% до 2,7 трлн. грн., однак пояснюється це зокрема вимушеною емісією гривні з боку НБУ для забезпечення покриття потреб, що виникли під час війни. За I квартал 2023 року активи зросли на 8,6%, спостерігається зниження обсягів чистих кредитів та зниження чистого корпоративного кредитного портфелю, відзначається збільшення питомої ваги непрацюючих кредитів на 4,5 в. п. до 38,1% за квартал. Продовжується суттєве скорочення мережі відділень: за 2022 рік кількість підрозділів зменшилась з 6685

до 5336 (на 20%). З ринку вибуло 6 банківських установ з початку війни, але таку статистику можна вважати в межах норми навіть для довоєнного часу; показники рентабельності начебто демонструють позитивну картину, рентабельність капіталу (ROE) по системі знаходиться на рівні 10,9%, однак джерелом прибутків банків в основному є процентні доходи від вкладень в ОВДП та депозитні сертифікати НБУ. За підвищеного резервування під кредитні ризики (на рівні, що був би цілком адекватним економіці воєнного часу), фінансові результати могли б бути значно менш оптимістичними. За 2022 рік 21 банківська установа (в тому числі 2 державні банки) задекларували збиток сумарно на 20,8 млрд. грн., хоча в цілому банківський сектор показав прибуток у 17,3 млрд. грн. у IV кварталі 2022 року [1, 7]. На сьогоднішній день ще не спостерігається активного висвітлення поточного стану банківської системи вітчизняними вченими, однак з наявних досліджень можна виділити роботу Еркес О. зі співавторами, що присвячена аналізу трансформаційних процесів банківського ринку, що виникли чи активізувалися під час війни в Україні [8]. Дисертаційне дослідження зосереджене на між-кризовому інтервалі, з I кварталу 2015 року по IV квартал 2021 року, так як ми вважаємо, що поточний період глобальної економічної нестабільності потребує проведення окремого ґрунтовного дослідження але вже після завершення війни, для забезпечення об'єктивності, цільності, адекватності як теоретичних так і практичних результатів такого дослідження.

В цілому, якщо говорити про досліджуваний період 2015-2021 років, незважаючи на деякі позитивні тенденції, що були розглянуті вище, в банківському секторі України спостерігались проблемні аспекти, які не дають можливості стверджувати про однозначну надійність та ефективність комерційних банків в системі: все ще висока частка неякісних активів, зокрема кредитів; високий рівень кредитного ризику в цілому; низька ефективність роботи частини системно важливих банків; періодичне «просідання» рентабельності активів та капіталу, та нестабільність даних показників. Для досягнення рівня стандартів ЄС українська банківська система продовжує виконання поставлених завдань навіть в умовах війни, однак попереду все ще лежить довгий шлях

1.2. Сутність, особливості та аналіз фінансового стану комерційних банків

Побудова системи моделей аналізу фінансового стану комерційних банків, перш за все, має ґрунтуватись на чіткій та однозначній теоретичній базі, системі категорій та понять на яких базується предметна область дослідження. Тому, першим важливим і невід'ємним кроком дослідження є ознайомлення з понятійним апаратом, що використовується в процесі його проведення та викладення. Серед ключових категорій в даній роботі є наступні: фінансовий стан, ефективність діяльності, стійкість, надійність, платоспроможність, невизначеність та ризик, прибутковість. В наступних підрозділах ми детальніше ознайомимося з даними поняттями та побудуємо схему взаємозв'язку категорій.

В першу чергу, розглянемо поняття **фінансового стану** як основоположне в рамках даної роботи, і яке, як можна буде прослідити пізніше, є тісно пов'язаним з іншими категоріями категоріального апарату дослідження.

Його визначення можна зустріти в багатьох роботах. Так, Поддєрьогін А. М. зі співавторами визначають фінансовий стан як «комплексне поняття, що є результатом взаємодії всіх елементів системи фінансових відносин підприємства» [9]. На їх думку, фінансовий стан визначається рядом факторів виробничо-господарського характеру, та рядом показників що описують наявність, розміщення та використання фінансових ресурсів.

Білик М. Д., розглядаючи поняття фінансового стану, більше уваги приділяє здатності підприємства розраховуватись за власними зобов'язаннями, забезпечити стабільне фінансування операційної діяльності, сталого [10], а Цал-Цалко Ю. С. пов'язує фінансовий стан з фінансовою стійкістю підприємства [11].

Досить глибокому дослідженню теоретичних основ фінансового стану підприємства присвячено статтю Чепки В. В. [12]. В ній же представлено схему співвідношення різних фінансових категорій, в якій фінансову стійкість визначено як складовий елемент фінансового стану.

Перейдемо до більш вузького поняття фінансового стану, а саме до категорії **фінансового стану комерційного банку**. Розглянувши ряд джерел, присвячених

дослідженню даної категорії, можна зробити висновок, що в основному її розглядаються в двох аспектах. Перший: це ресурсний аспект. В роботах Загороднього А. Г. [13] та Дзюблюка О. В. [14] фінансовий стан комерційного банку характеризується грошовими, фінансовими ресурсами, що необхідні для ведення підприємством господарської діяльності, підтримки операційної стабільності, проведення розрахунків з контрагентами. В роботі ж Герасимовича А. М. [15] фінансовий стан комерційного банку розглядається як система певних показників, що характеризує ефективність управління його активами, пасивами, достатність капіталу, якість активів, ліквідність, ділову активність та ефективність менеджменту. Це є другий аспект. Більш детально теоретичні аспекти даної категорії розглянуто в статті Вінниченко А. В. [16].

Ми пропонуємо під фінансовим станом банку розглядати його здатність забезпечувати надійність та ефективність своєї роботи: забезпечувати ефективний ризик-менеджмент, підтримувати високу якість активів, здатність залучати ресурси, забезпечувати адекватний рівень капіталу та ліквідності для захисту депозитів на випадок паніки вкладників та позик на випадок падіння їх якості, і в результаті генерувати прибуток та забезпечувати достатній рівень маржі. При цьому, на нашу думку, в контексті нашого дослідження фінансовий стан все ж таки варто розглядати саме з позицій певних показників, що описують різні сторони діяльності банківських установ направлені на виконання вищевказаних завдань.

Перейдемо до поняття **ефективності діяльності**. Загалом, тема ефективності є досить популярною і поширеною, в тому числі в науковій літературі. Моделі та методи управління ефективністю діяльності банку з точки зору управління активами та пасивами банку розглянуто в підручнику за авторством Кузнецової С. А. [17], спроба адаптації загальноекономічних моделей для прогнозування показників ефективності банківської діяльності здійснена у роботі Шамріна Р. В. [18].

Ефективність діяльності підприємства є досить багатограним поняттям, що також згадується в літературі як економічна ефективність.

У підручнику за авторством Суліми Н. М. подане таке визначення ефективності: «інтегрована економічна категорія, яка відображає виробничі відносини щодо економії сукупних затрат уречевленої та живої праці на отримання кінцевого позитивного результату». Також в роботі наголошується на нетотожності понять «ефективність» та «результативність». Як пише автор, «по своїй суті результативність має цільову природу і належить до сфери оцінювання отриманих наслідків від реалізації господарських рішень чи діяльності за певний період. Водночас, ефективність використовується для характеристики використання виробничих ресурсів для досягнення поставлених цілей» [19].

У зарубіжних дослідженнях поняття ефективності часто пов'язують з правильним, оптимальним розміщенням ресурсів з метою уникнення перевиробництва чи дефіциту [20]; також часто під ефективністю розглядається «ефективність за Парето», тобто такий розподіл ресурсів, за якого будь-які додаткові зміни не будуть приносити маржинальної корисності [21, 22].

Перейдемо до більш специфічного поняття **ефективності банківської діяльності** та розглянемо можливі загальні підходи її оцінки.

У статті Коваленко В. В. представлена цікава думка, щодо виділення трьох рівнів банківської системи, стосовно кожного з яких застосовується своє поняття ефективності та її оцінки: «ефективність банківської діяльності розглядається на трьох організаційно-функціональних рівнях її прояву – система банків другого рівня, окрема банківська установа та структурний підрозділ банку. Між цими рівнями існує тісний взаємозв'язок і кожен із них впливає на результати діяльності як вищого, так і нижчого рівня» [23]. У статті розглянуто систему показників саме для верхнього рівня, тобто для банківської системи в цілому. Це показники операційної ефективності та показники рентабельності.

Поняття ефективності банківської діяльності розглядає Єріс Л. М. [24] у своїй статті у двох аспектах: у короткостроковому періоді ефективність банківської діяльності означає здатність адекватно управляти наявними грошовими потоками з метою максимізації прибутку, зокрема зниження ефективності в умовах конкуренції призводить до втрати потенційного прибутку та зниження фінансової

стійкості на ринку; у довгостроковому періоді ефективність банківської діяльності означає здатність забезпечити перетворення заощаджень економічних агентів в інвестиції з найменшими трансакційними витратами, що також впливає на конкурентні позиції банку на ринку. Автор робить важливий висновок про доцільність розгляду поняття ефективності як складного системного явища у потрібному вимірі «прибутковість – ризикованість – надійність». Тут ми цілком погоджуємося і вважаємо, що ефективність насправді тісно пов'язана з іншими категоріями. Зроблено також висновок про те, що «критеріями ефективності банківської діяльності повинні бути як безпосередньо фінансові результати його діяльності (дохід і прибуток), так і результативність (рентабельність), а також вся сукупність показників фінансового стану (стійкість, ліквідність, платоспроможність), досягнутих банком з урахуванням їх ціннісного або цільового значення як для самого банку, так і для соціально-економічного середовища його діяльності». Автор наголошує, що «сукупність критеріїв треба розглядати як систему і комплексну характеристику, яка відображає відповідність результатів діяльності банку як суб'єкта господарювання поставленим цілям на кожному часовому інтервалі його функціонування, і в цьому аспекті тільки досягнення всіх, а не декількох, критеріїв дозволить говорити про економічну ефективність банківської діяльності як якісного управління його грошовими потоками».

Питання дослідження ефективності різних аспектів банківської діяльності представлено також в роботах Герасимовича А. М. [25] та Парасій-Вергуненко І. М. [26].

На нашу думку, ефективність банківської діяльності це дійсно багатогранне поняття, економічна категорія що всебічно описує функціонування банківської установи зокрема чи банківської системи в цілому; це не тільки здатність суб'єкту господарювання ефективно використовувати ресурси (власні чи залучені) з метою максимізації прибутку, а й здатність забезпечувати високу надійність та адекватний механізм управління ризиками з метою підвищення надійності та конкурентоспроможності на ринку.

Також, розглянувши два поняття – фінансовий стан та ефективність діяльності комерційного банку – можна відзначити багато спільного: обидві категорії характеризують ресурси підприємства, та його здатність їх ефективно застосовувати; обидві категорії тісно пов'язані з набором певних показників, що описують різні сторони роботи підприємства: дохідність та прибутковість, якість активів та пасивів, їх збалансованість, ліквідність, достатність капіталу, ефективність залучення персоналу та роботи менеджменту тощо. Взаємозв'язок же між даними поняттями може бути не таким очевидним. Ми вважаємо, що ефективність комерційного банку обумовлена його фінансовим станом, з огляду на те, що нами було вирішено розглядати фінансовий стан з точки зору показників діяльності. Тому ми визначаємо ефективність діяльності як відповідність фінансового стану банківської установи такому рівню, що надавав би їй можливість вести безперебійну діяльність направлену на отримання прибутку через підтримання основних бізнес-ліній з проведення кредитно-інвестиційних, комісійно-розрахункових та торгових операцій, та збереження власної надійності та платоспроможності, серед іншого через підтримання рівня капіталу та ліквідності.

Поглиблений аналіз ефективності банківської діяльності було проведено науковцем Примосткою Л. О. [27]. Неабияку цінність представляє представлена автором логічно-структурна модель. Автор у своїх дослідженнях звернувся до підходів іноземних вчених, а саме до декомпозиційної моделі Дюпона, що зображена на Рис. 1.4.

Як видно з Рис. 1.4., на верхньому рівні моделі прибутковість, як і ризик, визначається одним з ключових компонентів, що відіграють важливу роль для загального аналізу ефективності банківської діяльності.

Розглянемо власне саме поняття **прибутковості**. Воно є досить однозначним і характеризує здатність підприємства генерувати позитивний фінансовий результат від будь-якої своєї діяльності. Такий результат утворюється внаслідок перевищення загальних доходів суб'єкта господарювання над його загальними видатками.

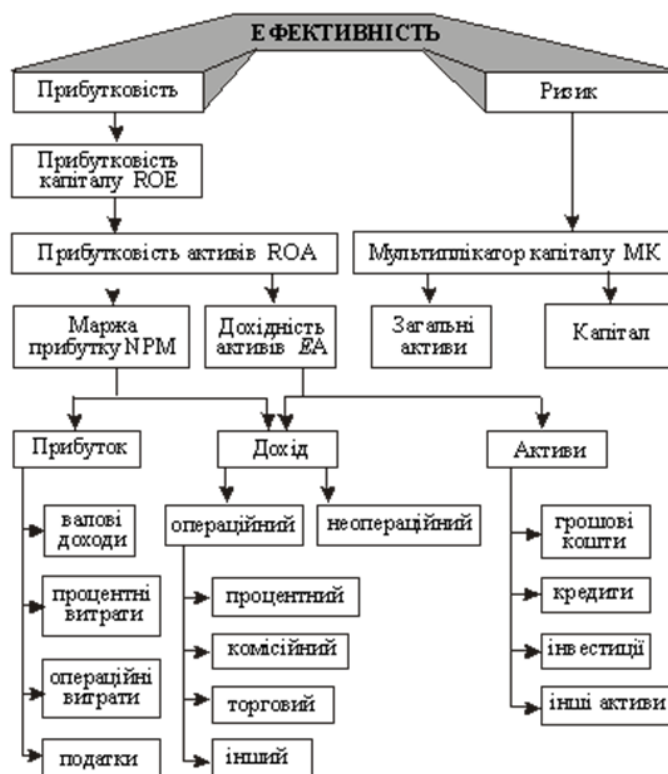


Рис. 1.4. Логічно-структурна модель факторного аналізу ефективності діяльності комерційного банку

Джерело: [27]

Специфічними особливостями прибутковості для банку є джерела отримання його прибутку. Це більшою мірою активні, кредитно-інвестиційні операції, основою яких є кредити фізичним та юридичним особам. Менш об'ємними зазвичай є торгівельні операції (наприклад, операції з купівлі/продажу валюти на валютному ринку) та комісійні операції, пов'язані з розрахунково касовим обслуговуванням. Перелічені операції можуть відповідно приносити банку процентні, торгівельні та комісійні доходи (або витрати). Прибутковість банківської діяльності буде характеризуватись перевищенням сумарного доходу від усіх видів діяльності над витратами.

В науковій літературі визначення прибутку добре подано у роботі Криклія О. А. та Маслака Н. Г., як «виражений у грошовій формі дохід власника на вкладений капітал, плата за працю та ризик ведення підприємницької діяльності, що є різницею між сукупним доходом і сукупними витратами в процесі здійснення

такої діяльності» [28]. Інші визначення прибутку та аналіз категорії прибутку комерційного банку в цілому, можна знайти в роботі Катана Л. І. [29].

Розглядаючи таку категорію як прибутковість, варто ознайомитись також з поняттям **фінансової стійкості**, що є більш широким, тобто можна сказати, що прибутковість, серед іншого, забезпечує фінансову стійкість підприємства. Дане поняття вже декілька разів згадувалось при розгляді ефективності банківської діяльності та фінансового стану, що наводить на думку про наявність зв'язку між цими категоріями.

Існує багато визначень фінансової стійкості, більш повно їх розглянуто, наприклад, у статті Гапака Н. М. [30]. Серед усіх визначень можемо виділити наступне: «Здатність суб'єкта господарювання функціонувати і розвиватись, зберігати рівновагу своїх активів і пасивів у мінливому внутрішньому і зовнішньому середовищі, що гарантує його платоспроможність та інвестиційну привабливість у довгостроковій перспективі в межах допустимого рівня ризику» [31].

При цьому варто враховувати, що фінансова стійкість банківської установи має свої особливості.

Фінансова стійкість банку – такий його стан, якому властиві збалансованість фінансових потоків, достатність коштів для підтримки своєї платоспроможності та ліквідності, а також рентабельність діяльності. Фінансова стійкість банку означає постійну здатність банку відповідати за своїми зобов'язаннями і забезпечувати прибутковість на рівні, достатньому для нормального функціонування у конкурентному середовищі. Згідно з Науменковою С. В. [32] розрізняють три основні види фінансової стійкості банку за стабільністю його роботи:

а) нормальна стійкість, що описується стабільною діяльністю, відсутністю затримки виконання своїх зобов'язань чи невиклат, стабільною рентабельністю;

б) нестійкий фінансовий стан, який характеризується невчасними платежами, неможливістю у вказаний термін виконувати певні зобов'язання, низьким рівнем рентабельності тощо;

в) кризовий фінансовий стан, який характеризується регулярними невиконаннями, наявністю простроченої заборгованості тощо. Кризовий фінансовий стан може стати причиною того, що банк стане неспроможним фінансувати поточну діяльність, проводити платежі та виконувати свої зобов'язання, а в кінцевому рахунку – збанкрутує.

Довгаль Ю.С. та Чамара Р.О. визначають фінансову стійкість комерційних банків як «такий стан банківської установи, при якому вона є платоспроможною, ліквідною, має достатню величину регулятивного капіталу для покриття всіх можливих ризиків, пов'язаних із банківською діяльністю, і при цьому залишається достатньо рентабельною, тобто її активи використовуються максимально ефективно» [33].

У роботі Рисіна В. В. [34] також розглядається теоретичні підходи до трактування поняття фінансової стійкості банку, а також проводиться систематизація внутрішніх та зовнішніх чинників впливу на фінансову стійкість вітчизняних банків. Крім іншого, в роботі зазначається, що дану категорію часто ототожнюють з категорією надійності, проте відмічається що дані поняття не є тотожні і поняття надійності є вужчим.

Проблема фінансової стійкості банків, як важливої умови для забезпечення ефективності функціонування банку, була також висвітлена у роботах Дзюблюка О. В. [35], Кочеткова В. М. [36] та Примостки Л. О. [37].

Отже, розглянувши визначення фінансової стійкості, подані різними авторами та дослідниками, можемо сформулювати власне визначення як здатність підприємства здійснювати основну та інші види діяльності безперебійно, не зважаючи на ризики й зміни в середовищі бізнесу, які можуть відбуватися в процесі господарської діяльності; забезпечувати достатній рівень доходу для покриття власних витрат і прибутковості для поступового і стабільного розвитку, підтримання рентабельності власних активів та капіталу; формувати достатній рівень капіталу та ліквідності для забезпечення надійності власних активів та ресурсів, підтримання власної платоспроможності; забезпечувати баланс між

власними та залученими ресурсами; підтримувати конкурентоспроможність на банківському ринку.

Фінансова стійкість банківської установи тісно пов'язана з його платоспроможністю та надійністю а саме поняття фінансової стійкості пов'язане з поняттям **ризик**у та **невизначеності**. Їх ми й розглянемо далі, розпочавши з останніх двох категорій.

Невизначеність – це фундаментальна характеристика недостатньої забезпеченості процесу прийняття економічних рішень знаннями стосовно певної проблемної ситуації [38].

Загалом, ризик можна описати як поєднання ймовірності та наслідків настання несприятливих подій. Іноді ризик визначається як конкретний інцидент, що може призвести до фінансових втрат.

Слід також звернути увагу на поняття **економічного ризику**. За визначенням Вітлінського В. В. [39] — це об'єктивно-суб'єктивна категорія у діяльності суб'єктів господарювання, що пов'язана з подоланням невизначеності та конфліктності в ситуації неминучого вибору. Вона відображає міру (ступінь) відхилення від цілей, від бажаного (очікуваного) результату, міру невдачі (збитків) з урахуванням впливу керованих і некерованих чинників, прямих та зворотних зв'язків стосовно об'єкта керування.

Розглянемо ризик, що притаманний сфері предметної області даного дослідження, а саме банківський ризик або ризик банківської діяльності.

Ризики банківської діяльності – комплексна економічна категорія, що характеризується ймовірністю того, що події, очікувані або неочікувані, можуть мати негативний вплив на капітал та/або надходження банку та величиною втрат капіталу банківської установи в результаті реалізації несприятливих подій.

Ризики банківської діяльності випливають із специфіки самої банківської діяльності, що здійснюється в умовах ринкових відносин, і означають імовірність одержання доходів, менших від очікуваних, зниження вартості активів. Високі банківські ризики призводять до значних фінансових утрат і, як наслідок, до банкрутства банків.

У нормативних документах, відповідно до Постанови 64, ризик визначається як імовірність виникнення збитків або додаткових втрат або недоотримання доходів, або невиконання стороною договірних зобов'язань унаслідок впливу негативних внутрішніх та зовнішніх факторів [40].

Питання ризиків банківської діяльності є досить популярним в науковій літературі. Моделюванню різних його видів присвячено багато праць як вітчизняних так і іноземних вчених.

Моделі та методи оцінювання банківського ризику ліквідності (в тому числі з використанням широко вживаної моделі VaR) подані у підручнику за авторством Примостки Л. О. [41]. Методи та моделі оцінки та управління ринковим ризиком було розглянуто у статті Бобиль В. В. [42], методам управління та хеджування відсоткових ризиків приділено увагу у підручнику Вовк В. Я. [43], дослідження операційних ризиків банку проводилися такими авторами як Євтушенко Г. В. [44] та Набок Р. М. [45]. Серед більш сучасних робіт, присвячених питанню банківських ризиків, можна виділити наступні: стаття Волкової Н. І. [46] присвячена аналізу впливу ризиків на фінансову стійкість вітчизняних банків в період з 2015 по 2019 роки, автором розглянуто ряд методів оцінки банківських ризиків та виділено дві головні стратегії механізму нейтралізації таких ризиків; причини і наслідки проявів банківських ризиків в Україні проаналізовано в статті Забчук Г. [47], особливу увагу приділено валютному ризику; досить цікавому та специфічному питанню ризиків легалізації доходів приділено увагу в роботі Мордань Є. Ю. [48]; стаття Кузніченко Я. [49] присвячена дослідженню впливу банківських ризиків на фінансову безпеку банків, представлено авторську структуру організації процесом управління ризиками, наголошується на важливості системи управління ризиками в комерційних банках; Шендрик В. В. [50] в статті детально розглядає корупційні ризики; системному ризику присвячена дисертаційна робота Корнилюка Р. В. [51], в якій досліджено його формування та реалізацію, розглянуто питання його оцінювання, розроблені та запропоновані практичні рекомендації щодо його регулювання.

Окрема увага приділяється кредитним банківським ризикам, які, зважаючи на те що кредитні операції є основним напрямком банківської діяльності, можна назвати найбільш значущими. Кількісному оцінюванню та методам моделювання кредитного ризику приділено увагу у підручнику за авторством і редакцією Вітлінського В. В. [52], питання доцільності моніторингу кредитних ризиків банків-кредиторів АПК з використанням бальних оцінок та інтегрального показника, розрахованого на їх основі, висвітлена у статті Онисько С. М. [53]. Проблеми управління кредитними ризиками банків, та застосування VaR для моделювання ризиків кредитного портфелю досліджені у роботі Мічинської О. Ю. [54]. У статтях авторів Артем'єва О.О. [55], Похилько С. В. [56] та Ларіонової К. Л. [57] можна ознайомитись з більш сучасними дослідженнями, присвяченими різним аспектам кредитного ризику.

Розглядаючи такі категорії як фінансова стійкість та ризик, варто звернути увагу також на такі поняття як надійність, конкурентоспроможність, ліквідність та платоспроможність комерційного банку.

Як вже зазначалось вище, надійність часто ототожнюють з фінансовою стійкістю, проте ми погоджуємось з думкою Рісіна В. В. [34], і також вважаємо, що категорія фінансової стійкості є ширшою за категорію надійності. Ми вважаємо, що надійний банк визначається діяльністю, що призводить до реалізації інтересів конкретного суб'єкта, і пропонуємо розглядати надійність як здатність банку виконувати свої зобов'язання перед вкладниками, кредиторами та іншими контрагентами в повному обсязі в обумовлені терміни, беручи до уваги можливі ризики банківської діяльності, підтримуючи для цього необхідний рівень капіталу та ліквідності.

Категорії, що залишились, не є ключовими в нашому дослідженні, проте тісно пов'язані з іншими категоріями, тому ми коротко ознайомимось з ними і перейдемо до аналізу взаємозв'язку в рамках всього категоріального апарату даного дослідження.

- **Конкурентоспроможність** комерційного банку можна визначити як вигідність його позицій за кількісними та якісними показниками в

порівнянні з іншими учасниками банківської системи, або як ступінь привабливості банку для вже існуючих та потенційних клієнтів.

- **Ліквідність** комерційного банку – спроможність забезпечувати виконання власних грошових зобов’язань через підтримання балансу між строками та сумами погашення розміщених активів та сумами виконання зобов’язань, а також строками та сумами інших джерел і напрямів використання коштів [58].
- **Платоспроможність** комерційного банку – спроможність банку виконати законні вимоги кредиторів, характеризується рівнем забезпеченості фінансових зобов’язань банку власним капіталом [59].

На даному етапі, коли нами був детально розглянутий понятійний апарат, що використовується в рамках даного дослідження, ми можемо спробувати встановити взаємозв’язок між усіма категоріями предметної області. Для цього звернімося до схеми, зображеної на Рис. 1.5.

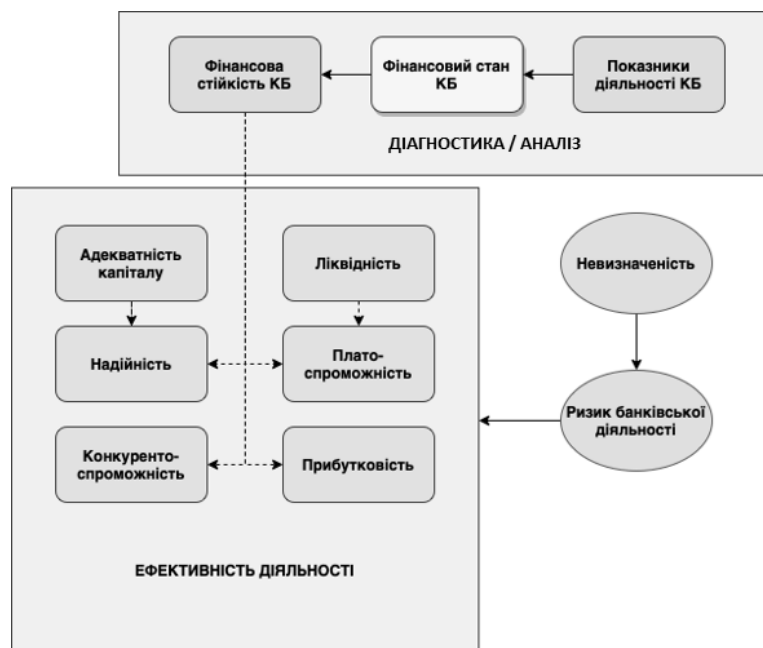


Рис. 1.5. Схема взаємозв’язку категорій предметної області дослідження

Джерело: розроблено автором

Як видно з Рис. 1.5, ключовому в нашому дослідженні є фінансовий стан комерційного банку. Як вже зазначалось раніше, ми притримуємось точки зору, що фінансовий стан визначається набором показників, що мають найбільш повно та всесторонньо описувати діяльність комерційного банку. Фінансову стійкість ми

розглядаємо як властивість комерційного банку вести свою господарську діяльність, проводити кредитно-інвестиційні, торгові та комісійно-розрахункові операції.

При цьому, як вже зазначалось вище, фінансова стійкість також характеризується здатністю банку в процесі господарської діяльності генерувати прибуток, забезпечувати власну платоспроможність, зокрема через підтримку рівня ліквідності та надійності через підтримку адекватності (достатності) власного капіталу, зберігати конкурентоспроможність на банківському ринку.

Прибутковість, конкурентоспроможність, адекватність капіталу та надійність, ліквідність та платоспроможність ми визначаємо як складові компоненти фінансової стійкості, які власне і характеризуються набором показників, що описують фінансовий стан комерційного банку. Як вже було зазначено, ефективність комерційного банку визначається його фінансовим станом: рівень показників, що відповідає оптимальному рівню та стандартам банківської діяльності дозволяє зробити висновок про високу ефективність банківської діяльності, і навпаки, незадовільний рівень показників може свідчити про низьку ефективність. Ризик в даній системі категорій визначається невизначеністю та формується під її впливом. Ризик є одним з ключових елементів і має безпосередній вплив на фінансову стійкість банку через її компоненти: так, різні види ризику в результаті реалізації можуть справити негативний (а іноді і позитивний) вплив на прибутковість, конкурентоспроможність (зокрема, через репутаційний вплив), рівень капіталу та ліквідності. Це спричиняє відповідні наслідки для ефективності діяльності банку та має прямий ефект на рівні показників його фінансового стану. Наша задача – це проведення аналізу саме фінансового стану, що означає дослідження фінансової стійкості комерційних банків через аналіз та моделювання набору показників, що її характеризують.

З проведеного огляду понятійного апарату можна зробити висновок, що в категоріальній системі предметної області нашого дослідження категорія ризику є однією з найбільш значущих, так як реалізація ризику може суттєво вплинути на будь-який компонент фінансової стійкості банку. Ми можемо також сформулювати

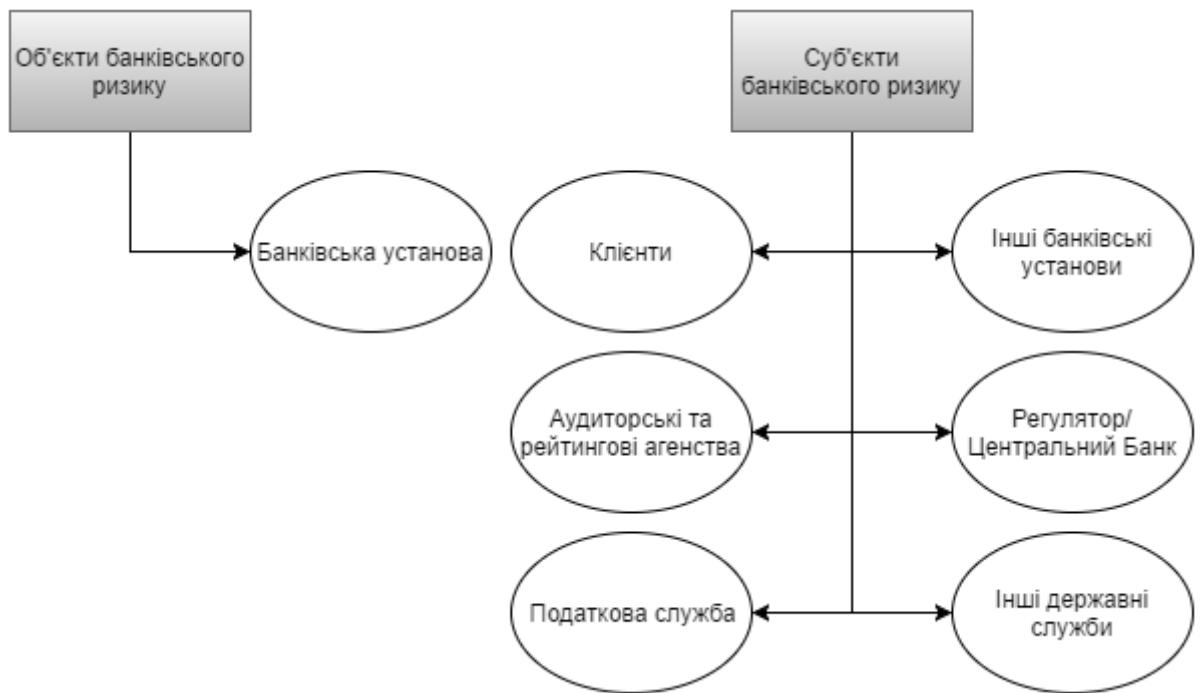
наступне: ризик лежить в основі діяльності будь-якої установи, будь-якого організаційного утворення, що веде фінансову діяльність в цілому, та банківську діяльність зокрема. В банківському секторі різного роду ризики є тісно пов'язаними з результатами діяльності банківських установ, а їх реалізація може спричинити суттєві зміни фінансового стану. В деяких випадках реалізація ризику може навіть призвести до дефолту банківської установи. Ризик є визначним фактором впливу на сам механізм функціонування організацій адже пріоритетним напрямком для будь якого підприємства, особливо в галузі банківської справи, стає встановлення системи ризик-менеджменту та чітке дотримання правил та механізмів управління ризиками. Беручи до уваги важливість ризику в контексті об'єкту нашого дослідження, вважаємо що необхідно зробити акцент на детальному розгляді саме цієї категорії.

Якісний аналіз банківських ризиків, як відзначив Долінський Л. Б. у своїй статті [60], має включати визначення та дослідження таких його елементів (Рис. 1.6):

- Об'єкт ризику – економічна система, ефективність та умови функціонування якої наперед точно невідомі.
- Суб'єкт ризику – особа (колектив), що зацікавлена в результатах керування об'єктом ризику і має компетенцію прийняття рішень щодо об'єкта ризику.
- Джерело ризику — це чинники (явища, процеси), які спричиняють невизначеність результатів, конфліктність у широкому сенсі цього поняття.

Відповідно до Рис. 1.6., об'єктом банківського ризику у нашому випадку є сама банківська установа, що приймає на себе певний ризик, проводячи банківські операції.

Суб'єктом може бути інший банк, що функціонує в банківській системі країни; потенційний або наявний клієнт банку; Центральний Банк, що проводить нагляд за діяльністю банку (НБУ); аудиторські та рейтингові агентства, що здійснюють оцінку діяльності банку; податкова та інші державні служби.



Джерело: авторські розробки на основі [60]

Рис. 1.6. Складові суб'єкту та об'єкту банківського ризику

Важливим питанням в теорії банківських ризиків є дослідження чинників (джерел ризику), що спричинили виникнення банківських ризиків.

Очевидним є той факт, що виникнення банківських ризиків серед іншого пов'язане з тим, що вкрай складним завданням є реалізація та дотримання встановленої банком стратегії, курсу розвитку та діяльності. Це спричинене тим, що існує велика кількість різноманітних чинників, які негативно впливають на діяльність банківської установи, і це, в свою чергу, породжує різного роду ризики в усіх сферах банківської діяльності.

Загалом, усі чинники ризику можна віднести до двох категорій: це чинники контрольовані та неконтрольовані. Контрольовані чинники – такі, на які може вплинути банк в рамках реалізації певних управлінських рішень. Неконтрольовані чинники – ті, що зумовлені випадковими подіями, зазвичай це зовнішні відносно до банку чинники. Вони ніяк не залежать від діяльності робітників банківської установи. У випадку, якщо існує повна та достовірна інформація стосовно чинників ризику, то більшість з них може бути достатньо точно та адекватно ідентифікована. В іншому ж випадку, якщо така інформація відсутня, то вони характеризуються як невизначені. Для того щоб ідентифікувати та провести оцінку таких чинників

ризик, використовується специфічний інструментарій, такий як методи математичної статистики та теорія нечітких множин. В реальності, визначені або невизначені чинники в чистому вигляді зустрічаються вкрай рідко, зазвичай вони мають змішаний характер.

Отже, можемо підсумувати, що чинники ризиків банківської діяльності – це деякі умови, що ведуть до появи певних ризиків в банківській діяльності.

В економічній літературі теж не сформувалося одностайної думки стосовно того, яка існує кількість факторів, що призводять до появи ризиків в діяльності банків та як їх розподілити на однорідні групи.

Наприклад, Вітлінський В. В. у своїй монографії [39] відзначає, що «до чинників ризику слід віднести виключно не детерміновані (ймовірнісні, розпливчасті) процеси, що протікають у банку та його економічному оточенні».

Владичин У. у своїй роботі [61] виділяє зовнішні та внутрішні чинники банківського ризику. Внутрішні включають: кваліфікаційний рівень працівників банку, комплексність структури управління банку, трансформаційні процеси в сфері управління персоналом та інше. Зовнішні ж чинники включають в себе: залучення інформаційних технологій, законодавчі зміни чи зміни ринкової кон'юнктури, певні непередбачені обставини та інше.

Виділяють також макро- та мікроекономічні чинники. Макроекономічні чинники можна охарактеризувати як такі, що є зовнішніми відносно банків, проте їх можна розподілити на два рівні: рівень держави та світовий рівень. Мікроекономічні чинники також можна розподілити на дві групи: такі, що пов'язані з діяльністю самого банку (внутрішні) та такі, що пов'язані з діяльністю контрагентів (зовнішні). На відміну від мікроекономічних, банк не зможе справляти зворотний вплив на макроекономічні фактори, але сам при цьому піддається їх впливу.

Як вже зазначалося раніше, важливо розуміти, що в «чистому» вигляді чинники банківського ризику зустрічаються вкрай рідко, зазвичай вони є змішаними. Також важливим є розуміння того факту, що неможливо сформувати повний, вичерпний та остаточний перелік ідентифікованих чинників ризику, так як

сучасний світ є швидкоплинним, економічні процеси змінюються, фінансовий сектор стрімко трансформується, і тому з'являються нові фактори, нові групи чинників. Отже, їх систематизація повинна обов'язково враховувати можливість періодичного поповнення та оновлення.

Очевидно, що чим більше різноманітних операцій здійснює банк в процесі своєї діяльності, тим більша кількість ризиків, з якою він стикається. Ризикованість діяльності при цьому може підвищуватись як за рахунок розширення спектру операцій, так і за рахунок збільшення кількості клієнтської бази.

Банківські ризики є вкрай багатоаспектними, вони проявляються в різних сферах банківської діяльності та присутні у різних типах банківських операцій. Різні джерела банківських ризиків спричиняють відповідно різні види ризиків банківської діяльності. Для структурування усіх можливих видів ризиків створюється їх класифікація. Класифікація може різнитися, тому для більш об'єктивної оцінки розглянемо декілька з них.

Перш за все, необхідно згадати Міжнародну конвергенцію вимірювання капіталу та стандарт капіталу (надалі – Базель II) – угоду про капітал, що базується на розширеному тлумаченні природи різних категорій ризиків. Згідно з Базель II, можна виділити три основні види банківських ризиків [62]:

- **Кредитний ризик** – можливість невиконання позичальником або протилежною стороною взятих на себе зобов'язань відповідно до встановлених умов. Метою управління кредитними ризиками є максимізація рівня регулювання ризиків повернення кредитів шляхом підтримання кредитних ризиків потенційних втрат у прийнятих межах. Банки потребують управління як кредитними ризиками, що притаманні портфелю активів в цілому, та і ризиками, що притаманні окремим кредитам або операціям. Кредитний ризик – ризик несплати позичальником суми основного боргу за кредитом і відсотків за ним.
- **Операційний ризик** – ризик, який виникає під впливом людського, технічного або технологічного чинника. Операційний ризик часто

асоціюється з ризиком невиконання, затримки платежів, а також з порушенням умов угод і невиконанням зобов'язань.

- **Ринковий ризик** – це можливість втрат, пов'язаних із несприятливим розвитком фінансових ринків, має макроекономічну природу, тобто джерелами ринкових ризиків є макроекономічні показники фінансової системи: індекси ринків, рівні відсоткових ставок.

На сьогоднішній день розвинуті країни вже здійснили поступовий перехід на наступну редакцію угоди, Базель III [63], а Україна наразі досить активно проводить відповідні реформи банківського сектору, необхідні для такого переходу. В новій версії угоди приділено увагу ще одному виду ризику – ризику ліквідності, проте чіткого визначення в документі не надано, а сам ризик в основному розглядається в контексті нових показників ліквідності LCR та NSFR

Розглянуті види ризику можна співставити з видами активних банківських операцій, як зображено на Рис. 1.7.

Як показано на Рис. 1.7., якщо розглядати банківські ризики в аспекті їх активних операцій, можна виділити операційний ризик (здійснення розрахунково-комісійних операцій для отримання комісійного доходу), кредитний ризик (здійснення кредитно-інвестиційних операцій для отримання процентного доходу), ринковий ризик (здійснення торгових операцій для отримання торгового доходу).

Відповідно до нормативних документів Національного банку України [40], існує наступна класифікація:

1. Кредитний ризик – це імовірність виникнення збитків або додаткових витрат або недоотримання запланованих доходів унаслідок невиконання боржником/контрагентом узятих на себе зобов'язань відповідно до умов договору [40].

Більш детально кредитний ризик диверсифікувався у Постанові 361 [64], що наразі втратила чинність і була замінена Постановою 64. Однак вважаємо за доцільне принаймні згадати деякі елементи кредитного ризику, що не були включені в новий нормативний документ: індивідуальний та портфельний кредитні ризики.

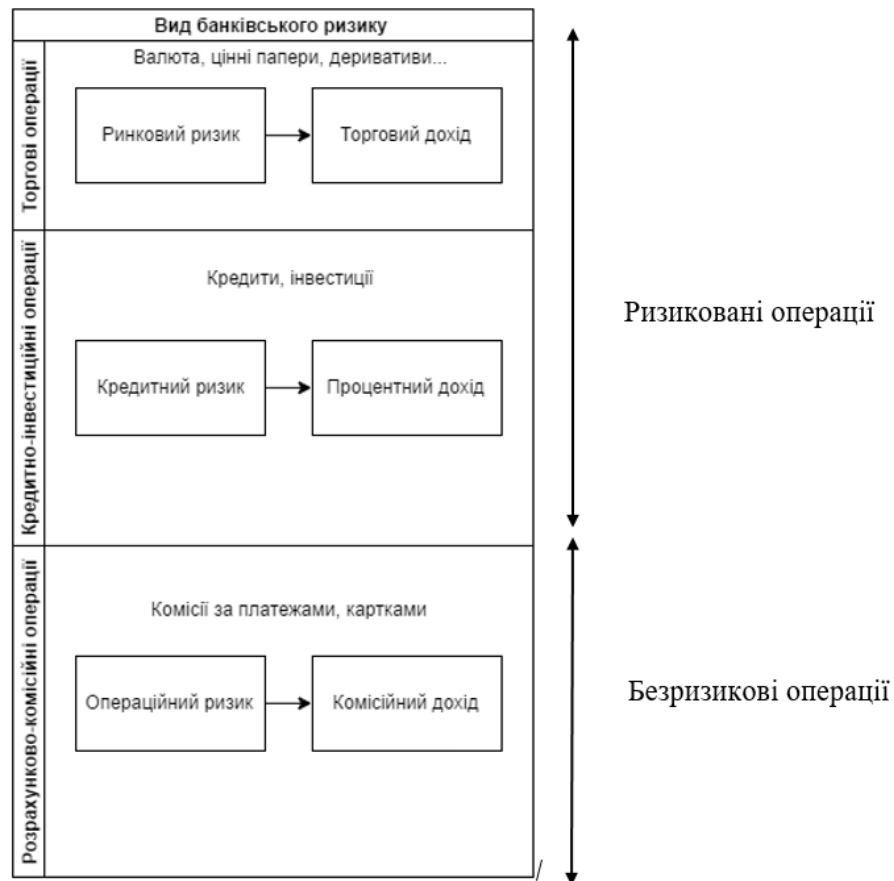


Рис 1.7. Класифікація за активними банківськими операціями

Джерело: Авторські розробки

В рамках оцінювання **індивідуального кредитного ризику** проводиться визначення кредитоспроможності даного суб'єкта, іншими словами, визначається його здатність виконати усі прийняті зобов'язання у встановлений термін та у повному обсязі. Для **портфельного кредитного ризику** визначається сукупна заборгованість перед банком за операціями, яким притаманний кредитний ризик (кредитний портфель, портфель цінних паперів, портфель дебіторської заборгованості тощо), передбачається оцінка концентрації та диверсифікації активів банку.

Кредитний ризик також включає в себе інші під-види ризиків. Якщо розглядати рівень міжнародного кредитування, то тут існує **ризик країни** – усі види ризиків, що виникають в економічному, політичному і соціальному середовищі країни реєстрації та ведення бізнесу боржника-нерезидента та можуть мати потенційний вплив на його спроможність обслуговувати борги; та **трансфертний ризик** – імовірність виникнення збитків або додаткових втрат або

недоотримання запланованих доходів унаслідок того, що у боржника немає можливості отримувати та/або перераховувати іноземну валюту для обслуговування боргу за кредитом. Також виділяються **ризик контрагентів**, пов'язаний з невиконанням другої сторони фінансового контракту своїх зобов'язань, та **ризик інвестицій в дочірні компанії**, як імовірність отримати відмінний від очікуваного результат від вливання фінансових ресурсів у підпорядковані материнській компанії структури.

В свою чергу, ризик контрагента поділяється на **перед-розрахунковий ризик** - імовірність виникнення збитків або додаткових втрат або недоотримання запланованих доходів унаслідок дефолту контрагента за договором до початку виконання своїх зобов'язань будь-якою зі сторін договору; та **ризик розрахунків** - імовірність виникнення збитків або додаткових втрат або недоотримання запланованих доходів унаслідок невиконання контрагентом своїх зобов'язань після того, як банк виконав свою частину зобов'язань.

2. Ризик ліквідності – імовірність виникнення збитків або додаткових втрат або недоотримання запланованих доходів унаслідок неспроможності банку забезпечувати фінансування зростання активів та/або виконання своїх зобов'язань у належні строки [40].

Також іноді виникає ситуація, коли банк неспроможний швидко закрити розриви власних позицій за існуючими ринковими ставками без певних втрат, і виникає ризик для капіталу та надходжень. Такий ризик називається **ризиком ліквідності ринку**.

3. Процентний ризик банківської книги (раніше – ризик зміни процентної ставки) – імовірність виникнення збитків або додаткових втрат або недоотримання запланованих доходів унаслідок впливу несприятливих змін процентних ставок на банківську книгу. Процентний ризик банківської книги впливає на економічну вартість капіталу банку та чистий процентний дохід банку [40].

Виділяють декілька під-видів даного ризику:

- **Ризик розривів**, виникає через різницю в строках погашення (для інструментів з фіксованою процентною ставкою) або зміни величини

індексу процентної ставки (для інструментів з плаваючою процентною ставкою) активів, зобов'язань та позабалансових позицій в банківській книзі.

- **Базисний ризик**, виникає через недостатність тісного зв'язку між коригуванням ставок, отриманих та сплачених за різними інструментами, всі інші характеристики яких щодо переоцінки є однаковими.
- **Опціонний ризик**, виникає через проведення банком операцій з опціонами (**автоматичний ризик опціонності**) або наявності вбудованих опціонів у стандартних продуктах банку (**поведінковий ризик опціонності**).

4. Ринковий ризик - імовірність виникнення збитків або додаткових втрат або недоотримання запланованих доходів унаслідок несприятливої зміни курсів іноземних валют, процентних ставок, вартості фінансових інструментів [40].

Ринкові ризики включають в себе наступні під-види ризиків:

- **Ризик дефолту**, що виникає через невиконання емітентом боргового цінного папера або іншого боргового інструменту, що міститься в торговій книзі банку, свої контрактних зобов'язань.
- **Процентний ризик торгової книги**, що виникає через несприятливі зміни ринкових процентних ставок, що впливають на вартість боргового цінного паперу або іншого інструменту з фіксованим прибутком, вартість похідного фінансового інструменту (деривативу), базовою змінною для якого є ринкова процентна ставка, що міститься в торговій книзі банку.
- **Ризик кредитного спреду**, що виникає через розширення кредитного спреду між дохідністю до погашення цінного папера або іншого фінансового інструменту з фіксованим прибутком, що міститься в торговій книзі банку, та без-ризиковою дохідністю до погашення з аналогічною дюрацією, яка впливає на ринкову вартість таких інструментів. Розширення кредитного спреду може бути пов'язане зі зниженням кредитної якості емітента або із загальними змінами ринкової кон'юнктури.

- **Фондовий ризик**, що виникає через несприятливі зміни ринкової вартості акцій та інших цінних паперів із нефіксованим прибутком, що містяться в торговій книзі банку.
- **Валютний ризик**, що виникає через несприятливі коливання курсів іноземних валют, що впливають на активи, зобов'язання та позабалансові позиції, що містяться в торговій та банківській книгах банку.
- **Товарний ризик**, що виникає через несприятливі зміни ринкової вартості товарів, включаючи дорогоцінні метали, що містяться в торговій та банківських книгах банку.
- **Ризик волатильності**, що виникає через несприятливі зміни волатильності ринкових цін, процентних ставок, ринкових індексів і валютних курсів, що призводить до зменшення вартості опціонів, включаючи вбудовані опціони в інші фінансові інструменти.

В старих нормативних документах валютний ризик також поділявся на підвиди [64]:

- **Ризик перерахування або трансляційний ризик.** Даний ризик виникає в результаті зміни величини еквівалента валютної позиції у звітності, що в свою чергу спричинене зміною обмінних курсів, що використовуються для перерахування залишків іноземної валюти у національну.
- **Трансакційний ризик.** Даний ризик виникає у випадку зміни валютного курсу, що може спричинити зниження реальної вартості відкритих валютних позицій.
- **Економічний валютний ризик.** Даний ризик виникає у випадку, коли має місце зміна конкурентоздатності банківської установи (або її підрозділів, структур) на зовнішньому ринку, що виникають в результаті певних помітних змін обмінних курсів.

5. Операційний ризик (раніше – операційно-технологічний ризик) - імовірність виникнення збитків або додаткових втрат або недоотримання запланованих доходів унаслідок недоліків або помилок в організації внутрішніх процесів, навмисних або ненавмисних дій працівників банку або інших осіб, збоїв

у роботі інформаційних систем банку або внаслідок впливу зовнішніх факторів [40].

В результаті реалізації даного ризику банк може понести фінансові втрати, через шахрайство, помилку персоналу чи технічну помилку, недотримання строків виконання робіт, перевищення співробітниками банку повноважень та порушення ними встановлених норм та обмежень. Необґрунтована стратегія банку, слабка внутрішня організація та неадекватний підхід до вибору та інтеграції інформаційних технологій також сприяють виникненню операційно-технічного ризику.

З іншого боку, операційний ризик може бути реалізований в результаті непередбачених подій, таких як стихійні лиха.

До операційного ризику включають **інформаційний ризик** - імовірність виникнення збитків/санкцій або додаткових втрат, недоотримання запланованих доходів внаслідок виникнення внутрішніх і зовнішніх подій щодо інформаційних систем банку та інших інформаційних ресурсів, що використовуються для досягнення цілей банку, недостатності внутрішнього контролю чи неадекватних або помилкових внутрішніх процесів банку у сфері інформаційно-комунікаційних технологій [40].

6. Комплаєнс-ризик - імовірність виникнення збитків/санкцій або додаткових втрат, недоотримання запланованих доходів або втрати репутації внаслідок невиконання банком вимог законодавства, нормативно-правових актів, ринкових стандартів, правил добросовісної конкуренції, правил корпоративної етики, виникнення конфлікту інтересів, а також внутрішньо-банківських документів банку [40].

7. Ризик репутації - імовірність виникнення збитків або додаткових втрат, недоотримання запланованих доходів внаслідок несприятливого сприйняття іміджу банку клієнтами контрагентами, акціонерами, наглядовими та контролюючими органами [40]. Ризик репутації може привести банк (або його керівників) до фінансових втрат або зменшення клієнтської бази. Цей ризик має

місце на всіх рівнях організації, і тому банки повинні ставитися до своїх відносин із клієнтами та суспільством належним чином.

8. Юридичний ризик - імовірність виникнення збитків або додаткових втрат, недоотримання запланованих доходів внаслідок невиконання сторонами умов договорів у зв'язку з їх невідповідністю вимогам законодавства [40].

Юридичний ризик виникає в тому випадку, коли банком встановлюються відносини з великою кількістю сторін. Це можуть бути посередники, клієнти, інші уповноважені особи та контрагенти. В результаті реалізації юридичного ризику банк може понести фінансові втрати, через сплату штрафів, відшкодування збитків, накладання санкцій. Можливе також недоотримання банком частини прибутку, через втрату частини клієнтів в результаті погіршення репутації.

Так, відповідно до даних спецпроекту НБУ [2], тільки з 2014 року було закрито 14 банків через незадовільну ділову репутацію власників та 28 банки припинили свою діяльність за результатами фінмоніторингу, за підозри шахрайства чи виявлення фінансових невідповідностей. Разом ці банки складають майже половину (48.8%) від усіх банків, що були закриті за період з 2014 року.

9. Стратегічний ризик - імовірність виникнення збитків або додаткових втрат, недоотримання запланованих доходів внаслідок неправильних управлінських рішень та неадекватного реагування на зміни в бізнес-середовищі [40].

Найбільш широку класифікацію, подано у посібнику за авторством Вовк В.Я. [43]. Розглянемо ті види ризику, що не були згадані вище. Це:

- **Ризик на рівні держав (міждержавний ризики)** – можливість понести збитки через неплатоспроможність, неможливість чи небажання іншої країни виконати власні зобов'язання перед позичальником (інвестором) з певних причин, відмінних від тих що виникають коли мова йде про звичайні банківські ризики. Ризик на рівні країни особливо значущим є для банків, що ведуть активну зовнішньоекономічну діяльність, встановлюють зовнішні відносини, клієнти яких ведуть активну експортну політику [43].

- **Політичний ризик** – можливість зазнати збитки чи втратити частину прибутку в результаті певних змін в політиці держави. Політичний ризик формується під впливом таких факторів як політика уряду країни та пріоритетні напрямки діяльності даного уряду. В країнах де законодавство є непостійним, часто змінюється, відсутні традиції та культура підприємництва – врахування політичного ризику є досить важливим [43].
- **Соціальний ризик** - зумовлений національними особливостями, релігійними розбіжностями, національними та ідеологічними конфліктами, дискримінаційними процесами [43].
- **Ризик інфляції** – можливість понести збитки чи недоотримати частину прибутку через знецінення національної валюти, банківських активів та власного капіталу банку [43].
- **Форс-мажорні ризики** – можливість понести збитки в результаті певних непередбачених подій: війни, епідемії, стихійного лиха [43].
- **Технологічний ризик** – потенційна чи наявна можливість понести збитки в результаті збою в роботі технічних засобів, програмного забезпечення, високотехнологічного обладнання, технології і т.д. [43].
- **Документарний ризик** – потенційна чи наявна можливість понести збитки в результаті виникнення помилки в документі, угоді під час її укладання, що як наслідок, може призвести до порушення певних пунктів даної угоди чи недотримання зобов'язань. Збитки банку від реалізації даного ризику можуть бути оцінені як розмір фінансових ресурсів витрачених на усунення документарних помилок, на участь у судових процесах, а також як сума недоотриманого прибутку. Для того, щоб зменшити вплив даного ризику, доцільним є зміцнити систему менеджменту, аудиту, провести автоматизацію процесу формування документів та документообігу, підвищити кваліфікацію персоналу фронт - і бек-офісів та ін. [43].
- **Ризик впровадження нових продуктів** – потенційна чи наявна можливість понести збитки в результаті нездатності банком досягти

цільових показників окупності банківських продуктів, послуг, технологій що виводяться на ринок. Ризик впровадження нових продуктів можна включити до складу маркетингового ризику, він виникає у випадках, коли банк розробляє та впроваджує новий продукт з метою розширення клієнтської бази, виходом на нові сегменти ринку тощо [43].

- **Ризик зловживань** – потенційна чи наявна можливість понести збитки в результаті шахрайства, недобросовісності співробітників або клієнтів банку, розкриття ними внутрішньої конфіденційної інформації, порушення банківської таємниці, використанням приватної інформації керівництвом у власних цілях тощо [43].

Ознайомлюючись з Базель III можна зустріти специфічний вид ризику, що називається **ризик втрати напрямку** (*wrong-way risk, WWR*) [63]. Це під-вид ризику контрагентів, що виникає, у випадку коли зростання вразливості до дій / залежності від контрагента корелює зі зростанням імовірності дефолту цього контрагента. Ризик втрати напрямку може бути спричинений як певними специфічними особливостями самого контрагента, так і макроекономічними факторами.

З усіх розглянутих вище ризиків ми пропонуємо об'єднувати політичний ризик, ризик інфляції та соціальний ризик у одну групу ризиків – ризиків на рівні держави.

Детально проаналізувавши вищенаведені класифікації, пропонуємо власну систематизацію видів банківських ризиків на Рис. 1.8.

На Рис. 1.8. цільними стрілками з'єднанні елементи, що включають під-елементи. Тобто, до складу кредитного ризику ми відносимо ризик країни, трансфертний ризик, ризик інвестицій в дочірні компанії, ризик контрагентів. Останній поділяється на перед-розрахунковий ризик та ризик розрахунків.

До складу операційного ризику можна віднести технологічний, форс-мажорний, документарний ризики, інформаційний ризик та ризик зловживань, ризик впровадження нової продукції, так як ці ризики пов'язані з операційною та інформаційно-технологічною діяльністю банківської установи. До ризику на

державному рівні ми відносимо політичний, соціальний ризики та ризик інфляції (зміни в законодавстві, вплив соціальної нерівності, девальвація національної валюти можуть негативно вплинути на діяльність банківської установи).

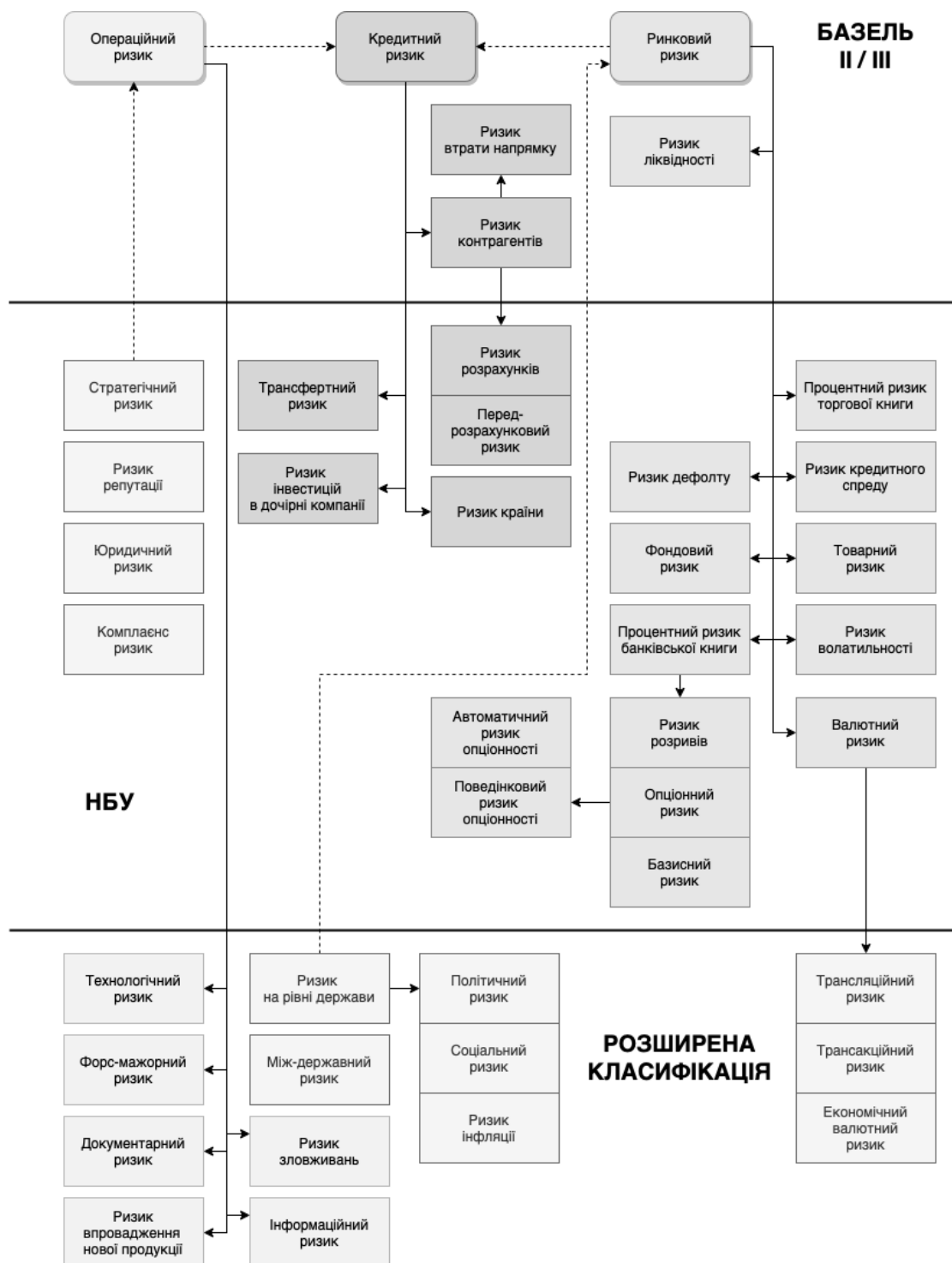


Рис. 1.8. Систематизація видів банківських ризиків

Джерело: Авторські розробки на основі [40, 43, 62, 63, 64]

До ринкового ризику відносимо процентний ризик торгової книги, ризик дефолту, ризик кредитного спреду, фондовий та торговий ризики, ризик

волатильності, валютний ризик, процентний ризик банківської книги та ризик ліквідності. Всі ці види ризику було вирішено віднести до ринкових ризиків, так як вони всі залежать від ринкових коливань. В свою чергу, валютний ризик поділяється на трансляційний, трансакційний та економічний валютний ризики. Процентний ризик банківської книги включає ризик розривів, базисний ризики та опціонний ризик. Останній також поділяється на автоматичний та поведінковий опціонні ризики.

Окремо виділяються ризик на рівні держави, міждержавний ризик, стратегічний ризик, юридичний ризик, ризик репутації та комплаєнс ризик.

Ми включили ризик ліквідності до області «Базель II / III», так як в новій редакції відносно багато уваги приділяється цьому виду ризику, хоча і важко сказати, що його виділяють як один з ключових ризиків, на рівні з кредитним, ринковим та операційним. Ризик контрагентів також включено до цієї області, так як даний ризик зустрічається в новому Базелі в секції огляду кредитного ризику.

Пунктирними стрілками позначений вплив одного ризику на інший (структурні зв'язки). З операційних ризиків більше всього на кредитний впливає ризик зловживань, зумовлений можливістю махінацій та спекуляцій робітниками банку, або незаконними діями клієнтів. На кредитний ризик також впливають коливання валют, процентної ставки (ринковий ризик). В свою чергу ризик на державному рівні сильно впливає на ринковий ризик (в результаті політичних та соціальних змін, інфляційних процесів – може змінюватися кон'юнктура ринку та ринкова рівновага, що призведе до коливання процентної ставки, вартості товарів на ринку, знецінення національної валюти).

Також, відповідно до Рис. 1.8. видно, що класифікація НБУ включає в себе ризики, виділені в Базель II, а розширена класифікація включає в себе класифікацію Базель II та НБУ.

Ми вважаємо, що авторська класифікація ризиків, представлена в даному розділі та візуально зображена на Рис. 1.8. має певну науково-теоретичну новизну, так як переважна більшість існуючих на момент проведення дослідження класифікацій має на меті висвітлити найбільш істотні для конкретного дослідження

види ризиків, деякі ж класифікації вже втратили актуальність в результаті змін в українському нормативному полі та стрімкого розвитку та впровадження Базеля III. Новизна наведеної нами класифікації на момент дослідження, по-перше, полягає саме в багатоаспектному, всесторонньому огляді банківських ризиків на основі сучасних, актуальних нормативних та дослідницьких джерел, що може претендувати на повноту. По-друге, нами було визначено ієрархію ризиків, взаємовплив та взаємозв'язок ризиків на основі як низки наукових джерел, так і загальноприйнятих нормативних документів: законів України та Базельських рекомендацій.

Відповідно до наведеної вище (Рис. 1.8) систематизації, проведено класифікацію ризиків за значущістю, відображену на Рис. 1.9.



Рис. 1.9. Класифікація за значущістю ризику

Джерело: Авторська розробка

На рис. 1.9. продемонстровано що значущими також є і ринкові ризики, що виникають під час торгових операцій банку. Операційні ризики реалізуються відносно рідше, проте збитки від них зазвичай можуть бути досить значними. Таким чином, не виникає сумнівів, що саме дані три види ризику потребують більш детального розгляду. Саме тому в наступному підрозділі пропонується сконцентрувати увагу на аналізі особливостей управління та моделювання саме цими видами банківських ризиків

1.3. Моделі та методи аналізу фінансового стану комерційних банків та ризиків банківської діяльності

Одним з найбільш важливих висновків з проведеного в попередніх підрозділах аналізу є визначення категорії ризику як ключового елементу дослідження. Ризик суттєво впливає на забезпечення комерційними банками стійкого фінансового стану через досягнення банківською установою ефективності у своїй діяльності. Забігаючи наперед, відзначимо, що в процесі аналізу фінансового стану банківських установ ми будемо розглядати ефективність діяльності банків та їх стійкість або вразливість до сукупного ризику банківської діяльності через призму набору різнопланових показників, які будуть включені в модель. Однак доцільно також розглянути інші підходи: як окремі більш класичні методи та моделі, що використовуються для управління ефективністю діяльності та оцінювання різних видів ризику, так і більш інноваційні методи, що можуть бути застосовані для комплексного аналізу фінансового стану комерційних банків.

Оцінка банківського ризику – один з найважливіших етапів ризик-менеджменту, на якому проводиться як вимірювання ймовірності настання певного банківського ризику, так і моделювання/прогнозування можливих збитків, що понесе за собою реалізація даного ризику. Варто відзначити, що моделей та методів оцінки ризиків банківської діяльності існує багато, і кожен вид ризику оцінюється певними специфічними методами. Для деяких видів ризику не існує стандартизованого підходу, і банк визначає самостійно, яким чином проводити його вимірювання. Зважаючи на багатоманітність моделей та методів оцінки ризиків, буде розглянуто оцінку тих ризиків, що були класифіковані у підрозділі 1.2.2, як найбільш значущі: кредитний, операційний ризики, ринковий ризик та декілька основних його під-видів.

Розпочнемо з **ринкових ризиків**. Згідно нашої класифікації, вони включають в себе усі ризики, що виникають в результаті коливань та нестабільності на ринку. Ми розглянемо ринковий ризик в цілому та деякі його під-види, що є найбільш

поширеними: це валютний ризик, процентний ризик (або процентний ризик банківської книги) та ризик ліквідності.

Валютний ризик. Розглянемо підхід, описаний у Базелі II [62] і впроваджений НБУ [65]. Відповідно до нього виділяється два етапи оцінювання валютного ризику.

На першому етапі визначаються позиції банку в окремій валюті та золоті (що класифікується як валюта). Для цього банк визначає всі свої спот-позиції в даній валюті, форвард-позиції, гарантії, захеджовані ненараховані проценти та дельта-еквівалента портфеля валютних опціонів. Від отриманої величини віднімається структурна валютна позиція в цій валюті (якщо така ведеться банком). На першому етапі чиста довга чи коротка позиція банку перераховується в еквівалент в національній валюті відповідно до актуального курсу.

На другому етапі необхідно врахувати ефекти диверсифікації та концентрації валютного ризику. Для цього використовується так званий «метод короткої руки». За цим методом, банк групує еквіваленти позицій окремих валют залежно від знаку у дві групи: довгі (+) та короткі (-) і визначає суму по кожній групі. З двох отриманих сум береться більша за модулем.

До значення отриманого на першому кроці додається 8% абсолютного значення більшої з суми довгих або коротких позицій та абсолютної чистої позиції в золоті. Отримане значення є показником валютного ризику. Розрахунок здійснюється окремо для кожної валюти та золота, для інструментів торгової та банківської книг.

В роботі Лавренюка В. В. [66] детально розглядаються теоретико-методологічні та практичні питання управління валютним ризиком банку: проводиться типізація, систематизація методів управління валютним ризиком. Моделювання валютного ризику проводиться з застосуванням VaR та прогнозних моделей

Процентний ризик. Для українських банків одним з основних інструментів оцінки процентного ризику є GAP - аналіз, що заснований на розбіжності величини

чутливих до зміни процентної ставки активів та пасивів (аналіз проводиться за строковими періодами). Розрахунок проводиться за наступною формулою:

$$GAP_t = \sum A_t - \sum L_t \quad (1.1)$$

де GAP_t – періодичний розрив,

$\sum A_t$ – сумарний обсяг процентних активів, що переоцінюються в періоді t ,

$\sum L_t$ – сумарний обсяг процентних пасивів, що переоцінюються в періоді t .

Від’ємний GAP свідчить про перевищення чутливих до процентної ставки пасивів над активами. Якщо процентна ставка збільшується, чистий дохід банку у вигляді відсотків зменшується. Додатній GAP свідчить про обернену ситуацію. Зниження короткострокових процентних ставок дає протилежний ефект. Дохід у вигляді відсотків зменшується сильніше, ніж витрати по виплаті відсотків, так що спред банку та чистий дохід у вигляді відсотків знижується. Нульовий GAP вказує на те, що чутливі до зміни процентної ставки активи й пасиви рівні між собою і однакові зміни процентної ставки не впливають на дохід банку [65].

Також виділяють кумулятивний GAP [65], що враховує ефект накопичення, та досить широко вживаний метод дюрації, що представляє з себе середньозважений час до погашення фінансового інструменту, а очікувана модифікована дюрація враховує ризик несвоєчасного погашення активу або пасиву банку. За методом дюрації процентний ризик визначається як різниця між середньозваженими модифікованими дюраціями активів та пасивів: чим нижчою є ця різниця, тим меншим вважається процентний ризик [65]. Даний підхід запропонований НБУ як альтернатива рекомендаціям Базеля щодо вимірювання процентних ризиків.

Наразі як у нормативних документах НБУ, так і у базельських угодах також рекомендується застосування таких методів як Довгостроковий метод оцінки процентного ризику на основі визначення економічної вартості капіталу (Economic Value of Equity - EVE) та Метод чутливості чистого процентного доходу (Net Interest Income – NII), представлені в роботах Грея Д. [67], А. Ресті та А. Сіроні [68] відповідно. Ці, та ряд інших методів, також розглядаються в роботах Михайличенко М. С. [68] та Бучко І. Є. [70].

Інший підхід, запропонований у Базель II розглянуто у статті Бобиля В. В. [71], відповідно до даного підходу процентний ризик може розраховуватись за наступною формулою:

$$\text{ПР} = \text{СПР} + \text{ЗПР} \quad (1.2)$$

Де СПР – специфічний процентний ризик (ризик несприятливої зміни поточної вартості фінансового інструменту через фактори, які стосуються окремого емітента),

ЗПР – загальний процентний ризик (ризик несприятливої зміни поточної вартості фінансового інструменту, пов'язаного з ринковими коливаннями процентних ставок).

Ризик ліквідності. Раніше даному виду ризиків не приділялось достатньо уваги, проте з впровадженням третьої версії Базеля, ризик ліквідності розглядається як один з основних.

Для аналізу ліквідності, згідно НБУ [40], також пропонується застосовувати GAP-аналіз за схожою методикою, що була розглянута вище для процентних ризиків, однак метою є визначення неузгодженості між вимогами та зобов'язаннями банку за контрактними або очікуваними строками погашення. GAP-аналіз для ризику ліквідності проводиться контрактний – виключно за контрактними строками, за основними інструментами та вбудованими опціонами; та прогнозний – включаючи припущення щодо нових можливих операцій.

Для моніторингу ліквідності банківської установи на сьогоднішній день використовуються нормативи НБУ [65], такі як:

- Норматив короткострокової ліквідності Н6 - визначається як співвідношення активів до зобов'язань з кінцевим строком погашення до одного року ($\text{не} < 60\%$).
- Коефіцієнт покриття ліквідністю (Liquidity coverage ratio, LCR) – норматив, що визначає мінімально необхідний рівень ліквідності, що необхідний для покриття чистого очікуваного відпливу грошових коштів протягом 30 днів з урахуванням стрес-сценарію ($= 100\%$).

- Коефіцієнт чистого стабільного фінансування (Net stable financing ratio, NSFR) – норматив, що визначає мінімально необхідний рівень стабільного фінансування, для забезпечення діяльності банку протягом одного року (= 100%).

Останні два нормативи впроваджені у відповідності до стандарту Базель III.

Методам та інструментам управління ризиком ліквідності присвячена стаття Коваленко В. В. [72], в якій розглядаються деякі інші підходи до оцінки та управління ризиком ліквідності. Цікавою також є робота Рую Цай та Мао Чжан [73], а якій проведено ґрунтовне дослідження впливу кредитного ризику на ризик ліквідності

Для оцінки **ринкового ризику в цілому** часто застосовується підхід Value-at-Risk, VaR. Він дозволяє з рівнем надійності $A\%$ припустити, що втрати від реалізації ризику не будуть більшими за V одиниць за період у T днів. Величина V є невідомою і представляє собою оцінку ризику – VaR, і є функцією від часу та рівня довіри. Для розрахунку VaR застосовується наступна формула:

$$VaR = M * T \pm K_{\alpha} * \sigma * \sqrt{T} \quad (1.3)$$

де M – середня відносна зміна ринкових процентних ставок,

T – горизонт часу,

K_{α} – коригуючий коефіцієнт (визначається залежно від значення α),

σ – середньоквадратичне відхилення відносних змін ринкових процентних ставок.

Серед альтернативних підходів можна виділити авто-регресійні моделі гетероскедастичності або ARCH моделі (Auto Regressive Conditional Heteroscedasticity), представлені Робертом Енегелем у своїй праці [74], що є перспективними до застосування для моделювання ринкових ризиків в цілому для підприємств, і в тому числі банківських установ. Такі моделі можуть бути застосовані для прогнозування волатильності певних активів, які знаходяться на балансі банку, і потенційна можливість зміни реальної вартості яких призводить до виникнення ринкового ризику. Також широко вживаними є модифікації, такі як GARCH-моделі (Generalized ARCH) та EGARCH-моделі (Exponential GARCH),

запропоновані Тімом Болерслемом [75] та Д. Нельсоном [76] відповідно, що підвищують ефективність оригінальної моделі та усувають деякі з її недоліків.

Перейдемо до розгляду **операційних ризиків** банківських установ. Згідно статті Кузьмака О.М. [77] «до основних причин операційних ризиків відносяться збої в комп'ютерній техніці (56% банків), помилки банківських працівників (36%), шахрайство (8%)». Серед основних причин також виділяють: несанкціоновані торговельні операції, помилки при застосуванні деривативів. Система регулювання/оцінки операційних ризиків на сучасному етапі зазвичай присутня лише у великих банках і впровадження таких систем є складною задачею. Основною проблемою оцінки операційних ризиків є відсутність єдиної методології управління даними ризиками. Одним з методів є статистичний, заснований на зборі інформації щодо збоїв та інших випадків реалізації операційного ризику. Але такий підхід є спрощеним і обмеженим.

Відповідно до Базель II, для банків доцільним є інтеграція повноцінної системи управління операційними ризиками, що має включати визначення політики у сфері операційних ризиків, формування рівнів відповідальності (ролевих позицій), класифікацію операційних ризиків, аналіз операційних ризиків за напрямками діяльності та моніторинг стану операційних ризиків [77].

Для того, щоб запобігти непередбаченим втратам від реалізації операційного ризику, банком має виділятися капітал під операційний ризик. Базельській комітет виділяв три підходи до оцінки такого капіталу: підхід на основі Основного Індикатора, стандартизований підхід та підхід вдосконаленого вимірювання. Більш детально всі ці підходи описані та розглянуті у статті Бобиля В. зі співавторами [78].

Безель III пропонує замінити три розглянуті підходи на один стандартизований підхід, за яким операційні ризики визначаються як добуток двох компонентів: Бізнес індикатор (Business Indicator Component, BIC) та Мультиплікатор внутрішніх втрат (Internal Loss Multiplier, ILM). Бізнес індикатор представляє з себе агреговане значення добутків доходів банку від трьох компонентів: компоненту процентів, дивідендів та доходів від оренди, компоненту

послуг та фінансового компоненту – на спеціальні маржинальні коефіцієнти, які встановлюються відповідно до рівня прибутку. Мультиплікатор внутрішніх втрат є функцією від Бізнес індикатору та Компоненту втрат (Loss Component, LC), що в свою чергу рівний 15-кратним середнім історичним втратам банку за останні 10 років. Регулятор може встановлювати значення MBV на рівні 1 для банків в своїй юрисдикції на свій розсуд. Це буде означати, що вимоги до капіталу будуть визначатись виключно компонентом БІ.

Відповідно до Постанови 64 НБУ [40], вітчизняні банки для виявлення та вимірювання операційного ризику розраховують та відстежують ключові показники ризику (Key Risk Indicators, KRI), проводять самооцінку ризику: аналіз бізнес-процесів та ризиків з ними пов'язаних, проводять сценарний аналіз шляхом формування судження щодо визначення можливих подій операційного ризику та їх кількісної оцінки на підставі власної експертної думки, а також експертної думки представників підрозділів першої лінії захисту.

Питання попередження операційних ризиків та деякі конкретні шляхи запобігання їх реалізації розглянуті у статті Євтушенко Г. В. [79]. Дослідження підходів до моделювання операційного ризику банківських установ, а саме через розподіл діяльності банку на бізнес-лінії та застосування актуарних моделей представлено в роботі Краснови І. В. [80].

Останнім видом ризику, який ми розглянемо, і який можна визначити як найважливіший та найбільш значущий, є **кредитний ризик**. Приймаючи саме цей ризик, банки отримують основну частку свого доходу. Методів та підходів оцінки кредитного ризику існує досить багато, тому розглянемо лише деякі з них.

В першу чергу, варто відзначити один з найпопулярніших методів – VaR, який ми вже коротко розглядали вище, і який застосовується для оцінки ризику кредитного портфелю, проте специфіка розрахунку дещо відрізняється. Поняття кредитного ризику на основі методології VaR можна сформулювати як максимально можливі збитки по кредитному портфелю для заданого ступеню довіри. Максимальні збитки при цьому розподіляються на очікувані (Expected Loss, EL_p)

та неочікувані (Unexpected Loss, UL_p) збитки по портфелю. Графічно це зображено на Рис. 1.10.

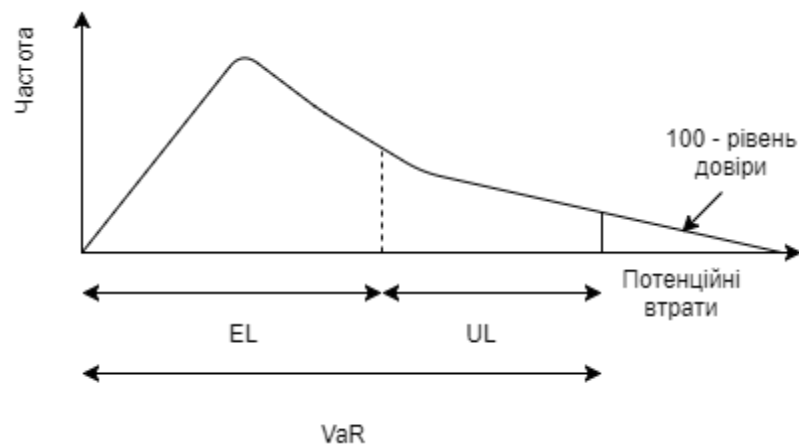


Рис. 1.10. Розподіл витрат за кредитним портфелем

Джерело: розроблено автором на основі [81]

Очікувані втрати за кредитним портфелем розраховуються за наступною формулою:

$$EL_p = \sum_{i=1}^N PD_i * CE_i * (1 - RR_i) \quad (1.4)$$

де PD_i – імовірність настання дефолту позичальника,

CE_i – вартість активів, що знаходяться під ризиком в момент настання дефолту,

RR_i – рівень відшкодування збитків, тобто доля заборгованості, що у випадку дефолту позичальника можливо буде повернути шляхом реалізації застави, гарантій і т. д.

Важливою задачею при визначенні рівня очікуваних втрат є саме оцінювання імовірності дефолту позичальника. Для цього застосовуються моделі дискримінантного аналізу, логіт та пробіт моделі, нейронні мережі, рейтингові системи і т. д.

Неочікувані збитки визначають рівень кредитного ризику і розраховуються як $UL = VaR^a - EL$.

Одним з методів для розрахунку самого показника VaR є імітаційний метод моделювання Монте-Карло, суть якого полягає в багатократній реалізації певного випадкового процесу, при заданих за умовою задачі імовірнісних характеристиках.

Більш детально його використання з метою моделювання кредитного ризику описане в статті Фартушного І. Д. [81].

Розглядаючи методи та підходи оцінки кредитних ризиків, варто згадати **IRB-підхід**. Це підхід до визначення кредитних ризиків банків для оцінки достатності регулятивного капіталу, заснований на використанні внутрішніх рейтингів позичальників, тобто рейтингів, що встановлюються самими банками. Підхід запропонований в Базелі II як альтернатива стандартизованому підходу і адаптований НБУ для застосовування в Україні в рамках Постанови № 351 Положення «Про визначення банками України розміру кредитного ризику за активними банківськими операціями» від 30 червня 2016 року [82].

Теоретичні аспекти даного питання детально розглянуті в статті Ониськіва Л. М. [83], а ми розглянемо основні аналітичні аспекти нової методики розрахунку кредитного ризику. Розрахунок кредитного ризику здійснюється відповідно до рекомендацій Базельського комітету, і включає визначення трьох компонентів: ймовірності дефолту боржника (Probability of default – PD), рівень втрат у випадку дефолту (Loss given default – LGD) та вимоги під ризиком дефолту (Exposure at default – EAD).

Загальна формула для розрахунку кредитного ризику має наступний вигляд:

$$CR = PD * LGD * EAD \quad (1.5)$$

де CR – розмір кредитного ризику за певним активом,

PD – показник імовірності дефолту позичальника за певним активом,

LGD – втрати при дефолті. Розраховується наступним чином:

$$LGD = 1 - RR \quad (1.6)$$

де RR – показник, що характеризує частину боргу, що буде повернена за рахунок реалізації забезпечення. Розраховується наступним чином:

$$RR = (\sum_{i=1}^j (CV * k)_i + RC) / EAD \quad (1.7)$$

де EAD – вимоги під ризиком дефолту,

CV_i – вартість i -го виду забезпечення,

ki – коефіцієнт ліквідності i -го виду забезпечення. Визначається відповідно до Додатку 6 Положення [82],

RC – інші надходження.

За новою методикою клас боржника, наприклад, боржника-юридичної особи, визначається з використанням інтегрального показника Z , загальний порядок розрахунку якого має наступний вигляд:

$$Z = \beta_0 + \sum_{i=1}^n \beta_i x_i \quad (1.8)$$

де Z – інтегральний показник,

x_i – фінансові коефіцієнти, які розраховуються за даними з фінансової звітності боржника-юридичної особи,

β_i – коефіцієнт логістичної моделі, що розраховується з урахуванням вагомості показників,

β_0 – вільний член логістичної моделі, визначається НБУ.

Вигляд моделі для розрахунку інтегрального показника відрізняється в залежності від виду економічної діяльності, яку веде боржник юридична-особа а також від масштабу підприємства (мале, середнє або велике).

При встановлені класу боржника також вагомими є наступні фактори:

- Чи є борг простроченим (прострочення від 31 до 60 днів – клас боржника не вище 5; від 61 до 90 днів – не вище 8; від 91 дня та більше – не вище 10).
- Чи належить позичальник до групи юридичних осіб під спільним контролем пов'язаних осіб.
- Чи наявні ознаки високого кредитного ризику позичальника.
- Чи є в банку інформація про кредитну історію позичальника.

На основі визначеного класу боржника, йому привласнюється показник PD з відповідного діапазону.

Детальніше про специфіку визначення величини кредитного ризику за новою методикою описано в статті Коцовської Р. Р. [84].

Основою для визначення кредитного ризику зазвичай є оцінка кредитоспроможності позичальника. Для такої оцінки використовуються

відповідні моделі, що загалом можна поділити на класифікаційні моделі та моделі комплексного аналізу.

Класифікаційні моделі дозволяють за допомогою математичних, статистичних та експертних методів розподілити клієнтів на умовні групи. Вони включають **рейтингові** та **прогнознi** моделі оцінки кредитоспроможності. Рейтингові моделі представлені наступними видами: бальна (рейтингова) оцінка та кредитний скоринг як модифікація рейтингової оцінки. До прогнозних моделей належать: МДА – множинний дискримінантний аналіз, система визначення показників (модель Бівера), модель CART (Classification and regression trees – “класифікаційні та регресійні дерева”).

Рейтингова оцінка полягає в тому, що кожному позичальнику присвоюють відповідний рейтинг. Згідно з Законом України «Про державне регулювання ринку цінних паперів в Україні» кредитний рейтинг — це умовний вираз кредитоспроможності об'єкта рейтингування в цілому та/або його окремого боргового зобов'язання за національною шкалою кредитних рейтингів [85].

Даний підхід до оцінки кредитоспроможності є досить популярний як в іноземній, так і у вітчизняній практиці. Його було розглянуто, наприклад, в статті Притули Н. І. [86].

Скорингових моделей існує велика кількість, вони різняться між собою наборами факторів, що включаються до моделей, і які характеризують ризик, що виникає при кредитуванні. В результаті моделювання отримується гранична оцінка, що розподіляє позичальників на дві категорії, зазвичай це «надійні» та «ненадійні». Кожний боржник отримує унікальну оцінку кредитного ризику засновану на певних історичних даних. Для фізичних осіб це зазвичай кредитна історія, соціальний статус, професійна приналежність та стаж, демографічна характеристика та багато іншого. Для юридичних осіб це певні фінансові показники діяльності.

Дослідження в області моделей аплікаційного скорингу та визначення найбільш значимих характеристик для побудови таких моделей при аналізі споживчого кредитування проводилися Матвійчуком А. В. [87].

Детально розглянуто специфіку застосування кредитного скорингу у статті Піщуліна А. [88], що також визначив і основні недоліки даного підходу: це технічна складність та невідповідність наявних програмних засобів банків вимогам, виконання яких необхідне для якісного моделювання; відсутність необхідних для моделювання даних; некоректність та розрізненість даних, що зазвичай збираються з різних джерел.

Розглянемо прогностні класифікаційні моделі. До них в першу чергу відносяться моделі дискримінантного аналізу, найбільш відомими з яких є: **модель Альтмана, модель Р. Таффлера, модель Р. Чессера**. Не будемо детально зупинятись на особливостях розрахунку та застосування даних зарубіжних моделей. Варто проте відзначити, що їх адаптація до вітчизняних реалій є неможливою, що пояснюється значними відмінностями у розвитку фінансово-господарської сфери. Наприклад, в основу моделі Альтмана покладений рівень рентабельності підприємства. Для країн з розвинутою економікою, дійсно, рівень рентабельності є визначальним індикатором фінансово-господарського стану, але в українських умовах ключова роль цього показника є сумнівною.

Якщо мова йде про прогнозування банкрутства в умовах транзитивної економіки, то варто згадати двох-факторну **модель М.А. Федотової та модель Р.С. Сайфуліна та Г.Г. Кадикова**, що включає в себе п'ять факторів.

Усі розглянуті вище моделі викладено та детально описано на електронному ресурсі [89].

Серед вітчизняних моделей можна виділити наступні: **універсальна модель Терещенка та галузева модель Терещенка** (описана розробником моделі в статті [90]). Автор розробив таку модель, яка показувала б адекватні результати при застосуванні для вітчизняних підприємств. На думку Терещенка, оптимальна функція дискримінантного аналізу має включати в себе повний спектр коефіцієнтів, що описують фінансовий стан позичальника. Серед них мають бути показники, що описують прибутковість, ліквідність, оборотність коштів, структуру активів, капіталу тощо.

Окрім багатофакторних дискримінантних моделей, для визначення імовірності банкрутства позичальника застосовуються також більш прості моделі. Це так звані моделі засновані на системі визначених показників, наприклад, **система показників Бівера** [89].

Широкого застосування в банківській практиці на заході набула **модель CART (Classification and regression trees)**. Вона використовується для визначення кредитоспроможності клієнтів банку, оцінювання ризикованості надання позик та класифікації наданих кредитів. Модель є непараметричної. Усі позичальники за даною моделлю розбиваються на «гілки» відповідно до значень фінансових коефіцієнтів. Кожна з таких «гілок» також розбивається на «гілки» відповідно до значень інших коефіцієнтів і т. д. Точність класифікації складає близько 90 %.

Проте перелічені вище моделі мають певні недоліки, що властиві класифікаційній групі в цілому:

- Фінансові коефіцієнти неповно характеризують кредитоспроможність позичальника, і надавати перевагу окремій групі показників не є доцільним.
- Показники, що розраховуються за минулі періоди і використовуються як розрахункова база, можуть неадекватно описувати платоспроможність та кредитоспроможність боржника.
- Уніфікованої бази нормативних значень фінансових коефіцієнтів не існує, і вони не завжди можуть бути адаптовані до окремих видів економічної діяльності і не завжди можуть бути адекватними рівню розвитку економіки.
- Розраховані коефіцієнти є статичними і не враховують поточного стану позичальника на момент отримання кредиту. Відсутність інформації про динаміку зазначених показників діяльності підприємства підвищує ризик у процесі кредитування позичальника.

Перейдемо до розгляду **моделей комплексного аналізу**. Вони поєднують в собі кількісні та якісні характеристики позичальника. Найбільш відомими, що широко застосовуються в іноземній банківській практиці – це **правило шести «Сі»**

(проводиться оцінка позичальника за особистісними якостями, фінансовими можливостями, забезпеченням позики, капіталом, зовнішніми умовами, аналізом грошового обігу або cash flow) та CAMPARI (проводиться оцінка позичальника за характеристикою клієнта та його репутації, здатністю повернути позику, маржою за позику, цільовим призначенням виданої позики, обсягом позики, умовами погашення позики, та заставою під позику). Джерелами інформації для проведення аналізу за правилом шести за даними моделями можуть бути дані, надані позичальником; історична інформація, зібрана банком; інформація від контрагентів; інформація надана держустановами. Серед інших моделей можна виділити такі як **PARSER, PARTS, MEMO RISK** та **система 4FC**.

В науковій літературі викладки в сфері аналізу кредитоспроможності можна зустріти в статтях таких авторів як Чайковський Я. І. [91] та Смолевої Т. М. [92]. Проте детальних досліджень в даній області, особливо на тему аналізу комплексних методів, проводиться небагато.

Відзначимо, що загалом класичні моделі також широко застосовуються в зарубіжних наукових дослідженнях. Так, в роботі Міхаловіца М. [93] розглядається модель, побудована на основі множинного дискримінантного аналізу і порівнюється з логістичною регресією, в контексті аналізу банкрутства словацьких банків. Схоже дослідження проводилося в роботі Афес та Хентаті-Кафеля [94]: порівняння ефективності логістичної регресії і класичного дискримінантного аналізу для прогнозування банкрутства банків США. У викладках Лу К. [95] застосовується факторний аналіз для оцінки ефективності комерційних банків Китаю.

Моделі комплексного аналізу стану позичальника є відносно привабливими для використання вітчизняними банками, однак механічно переносити іноземний досвід в український економічний простір, без урахування специфіки функціонування українських підприємств, не є ефективним та доцільним. Крім того, необхідно враховувати ряд недоліків, що виникають при здійсненні комплексного аналізу кредитоспроможності. Це недостатнє використання

математичного апарату, суб'єктивізм експертів та брак методологічних матеріалів для проведення аналізу.

На сьогоднішній день для вирішення різноманітних типів задач, в тому числі для визначення та розрахунку кредитного ризику, використовуються нові інноваційні методи. Серед них можна виділити використання моделей на основі нечіткої логіки, моделі на основі аналізу виживання, штучні нейронні мережі. Зважаючи на складність та об'ємність математичного апарату, не будемо детально зупинитись на його огляді, а наведемо тільки короткий опис.

Моделі на основі нечіткої логіки. Такі моделі останнім часом набули популярності і є досить перспективними для проведення дослідження в банківській сфері. Математичний апарат нечіткої логіки заснований на поєднанні класичної логіки та теорії нечітких множин; він надає можливість проводити формалізацію залежностей різної складності та працювати як з кількісними так і якісними параметрами, опис залежностей між якими, в свою чергу, описується природною мовою. Висока популярність зумовлює велику кількість наукових робіт, серед яких можна назвати статті Прімерова О. К. [96] та Кожухівської О. А. [97], в яких викладено аргументи щодо доцільності застосування нечіткого інструментарію до моделювання банківського ризику, та зроблено спроби апробації моделей на реальних даних.

Аналіз виживання. Аналіз виживання – це клас статистичних моделей, що надають можливість оцінити ймовірність настання певної події. Такі моделі іноді застосовують в дослідженнях в області кредитного скорингу. Статичні скорингові моделі можуть бути не дуже надійними, саме тому в деяких дослідженнях пропонується включати часовий параметр в модель, і розглядати дані в динаміці. Більш детально дане питання розглянуто, наприклад, в статті Писанця К. К. [98].

Штучні нейронні мережі – це математична модель, що представляє собою множину об'єднаних між собою вузлів, що називаються штучними нейронами, кожне з'єднання між якими здатне передавати сигнали. Особливістю і однією з ключових переваг нейронних мереж над традиційними алгоритмами є те, що вони не програмуються апіорі, а навчаються в процесі своєї роботи, що по суті полягає

у підборі коефіцієнтів зв'язків між нейронами. В процесі навчання нейронна мережа здатна до виявлення залежності між вхідними та вихідними даними, виконувати узагальнення. Так, мережа може надати адекватні результати та прогнози навіть після навчання на неповних, частково викривлених і зашумлених даних.

Нейронні мережі застосовуються для вирішення складних задач, які потребують аналітичних обчислень подібних до тих, що проводяться людським мозком. Штучні нейронні мережі останнім часом стають все більш популярними, їх математичний апарат безперервно досліджується, доповнюється та модифікується. Досліджень на дану тему існує досить багато, серед них статті Великоіваненко Г. І. [99] та Кузнецова Н. В. [100].

Не будемо детально зупинятись на описі математичного апарату та специфічних особливостях роботи різних типів нейронних мереж, так як це виходить за рамки даного дослідження. Більш детально теоретичні та математичні аспекти такого інструменту як штучні нейронні мережі описано, наприклад, в монографії Матвійчука А. В. [101].

Розглядаючи більш сучасні методи, такі як нейронні мережі, нечітку логіку та генетичні алгоритми в контексті загального дослідження роботи банків можна виділити роботу Думітреску та співавторів [102], в якій розглядається машинне навчання на основі модифікованої логістичної регресії, що показала непогані результати у вирішенні задачі кредитного скорингу. Ансамблеві моделі показують високу ефективність у задачах аналізу банківської ефективності, наприклад у Ердала та Караханоглу [103] розглядається метод беггінгу на основі трьох моделей типу дерева рішень, що застосовуються для передбачення ключових факторів прибутковості турецьких банків.

Порівняльний аналіз класичних методів і технік машинного навчання був проведений у роботі Ле та Вів'ані [104]. Автори провели порівняння дискримінантного аналізу та логістичної регресії з нейронною мережею (Support Vector Machine та K-nearest neighbors) проаналізувавши їх ефективність у

прогнозуванні банкрутства на основі даних по 3000 американських банків і 31 коефіцієнті.

Отже, нами було досить детально та повно розглянуто існуючі моделі, підходи та методи управління найбільш поширеними та значущими ризиками у діяльності комерційних банків. Розглянуті інноваційні методи якісно виділяються на фоні класичних моделей рядом особливостей. Серед них можна назвати високу чутливість до статистичних даних, на основі яких проводиться моделювання; здатність враховувати внутрішні приховані та неочевидні зв'язки між факторами; можливість використання для вирішення задач з неповною інформацією. Зважаючи на це, та ряд інших переваг, що ми розглянемо в наступному, останньому підрозділі, нами було вирішено відійти від класичних методів та моделей і звернутись до більш сучасного інструментарію, а саме до методів машинного навчання: алгоритмів зниження розмірності та кластеризації. Перш ніж переходити до наступного розділу даної роботи, ознайомимося трохи детальніше з теоретичним підґрунтям напрямку машинного навчання та інструментарію вказаних алгоритмів.

Перш за все, розглянемо саме поняття машинного навчання. **Машинне навчання** (англ. Machine learning) – клас методів штучного інтелекту, особливістю якого є непрямий підхід до вирішення задачі через процес навчання в ході вирішення схожих складних задач. В свою чергу, **штучний інтелект** є підрозділом інформатики, що охоплює питання формалізації проблем подібних до тих, що виконує людина. В рамках реалізації методів машинного навчання часто застосовуються елементи математичної статистики, чисельних методів, математичного аналізу, методів оптимізації, теорії ймовірності, теорії графів.

Серед найбільш поширених задач, які можуть бути вирішені за допомогою машинного навчання, можна виділити:

- **Класифікація.** Задача, що полягає в розподілі об'єктів на заздалегідь визначені класи, на основі вхідних параметрів, що описують кожний з об'єктів. До даної групи задач можна віднести також задачу розпізнавання образів, що можуть бути представлені у вигляді символів, зображень,

аудіозаписів і т. д. Для вирішення даної задачі проводиться навчання мережі на основі історичних даних, які містять характеристики (параметри) об'єктів та приналежність до певного класу яких вже визначена.

- **Прогнозування.** Проведення узагальнення та визначення прихованих залежностей між вхідними та вихідними даними. На основі таких залежностей можливе формування прогнозу для певної послідовності вхідних значень. Частковим випадком задачі є регресія.
- **Кластеризація.** Суть задачі кластеризації полягає у розбитті об'єктів на класи, але на відміну від задач класифікації, класи, до яких відносяться об'єкти, заздалегідь не відомі і формуються в процесі вирішення задачі.
- **Зниження розмірності даних та візуалізація.** Задача полягає у обробці багатовимірному масиву даних для отримання масиву даних меншої розмірності. Часто проводиться зниження розмірності даних до двох або трьох вимірів з метою візуалізації багатовимірному масиву.

Серед інших задач можна виділити: встановлення щільності розподілу ймовірності за набором даних, однокласова класифікація та встановлення новизни, побудова рангових залежностей, визначення аномалій.

Машинне навчання можна поділити на 2 ключові типи в залежності від способу його організації та проведення. За **навчання з учителем** передбачається, що кожному вхідному вектору призначений відповідний вектор, що має бути отриманий на виході, і який називається цільовим. Пара таких векторів називається навчальною парою. Зазвичай для навчання використовується деяка кількість таких пар, уся сукупність яких є навчальною вибіркою. В процесі навчання реально отримані вихідні значення порівнюється з цільовим вектором і розраховується відхилення від нього, що називається помилкою. Після обчислення помилки змінюється вага векторів відповідно до обраного алгоритму, з метою зменшення відхилення отриманих вихідних векторів від цільових значень. Дана процедура повторюється ітеративно, задану кількість ітерацій, або доки помилка не буде мінімізована до заданого дослідником рівня. Загалом, до навчання з учителем

можна віднести всі алгоритми та методи машинного навчання, які призначені для вирішення задачі класифікації. За **навчання без учителя** не потребується порівняння з цільовими, еталонними векторами. Навчальна вибірка складається виключно з вхідних векторів. В процесі навчання ваги підлаштовуються таким чином, щоб вектори, що подаються на вхід, і є достатньо подібними один до одного, в результаті давали однаковий вихід. Тобто вихідні вектори мають бути узгоджені. Отже, в результаті навчання в навчальній вибірці виділяються певні статистичні характеристики, на основі яких і проводиться розподіл подібних вхідних векторів на класи. Загалом, до навчання без учителем можна віднести всі алгоритми та методи машинного навчання, які призначені для вирішення задачі кластеризації. Дослідження в даному напрямку розвивались Т. Кохоненом [105] та багатьма іншими дослідниками.

Серед інших типів, що зустрічаються в деяких розширених класифікаціях, можна виділити: навчання з підкріпленням, активне навчання, навчання з частковим залученням учителя, трансдуктивне навчання, багатозадачне навчання, багатоваріантне навчання, бустінг, байесова мережа.

В рамках нашого дослідження нами було вирішено розглядати задачу кластеризації для аналізу фінансового стану вітчизняних комерційних банків. У випадку класичної задачі класифікації зазвичай розглядаються деякі позитивно класифіковані та деякі негативно класифіковані спостереження, в нашому випадку це могли б бути активні банки та неплатоспроможні банки. Однак, статистичних даних по неплатоспроможним банкам за всі доступні серед відкритої статистики НБУ періоди значно менше ніж по активним установам, що теоретично може негативно вплинути на якість подальшого моделювання за використання такого набору даних. Крім того, ми більш схильні до підходу до аналізу через визначення відносного стану банківських установ через аналіз їх розміщення один відносно одного відповідно до показників ефективності, надійності діяльності; через порівняння груп банків та їх характеристик між собою; через аналіз динаміки цих показників різних груп банків в часі; через аналіз впливу певних показників чи груп показників на формування груп (кластерів) банків. В свою чергу, вирішення задачі

класифікації зазвичай дозволяє отримати абсолютні дані щодо певної властивості об'єкту, наприклад провести прогнозування щодо платоспроможності певного банку в майбутніх періодах. Але визначити деякі ризикові банки може бути вкрай складно, а з урахуванням обмеження статистичних даних, навіть неможливо: класифікаційна модель може надавати недостовірні дані щодо платоспроможності деяких банків, і ми припускаємо, що для нашого дослідження помилка моделювання може бути неприйнятною. Тому нами було вирішено застосувати підходи картографування (через отримання двовимірних проекцій за допомогою алгоритмів зниження розмірності) з подальшим аналізом, та кластеризації при алгоритмах зниження розмірності. Зважаючи на поставлені цілі дослідження і специфіку задач класифікації та кластеризації, ми вважаємо, що вказані вище підходи будуть найбільш ефективними для побудови моделі аналізу фінансового стану банків.

Спочатку розглянемо основні, найбільш вживані алгоритми зниження розмірності. Загалом алгоритмів зниження розмірності існує досить велика кількість, але так як метою даної роботи не є дослідження саме даного типу алгоритмів, ми коротко розглянемо лише деякі з них: ті, які ми тестували та застосовували в рамках даного дослідження. Суть поняття зниження розмірності, постановка задачі, причини застосування більш детально розглянуті в Розділі 2.3 даної роботи.

Метод головних компонент (Principal Components Analysis, PCA). Винайдений Карлом Пірсоном у 1901 році, це один із найпоширеніших і найчастіше вживаних методів зниження розмірності багатовимірних даних, що направлений на збереження важливої інформації та максимізацію варіативності даних. Виконання алгоритму передбачає стандартизацію даних, розрахунок матриці коваріації (що показує парну кореляцію між кожною з ознак), розклад матриці коваріації, сортування власних значень за спаданням. Власний вектор з найбільшим власним значенням є першим головним компонентом. Більші власні значення вказують на вище значення загальною поясненої дисперсії. В залежності від обраної дослідником кількості вимірів обирається N власних векторів, що

стають N головними компонентами. Зазвичай, при виборі кількості компонент керуються таким правилом: додаткові головні компоненти мають пояснювати суттєву частку загальної дисперсії.

Основні переваги: в результаті роботи дозволяє отримати простір меншої розмірності, сприяючи підвищенню ефективності машинного навчання, усуненню кореляції та пом'якшенні ефекту перенавчання; відносна простота алгоритму, ресурсоефективність.

Основні недоліки: отримані головні компоненти складно інтерпретувати, визначити які саме сторони досліджуваного об'єкта описують нові отримані вектори є нетривіальною задачею; алгоритм базується на наявній кореляції між ознаками, якщо її немає – алгоритм буде неспроможний побудувати головні компоненти; PCA є дуже чутливим до викидів даних та вимагає обов'язкової попередньої стандартизації даних; алгоритм вкрай неефективно працює з даними нелінійного характеру та з пропущеними у вибірці значеннями.

Короткий огляд алгоритму представлено у роботі Карамізаде та співавторів [106].

Для подолання одного з недоліків PCA, а саме лінійної обмеженості алгоритму, було впроваджено розширення **ядерний метод головних компонент** (Kernel PCA), що може бути застосоване для зниження розмірності нелінійних даних через застосування ядер. В якості ядра можуть бути застосовані різні функції, проте частіше за все застосовують RBF-ядро (радіально-базисна функція Гауса):

$$k(x, y) = \exp(-\gamma \|x - y\|^2) \quad (1.9)$$

Робота алгоритму подібна до роботи класичного PCA, однак на першому етапі проводиться проектування даних у простір вищої розмірності, в якому поліноми вищих порядків стають лінійними функціями, і нелінійні дані можуть бути розділені гіпер-площиною. При цьому, ядерна функція застосовується для отримання власних векторів – головних компонент для простору вищої розмірності, уникаючи збільшення кількості ознак у результаті утворення поліноміальних комбінацій оригінальних ознак. Такий підхід, за якого

застосування ядерних функцій дозволяє проводити зниження розмірності нелінійних даних, називається «ядерний трюк» (kernel trick).

Алгоритм має ті ж самі переваги та недоліки, що і класичний PCA, за виключенням обмеження в роботі з нелінійними даними. Більш детально модифікація описана в роботі Шолькопфа та співавторів [107].

Багатовимірне шкалювання (Multidimensional Scaling, MDS). Класичний метод багатовимірного шкалювання також відомий як Метод головних координат (Principal Coordinates Analysis). Це засіб візуалізації рівня подібності індивідуальних спостережень набору даних. MDS застосовується для переведення інформації про попарні відстані між наборами з n об'єктів чи спостережень у конфігурацію з n точок, нанесених на вісь координат (так званий, картезіанський простір).

З більш технічної точки зору, MDS відноситься до групи зв'язаних ординаційних методів, що застосовуються для візуалізації інформації, зокрема для відображення інформації, що зберігається у матриці відстані. Представляє з себе алгоритм нелінійного зниження розмірності.

Отримуючи на вхід матрицю відстаней з відстанями між кожною парою об'єктів у наборі даних, і бажану кількість вимірів N , алгоритм MDS розміщує кожен з об'єктів у N -вимірному просторі (низькорозмірне представлення) таким чином, щоб між-об'єктні відстані збереглись якомога краще.

Класичний MDS заснований на факті, що бажана матриця координат X може бути отримана з розкладення власних значень з $B = XX'$, а матриця B може бути обчислена на основі матриці близькості D методом подвійного центрування.

Більш детально алгоритм MDS та його математична база описана в роботі Джеймса Рамсі, що зробив ключовий теоретичний внесок у розвиток алгоритму [108]. Крім класичного MDS існують також модифікації: метричне багатовимірне шкалювання, неметричне багатовимірне шкалювання, узагальнене багатовимірне шкалювання.

Основними недоліками методу є висока вимогливість до обчислювальних ресурсів, в порівнянні з іншими використовуваними методами MDS є

найповільнішим; так як метод є оптимізаційним по суті, в результаті оптимізації він може застрягати в локальному оптимумі; на вхід алгоритму приймається виключно симетрична матриця. MDS зазвичай використовується у випадках коли відстань між даними має важливу роль і задача полягає у якомога точнішому збереженні відстаней між даними після зниження їх розмірності.

Ізометрична проекція (ISOMAP). Може розглядатись як розширення алгоритмів MDS або KPCA. Відноситься до класу алгоритмів для вирішення задач на основі багатовидів (manifolds). Алгоритм намагається вбудувати дані в менш розмірне середовище, зберігаючи геодезичну відстань між всіма точками. Геодезична відстань застосовується в алгоритмах заснованих на спектральній теорії графів. Формально, це відстань між двома вершинами в графі, що виражена у кількості ребр на найкоротшому шляху між цими двома вершинами.

ISOMAP починає роботу з визначення найближчих сусідів для кожної точки. Будується мережа (граф) сусідства. Визначається найкоротший шлях між двома вершинами (наприклад, через алгоритм Дійкстри). Сума ваг ребр впродовж цього шляху визначається як геодезична відстань. Будується матриця геодезичних відстаней. Перші n власних векторів такої матриці представляють координати в новому n -вимірному евклідовому просторі.

Використання евклідової відстані у нелінійних багатовидах є адекватним коли структура сусідства може бути представлена в лінійному вигляді. Геодезична відстань дає змогу оцінити розміщення даних на «розгорнутому» багатовиді, що дає адекватну оцінку відстані точок на нелінійних багатовидах.

Більш детально про алгоритм та його роботу можна прочитати в роботі Дж. Тененбаума [109].

Перевагою алгоритму є здатність до збереження як локального так і глобального розташування точок завдяки апроксимації геодезичної відстані між двома точками. Також алгоритм добре справляється з нелінійними багатовимірними даними. Алгоритм не є ітераційним, отже є ресурсно-ефективним. **Недоліки** алгоритму проявляються у випадку коли набір даних в

багатовимірному просторі представляє з себе не неперервний багатовид, а групи точок з прогалинами між ними.

ISOMAP відноситься до класу так званих ядерних методів головних компонент заснованих на графах (Graph-based Kernel PCA), так як заснований на побудові низькорозмірного представлення даних із застосуванням функції ціни (втрат) що зберігає локальні властивості даних і заснований на теорії графів.

Спектральне вкладення (Spectral Embedding, SE) - відноситься до класу алгоритмів для вирішення задач на основі багатовидів. Алгоритм проводить пошук низько-розмірних представлень даних з використанням спектрального розкладу графу Лапласа. Отриманий граф можна вважати дискретним наближенням низько-розмірного багатовиду у просторі вищої розмірності. Мінімізація функції втрат заснована на цьому графі забезпечує, що точки що знаходяться поруч на багатовиді, в низько-розмірній проекції також будуть знаходитись поруч, зберігаючи локальні відстані. Алгоритм також має іншу назву: власні карти Лапласа (Laplacian eigenmaps).

Першим кроком алгоритму є побудова зваженого графу: перетворення даних в графове представлення з використанням матриці схожості (сусідства). Другим кроком є побудова графа Лапласа. Третім кроком частковий розклад на власні значення: проводиться розклад на власні значення на графі Лапласа [110, 111].

Spectral Embedding, маючи графову природу, часто застосовується для аналізу медіа даних: фото, відео, аудіо.

Загалом, нами було протестовано 6 різних алгоритмів зниження розмірності, окрім розглянутих вище ми також проводили експерименти з алгоритмами t-SNE та UMAP. Ми не будемо зупинятись на їх огляді в рамках даного Розділу, так як детально будемо розглядати їх в підрозділі 2.3. Однак, забігаючи наперед відзначимо, що саме ці два алгоритми були обрані нами для подальшої побудови комплексної моделі аналізу фінансового стану комерційних банків України. Обираючи алгоритм зниження розмірності, ми спирались на декілька критеріїв: висока якість візуалізації, роздільна здатність даних на групи; висока ефективність збереження локальної та глобальної структури даних, бажано: можливість

впливати на стратегію збереження структури даних; відносна швидкість алгоритму – хоча ми й не оперуємо над-великими масивами даних, проте все ж набір даних нараховує декілька тисяч спостережень, а отже алгоритм не повинен бути надто ресурсоемним; алгоритм повинен бути ефективним при роботі з нелінійними структурами та з шумом чи викидами в даних, так як ми маємо припущення про нелінійну природу даних та точно знаємо, що набір даних містить викиди; алгоритм повинен бути ефективний в роботі з перервними формами даних, тобто з даними що представляють з себе групи точок. Задоволення перших двох критеріїв має також сприяти досягненню прийнятного рівня якості кластеризації.

Питання застосування алгоритмів зниження розмірності для вирішення задач в області банківської справи не є широко освітленими у вітчизняній науковій літературі. Можна виділити статтю Ткаченко Н., присвячену застосуванню методу головних компонент для оцінювання впливу роботи банків з участю транснаціонального капіталу на банківську систему [112], серію статей Заруцької О. П. [113, 114, 115] в яких застосовується ще один алгоритм, проекція Саммона, в рамках алгоритму карт самоорганізації Кохонена для виконання кроку зниження розмірності вхідних даних.

У іноземній літературі трохи більше уваги присвячено даному питанню, застосовуються різні алгоритми в контексті дослідження різного роду проблем банківських секторів різних країн. Так, наприклад, у статті Орсеніго [116] проводиться порівняння лінійних та нелінійних методів зниження розмірності для прогнозування кредитного рейтингу банків, порівнюються PCA та різновид ISOMAP. В роботі Валарматі Б. та співавторів [117] застосовується специфічний метод CfsSubsetEval для зниження розмірності через оцінювання під-набору ознак на основі прогнозної здатності алгоритму і визначення дублікатів серед відібраних ознак.

Однак, алгоритми зниження розмірності рідко застосовують поодиночі для вирішення яких-небудь задач. Часто, вони підкріплюються саме алгоритмами кластеризації, і в парі застосовуються з метою досягнення більшої ефективності кластерного аналізу даних. Окрім іншого, для вирішення нашої задачі ми

спираємось на комплексний підхід. Тож далі ми розглянемо найбільш вживані та поширені методи і алгоритми кластеризації. Ми не будемо детально розглядати самі алгоритми та їх математичний апарат, так як в рамках нашого дослідження нами буде використаний лише один з них, і ми вважаємо що більше уваги необхідно приділити саме тому, який буде використаний для подальшого моделювання. Однак нами будуть вказані відповідні роботи авторів алгоритмів, чи дослідників що присвятили роботи застосуванню розглядуваних алгоритмів.

К-середнє (K-means). Відноситься до класу алгоритмів, що базуються на розподілі даних (based on partition). Основна ідея таких алгоритмів - розглядати центри скупчень точок як центри відповідних кластерів. Алгоритм виконується в декілька кроків:

1. Визначається кількість кластерів (суб'єктивно, на розсуд дослідника).
2. Обирається k спостережень випадковим чином, що на першому кроці приймаються за центри кластерів.
3. Кожне спостереження призначається одному з n кластерів, того до якого відстань найкоротша.
4. Розраховується новий центр кожного кластера як середнє арифметичне ознак об'єктів, що входять у цей кластер.
5. Відбувається така кількість ітерацій (повторюються кроки 3-4), поки кластерні центри стануть стійкими (тобто при кожній ітерації в кожному кластері опинятимуться одні й ті самі об'єкти), дисперсія всередині кластера буде мінімізована, а між кластерами — максимізована.

Переваги. Простота та швидкість алгоритму. Метод k-means більш зручний для кластеризації великої кількості спостережень. Алгоритм переважно формує більш щільні кластери. Він також є ефективним при роботі з сферичними кластерами.

Недоліки. Результат кластеризації сильно залежить від випадкових початкових позицій центрів кластерів; алгоритм чутливий до викидів, що викривляють середні значення; кількість кластерів має бути задана наперед; алгоритм неефективний якщо дані мають неопуклу природу.

Ключові засади алгоритму описані в роботах Гуго Штейнгауза [118], Стюарта Ллойда [119], Маккуїна [120].

Заснована на щільності просторова кластеризація для застосування з шумами (Density-based spatial clustering of applications with noise, DBSCAN).

Відноситься до алгоритмів, що засновані на щільності (based on density). В алгоритмах даного типу для заданої множини точок у деякому просторі в одну групу відносяться точки, які розташовані найбільш щільно (точки з багатьма сусідами) та відмічаються як викиди точки, які лежать в областях з невеликою щільністю (чиї сусіди розташовані занадто далеко).

Алгоритм задається двома параметрами: розмір околу (eps) та мінімальна кількість точок необхідних для утворення щільної області ($minPts$).

Метою кластеризації процедурою DBSCAN є поділ точок на ядрові, щільно-досяжні та викиди. Ядрова точка p утворює кластер разом з усіма точками досяжними з неї (ядровими чи ні). Кожен кластер містить принаймні одну ядрову точку. Неядрові точки можуть бути частиною кластеру, тоді вони утворюють «ребро» кластеру, оскільки з них не будуть досяжні інші точки. Досяжність не є симетричним відношенням. Дві точки p та q є щільно-зв'язними, якщо існує точка o з якої вони є досяжними. Щільна-зв'язність є симетричною. Всі точки кластера взаємно щільно-зв'язні, якщо точка є щільно-зв'язна з будь-якою точкою кластера, то вона так само належить кластеру. Всі точки не досяжні з будь-якої іншої точки є викидами.

Алгоритм має наступні кроки:

1. Знайти точки в кожному ε -околі кожної точки та визначити ядрові точки, у яких більше ніж $minPts$ сусідів.
2. Знайти компоненти зв'язності для ядрових точок на графі сусідів, виключивши з розгляду точки, які не є ядровими.
3. Приєднати кожну не ядрову точку до найближчого кластера, за умови, що кластер знаходиться в ε -околі, інакше помітити як шумову.

Переваги. Немає необхідності задавати кількість кластерів апіорно; алгоритм знаходить кластери довільної форми; надійно виявляє аномалії

керуючись поняттям шуму; здебільшого нечутливий до впорядкування точок; має всього два керуючих параметри, які при розумінні суті даних можуть бути визначені експертним шляхом.

Недоліки. Алгоритм не є детерміністичним: точки на межі, які досяжні з декількох кластерів, можуть належати одному або іншому кластеру в залежності від порядку обробки даних (для більшості наборів даних та областей, така проблема з'являється нечасто і має незначний вплив на результат кластеризації бо шумові та ядрові точки визначаються однозначно); якість кластеризації залежить від функції відстані, а використовувана евклідова відстань може бути неефективною при кластеризації багатомірних даних через «прокляття розмірності»; не може кластеризувати набори даних з великим перепадом щільностей, оскільки неможливо підібрати поєднання значень *minPts* – *eps*, яке б відповідало різним кластерам; якщо дані та масштаб не дуже зрозумілі, то вибір адекватного порогового значення відстані *eps* може бути складним.

Більш детально алгоритм розглянутий в роботі Еріка Шуберта в співавторстві з іншими дослідниками [121].

Алгоритм має різні модифікації, серед яких можна виділити OPTICS та HDBSCAN.

Впорядкування точок для визначення кластерної структури (Ordering points to identify the clustering structure, OPTICS). Його основна ідея схожа на DBSCAN, але він вирішує одну з основних слабостей DBSCAN: проблему визначення значущих кластерів в наборах даних різної щільності. Для цього об'єкти бази даних повинні бути впорядковані так, що об'єкти, які просторово близькі, будуть сусідами в упорядкуванні. Крім того, особлива відстань зберігається для кожної точки, яка являє собою щільність, яка повинна бути прийнятна для кластера, щоб обидві сусідні точки належали до тієї ж групи. Це представляється у вигляді в дендрограми. Модифікація представлена Міхаелом Анкерстом та співавторами [122].

Ієрархічний DBSCAN (Hierarchical DBSCAN). Алгоритм є розширенням DBSCAN, що перетворює його в ієрархічний алгоритм кластеризації, після чого

використовується техніка «витягування пласкої кластеризації», що базується на стабільності кластерів.

У порівнянні з іншими алгоритмами, що засновані на щільності точок, HDBSCAN є швидшим та більш адекватно реагує на дані з різною щільністю даних. Алгоритм представлений Кампело з співавторами у роботі 2013 року [123].

Алгоритм зміщення середнього значення (Mean Shift). Mean Shift можна віднести до класу алгоритмів, що орієнтуються на щільність точок. Також даний алгоритм відносять до центроїдних, тобто таких що працюють, оновлюючи набір кандидатів в центроїди таким чином, щоб вони ставали середнім значенням всіх точок в межах визначеної області. Після цього кандидати фільтруються на етапі після-обробки для виключення даних наближених до кандидатів (дублікати) та формується фінальний набір центроїдів.

Для вирішення задачі кластеризації Mean Shift застосовується наступним чином. Розглядається набір точок в двовимірному просторі. Розглядається кругле вікно з центром в C з радіусом r в якості ядра. Mean Shift є алгоритмом пошуку екстремуму, що зсуває ядро ітеративно до області з більшою щільністю, доки процес не досягне сходження. Будь-який зсув визначається вектором зсуву середнього значення. Цей вектор завжди вказує в напрямку максимального збільшення щільності. На кожній ітерації ядро зсувається до центру тяжіння (центроїду) чи середнього значення точок всередині нього. Метод обчислення цього середнього визначається вибором ядра. Якщо обране гаусове ядро (Gaussian kernel) замість плаского ядра (flat kernel), то кожній точці буде призначена вага, що зменшуватиметься експоненційно зі збільшенням відстані до центра ядра. Коли процес досягне сходження, не залишиться напрямку в якому зсув зможе забезпечити вміщення більшої кількості точок всередині ядра. Алгоритм описаний в роботі Їзонга Ченга [124].

Переваги. Здатність до кластеризації скупчень неопуклої форми; немає необхідності вказувати кількість кластерів апріорі, є тільки один параметр *bandwidth* що описує локальний масштаб; алгоритм є робустним, він не має

локальних мінімумів а отже не потребує повторної ре-ініціалізації; алгоритм є слабо-чутливим до викидів.

Недоліки. Алгоритм не придатний до кластеризації у високій розмірності; якщо задача потребує задання кластерів апіорі – досягнення цільової кількості кластерів через перебір параметру *bandwidth* є обчислювально обтяжливим процесом; в процесі роботи можуть бути сформовані «хибні» кластери, знаходження яких поруч не має адекватного економічного обґрунтування.

Спектральна кластеризація (Spectral clustering). Спектральна кластеризація формує собою окремий клас алгоритмів. Алгоритми даного класу використовують спектр (власні значення) матриці подібності даних для зниження розмірності перед кластеризацією в просторі меншої розмірності. Матриця подібності подається в якості входу і складається з кількісних оцінок відносної схожості кожної пари точок даних.

Інший опис надається у Вон Люксбурга [125]. Основною ідеєю даного виду алгоритмів кластеризації є відношення до об'єкту як до вершини, а до подібності між об'єктами як до зваженого ребра з метою перетворення задачі кластеризації на задачу розбиття графу. В основі лежить пошук методу розділення графу таким чином, щоб ваги зв'язків між різними групами були найменшими з можливих а загальна вага зв'язків між ребрами в межах найменшої групи – найбільшою з можливих.

Алгоритм спектральної кластеризації, представлений у *sklearn*, виконує вкладення матриці близькості між прикладами у простір меншої розмірності, після чого виконує кластеризації (алгоритмом K-means за замовчуванням) компонентів власних векторів у просторі зниженої розмірності. *Sklearn* імплементує SM алгоритм, що більше підходить для роботи з зображеннями. Для двох кластерів алгоритм вирішує задачу випуклого послаблення нормалізованого зрізу (*convex relaxation of the normalized cut*) на графі подібності: розрізаючи граф навпіл таким чином, щоб ваги ребр зрізу були малими порівняно з вагами ребр всередині кластеру.

Переваги. Алгоритм слабо чутливий до шуму; алгоритм добре підходить для кластеризації даних високої розмірності; алгоритм добре визначає кластери неопуклої форми.

Недоліки. Алгоритм є ресурсо-неефективним; необхідне задання кількості кластерів апіорі.

Збалансоване ітеративне скорочення і кластеризація за допомогою ієрархій (Balanced iterative educing and clustering using hierarchies, BIRCH). Алгоритм інтелектуального аналізу даних без вчителя, що використовується для ієрархічної кластеризації на наборах даних великого об'єму. Алгоритм відноситься до класу ієрархічних. Основна ідея алгоритмів даного типу – побудувати ієрархічний взаємозв'язок між даними для кластеризації. Припустивши, що кожна точка на початку процесу кластеризації є окремим кластером, далі можна з'єднувати кластери, що знаходяться поруч до тих пір поки не залишиться один кластер (або навпаки).

Алгоритм детальніше представлено та описано в роботі Жанга Т. та співавторів [126].

Переваги. Перевагою BIRCH є здатність методу динамічно кластеризувати по мірі надходження багатомірних метричних точок даних, намагаючись отримати кластеризацію кращої якості для існуючого набору ресурсів (пам'яті / часу); BIRCH ефективно оброблює шум; алгоритм добре працює з даними довільної форми. Добре підходить для роботи з зображеннями, для компресії зображень. BIRCH в більшості випадків потребує лише одного проходу по даним для кластеризації.

Недоліки. Ресурсо-затрати для роботи алгоритму відносно високі; необхідно зазначати кількість кластерів апіорі; працює виключно на числових даних.

Моделі гаусових сумішей (Gaussian Mixture Models, GMM). GMM можна віднести до класу алгоритмів, що базуються на розподілі даних. Головною ідеєю таких алгоритмів є те, що дані з тим самим розподілом мають належати до одного й того самого кластеру у випадку, якщо в наборі даних існують декілька розподілів.

Іноді GMM відносять до так званого класу алгоритмів, що базуються на моделі. Основною ідеєю таких алгоритмів є вибір певної моделі для кожного кластеру з метою пошуку найвищої ефективності розподілу даних на кластери.

GMM – це імовірнісна модель, що припускає що всі точки даних згенеровані сумішшю обмеженої кількості Гаусових розподілів з невідомими параметрами. Можна розглядати суміш моделей як узагальнення кластеризації з K-means для включення інформації про коваріаційну структуру даних, а також про центри прихованих гаусіан.

Основною складністю для алгоритму GMM у випадку навчання без міток даних (навчання без вчителя) є те, що зазвичай невідомо які точки від яких прихованих компонент походять (якщо є доступ до такої інформації, стає дуже просто застосувати певний гаусовий розподіл до кожного набору точок).

Для вирішення даної проблеми застосовується добре відомий статистичний ітеративний алгоритм очікування-максимізація (expectation-maximization). Перша частина алгоритму припускає наявність випадкових компонент (випадково центровані щодо точок даних, визначені через k-means або навіть просто нормально розподілені навколо початкових даних) і розраховує імовірність можливості того, що вони згенеровані кожним компонентом моделі.

Потім можна налаштувати параметри, щоб максимізувати ймовірність даних, що отримують ці призначення. Повторення даного процесу гарантує, що він завжди зійдеться до локального оптимуму.

Переваги. На відміну від k-means, точки при GMM кластеризації можуть належати до декількох кластерів з різною мірою приналежності, можна оцінити міру належності кожної точки до кожного кластеру; кластери описуються невеликою кількістю параметрів; GMM більш гнучкий з точки зору коваріації кластерів (у випадку з k-means, коваріація при кожному вимірі наближається до 0, тому точки приписуються до кластеру найближчого до точки; для GMM кластери можуть мати необмежену коваріаційну структуру і через розподіл даних може приймати довільну форму (не тільки сферичну)).

Недоліки. Необхідність задавати кількість компонент апріорі, що по суті є аналогією кількості кластерів; алгоритм може потрапити в локальний оптимум, що може бути подолане через багатократний запуск алгоритму з випадковою ініціалізацією параметрів; у випадку, коли кількість компонентів велика – алгоритм може мати високу обчислювальну складність; використання алгоритму не є ефективним якщо кількість точок спостережень є незначною; складно об’єктивно оцінити оптимальну кількість кластерів; алгоритм підходить для кластеризації виключно опуклих / пласких наборів даних.

Алгоритм GMM, специфіку кроків, що виконуються в процесі його виконання, та його математичну базу детально описано у роботі Адріана Біарнса [127].

В процесі відбору алгоритму кластеризації, ми розглядали їх сильні та слабкі сторони, і проводили порівняльний аналіз через призму різних факторів. Серед ключових факторів, що ми враховували: адекватність кластеризації даних переважно типової форми, тобто у вигляді скупчення точок, однак з урахуваннями можливої відмінності у щільності різних кластерів та різної віддаленості самих груп даних між собою; адекватність кластеризації даних, що містить шуми або викиди; адекватність кластеризації багатовимірних даних (так як ми апріорі не знаємо, зниження розмірності до якого рівня забезпечить кращий результат проектування даних і маємо розглядати потенційну необхідність кластеризації даних розмірністю вище 2-ох вимірів); складність налаштування алгоритму, з точки зору кількості гіпер-параметрів, що потребують встановлення апріорі; швидкість та ресурсо-ефективність алгоритму. Крім того, ми проводили численні тести, аналізуючи різні комбінації алгоритмів зниження розмірності та кластеризації, звертаючи увагу на такі метрики якості кластеризації як силует, оцінка Калінського-Харабаза та оцінка Девіса-Боулдіна; на адекватність ідентифікації алгоритмом кластерів на карті (у випадку, якщо проводився кластерний аналіз 2-ох або 3-ох вимірних простору) та на наявність економічного сенсу у розподілі банків між кластерами.

В результаті проведеного аналізу нами було вирішено зупинитись на агломеративній ієрархічній кластеризації із застосуванням методу Уорда для визначення відстані між кластерами. Даний алгоритм із застосуванням методу Уорда добре підходить для кластеризації даних у формі скупчень різних видів, добре справляється з викидами в даних. Алгоритми ієрархічного типу загалом непогано справляються з даними високої розмірності. Алгоритм не потребує задання кількості кластерів апріорі і може керуватись одним ключовим гіперпараметром. Він є не дуже ефективним для роботи з великими масивами даних, проте наше дослідження не передбачає роботу з об'ємними наборами даних, а при застосуванні на наборах даних малого та середнього обсягу алгоритм продемонстрував високу швидкість обробки даних. Також алгоритм дозволяє сформувати унікальний формат візуалізації результатів кластеризації – дендрограму. Її аналіз дозволяє інтерпретувати результати розподілу даних, а також наочно визначати оптимальну кількість кластерів. Візуальний, аналітичний та економічний аналіз отриманих результатів із використанням алгоритму ієрархічної кластеризації також показав доцільність його застосування в рамках нашої моделі. Більш детально обраний алгоритм буде розглядатись у Розділі 2 дисертаційної роботи.

Застосування кластерного аналізу в цілому є досить поширене серед вітчизняних науковців. Так, в роботах Негрей М. В. представлено застосування алгоритмів нечіткої кластеризації та K-means в рамках побудови методологічного підходу до оцінювання позичальників небанківських установ [128] та для підтримки управління портфелем з включенням компоненту стійкості на основі EGS-оцінок [129].

Якщо розглядати дослідження саме банківської системи за допомогою кластерного аналізу, можна відмітити, що багато уваги приділено задачі визначення бізнес-моделей національних банків. Даній тематиці присвячені статті Соцької Ю. І. [130], Рашкован В. [131], Примостки Л. О. [132]. Карти самоорганізації Кохонена застосовуються у роботі Мікуляк К. А. [133], в якій

проводиться кластерний аналіз банківських установ в контексті типів банківських операцій та обсягів здійснюваної банками діяльності. Серія статей Заруцької О. П. [113, 114, 115], що вже згадувалась вище, також висвітлює застосування кластерного аналізу, а саме карт самоорганізації, для вирішення задачі структурно-функціонального аналізу банківської системи, фінансової стійкості банківських установ. Проте, варто відзначити, що з моменту проведення більшості досліджень вітчизняної банківської системи дещо змінилася її структура: кількість банків значно зменшилася, НБУ змінила принципи розподілу комерційних банків на групи. Застосування саме ієрархічної кластеризації не є широко вживаним в роботах вітчизняних дослідників. Вартими уваги є дослідження Гур'янової Л. С., в якому застосовується алгоритм агломеративної ієрархічної кластеризації в рамках побудови методологічного підходу до класифікації регіонів України за рівнем соціальної напруги [134] та в рамках розробки комплексу моделей системного аналізу ринку освітніх послуг України [135].

У іноземних виданнях трохи більше уваги приділено питанню кластеризації банківських установ а спектр застосовуваних алгоритмів є ширшим. Так, стаття Гумбольта Т. з BNP Paribas [136] присвячена знову ж таки питанню кластеризації бізнес-моделей, але вже європейських банків за допомогою алгоритму ієрархічної кластеризації. У статті Негневіцького М. [137] розглядається задача визначення проблемних банків за допомогою нейромережі типу карт самоорганізації. Кластеризації українських банків з використанням агломеративної ієрархічної кластеризації типу Уорда з метою їх розподілу на три групи («зростання», «стабілізація», «спад») присвячена робота Колодіз'єва і співавторів [138]. Ієрархічний кластерний аналіз також був застосований в роботі Дао та Хана [139] для ідентифікації та відділення проблемних банків від інших на основі 12 показників їх роботи.

Національним банком України (НБУ) також було запропоновано підхід до кластеризації банків, який мав на меті їх поділ на державні; великі приватні; іноземні, що належать до банківських груп; російські; середні приватні; ринкові;

неринкові. Ринкові у свою чергу поділяються на бізнесові, ризикові, роздрібні, банки з іноземними інвесторами. Неринкові діляться на бізнесові, схемні, кишенькові та заморожені (неактивні). Цей підхід частково розглянуто та покращено у статті Коваленко та співавторів [140]. При цьому використовувані методи та алгоритми кластеризації не розкривалися, і з тих пір цей підхід так і не був впроваджений.

Згадані роботи Колодізієва, Негневіцького та Дао і Хана у яких фокус був зроблений на поділ банків на кластери виходячи з показників їхньої фінансової ефективності, є найближчими до нашого дослідження, однак у них використовуються невеликі «вузькі» набори показників, тобто такі що в основному спрямовані на одну певну сторону банківської діяльності. У роботі Колодізієва розглядається український ринок та використані в основному показники прибутковості, при цьому дослідження проводилося у 2016 році та його результати втратили актуальність, оскільки відтоді банківська система зазнала значних змін. Враховуючи також, що на даний момент в Україні на законодавчому рівні не запроваджено та не закріплено систему поділу банків, яка ґрунтувалася б на твердій економіко-математичній методології, ми б хотіли розширити представлені вище дослідження, на прикладі українських банків, використовуючи ширший спектр показників та запропонувати авторський підхід до поділу банків на групи/кластери та аналізу їх фінансового стану.

Висновки до розділу 1

В першому розділі дисертаційної роботи розглянуто ряд статистичних показників НБУ, на основі яких наведено характеристику поточного стану банківської системи, виділено та описано наявні в ній тенденції.

Проведено детальне та глибоке дослідження категоріального, понятійного апарату предметної області. Розглянуто поняття фінансового стану, ефективності діяльності банківської установи, фінансової стійкості, прибутковості, конкурентоспроможності, ліквідності, платоспроможності, невизначеності,

ризиків. Запропоновані авторські визначення таких понять як фінансовий стан комерційного банку, ефективність банківської діяльності, фінансова стійкість банківської установи. Представлене авторське бачення взаємозв'язку розглянутих категорій, що є контекстом для дослідження, що проводиться.

Категорія ризику виділена як ключова в питанні аналізу фінансового стану комерційних банків. Ризики та наслідки їх реалізації мають безпосередній та суттєвий вплив на результати діяльності банківських установ, а отже і на їх фінансовий стан.

Проведено ретельний якісний аналіз банківських ризиків, розглянуто основні класифікації ризиків, побудовано авторську розширену класифікацію ризиків, досліджено взаємозв'язки між різними видами ризиків. Кредитний, операційний та ринковий ризики виділено як основні види ризику, з якими мають справу суб'єкти банківської діяльності. Прийняття даних видів ризиків приносить найбільшу частку банківських доходів і загрожує найбільш відчутними збитками у випадку їх реалізації.

Розглянуто найбільш популярні вітчизняні та іноземні класичні моделі та методи дослідження ефективності та ризикованості діяльності підприємств в цілому, та банківських установ зокрема. Виділено основні недоліки таких моделей для проведення моделювання у сучасних реаліях. При цьому варто відзначити, що деякі розглянуті класичні методи аналізу можуть мати в своїй основі показники діяльності підприємств (показники ефективності чи ризикованості), що заслуговують на увагу. Такі показники можуть бути використанні для побудови моделі аналізу фінансового стану комерційних банків або для розширення цієї моделі.

Прийнято рішення зупинитись на алгоритмах та методах машинного навчання, виділивши їх серед усього різноманіття методів та моделей. Рішення на користь даного напрямку моделювання було прийняте на основі ряду переваг, що мають алгоритми та методи даного напрямку, як то інноваційність, універсальність, відносна простота реалізації (для деяких алгоритмів), висока

якість візуалізації та сприяння простоті подальшої інтерпретації результатів моделювання тощо.

Додатково до перелічених проміжних висновків, на основі ґрунтовного аналізу проведеного в Розділі 1 ми також дійшли до ряду загальних висновків і рекомендацій стосовно підходу до подальших досліджень:

- В рамках аналізу комерційних банків пропонується системний підхід, за якого діяльність банківської установи не розглядається ізольовано, а досліджується в середовищі предметної області, тобто в умовах функціонування інших комерційних банків. За такого підходу будуть враховані не тільки показники ефективності та ризикованості діяльності кожного банку а й зв'язки всіх банківських установ, що також можуть мати важливе значення в процесі вирішення задачі аналізу фінансового стану. Для реалізації такого підходу було вирішено проводити моделювання всієї банківської системи а не її відокремлених суб'єктів.
- Як один з результатів проведеного в теоретичному розділі даної роботи фундаментального аналізу, є висновок щодо суттєвого впливу ризиків та результатів їх реалізації на фінансовий стан комерційних банків та на показники ефективності їх діяльності. Зроблено акцент на важливості аналізу ризикованості діяльності банківських установ як ключового напрямку в задачі аналізу їх фінансового стану.
- Іншим висновком з проведеного фундаментального аналізу є те, що найбільш поширені методи і моделі оцінювання банківських ризиків призначені для аналізу виключно одного виду ризику; їх застосування відокремлено не дає можливості побачити всю картину ризиків банківської діяльності в цілому. А застосування методів і моделей оцінювання різних видів ризиків в комплексі частіше за все передбачає необхідність організації значної кількості складних процесів оцінювання, формування складних політик та нормативів застосування комбінації різних моделей, оптимізацію механізму ефективного застосування системи даних моделей в рамках конкретної організації. Все це означає

суттєве ускладнення, перевантаження самого процесу ризик-менеджменту. Ми вважаємо доцільним формування комплексного підходу до аналізу ризикованості банківських установ на основі ряду показників, коефіцієнтів що відображають різні сторони банківської діяльності. Таким чином, ми забезпечуємо повноту аналізу не через оцінювання різних конкретних видів ризику за допомогою складної комбінації різних прямо не пов'язаних між собою методів та моделей, а через застосування однієї моделі, побудованої на основі ряду коефіцієнтів, які характеризують основні, ключові напрямки діяльності комерційних банків. Такий підхід так само дозволяє врахувати різні види ризику та забезпечити ефективність вирішення поставленої нами задачі аналізу фінансового стану комерційних банків за використання меншої кількості ресурсів.

Основні результати Розділу 1 опубліковано у наукових працях автора: [141, 142, 143, 144, 145, 146, 147, 148].

РОЗДІЛ 2

СИСТЕМА МОДЕЛЕЙ АНАЛІЗУ ФІНАНСОВОГО СТАНУ КОМЕРЦІЙНИХ БАНКІВ УКРАЇНИ

2.1. Відбір показників та ознак до моделі аналізу фінансового стану комерційних банків

Аналіз фінансового стану комерційних банків представляє з себе нетривіальну задачу і її вирішення є складним і багатокроковим процесом, що потребує застосування ряду алгоритмів та економіко-математичних моделей. Для того, щоб відобразити та описати даний процес якомога прозоріше та зрозуміліше ми представимо його у вигляді дев'яти кроків, виконання яких дозволить сформувати та реалізувати необхідну систему моделей для досягнення поставленої мети дослідження.

Отже, розглянемо детальніше процес покроково а також схематично зобразимо його на Рис. 2.1.

Відповідно до Рис. 2.1, **крок 1 – це формування початкового набору показників**. Даний крок включає в себе огляд низки вітчизняних та іноземних джерел, в яких представлені показники діяльності та фінансового стану комерційних банків. Результатом виконання цього кроку є сформована база показників.

Крок 2. Відбір ознак. Застосування підходів та методів відбору ознак для очистки початкового набору показників від таких, які не можуть бути розраховані через обмеженість статистичної звітності НБУ, які дублюють один одного, які описують дуже подібні властивості об'єкта та які знижують загальну якість моделювання. Результатом виконання цього кроку є очищений набір показників, який використовується на наступних кроках для моделювання. Більш детально цей крок розглянуто в наступному підрозділі.

Перші два кроки забезпечують отримання моделі об'єкта дослідження, його еквівалента, який у математичній формі відображає його властивості. Фінансовий

стан комерційних банків як об'єкт моделювання є надзвичайно складним. Він залежить від великої кількості факторів, їх взаємозв'язку та безлічі внутрішніх та зовнішніх процесів якими супроводжується функціонування банківської системи. Саме тому нами було вирішено звернутись до інструментарію машинного навчання, алгоритми якого є досить сучасними та потужними для моделювання складних процесів, властивостей, взаємозв'язків, у тому числі прихованих, таких які не можуть бути виявлені класичними економетричними моделями. За допомогою цього інструментарію можливим є вирішення задачі аналізу фінансового стану комерційних банків через оцінювання як ефективності так і ризикованості діяльності банківських установ, а також можливе оцінювання загальної ситуації у банківській системі, визначення основних проблемних місць у роботі вітчизняних банків.



Рис. 2.1. Схема процесу вирішення задачі аналізу фінансового стану комерційних банків

Джерело: розроблено автором

На даному етапі ключовими елементами, що характеризують об'єкт нашого дослідження є набір вхідних показників, який найбільш повно та всебічно описує функціонування комерційного банку та його фінансовий стан.

Далі проводиться вибір алгоритму або алгоритмів машинного навчання для реалізації моделі. Як вже було зазначено вище, нами було вирішено звернутись до інструментів та методів машинного навчання, а саме - зниження розмірності та кластеризації. Ключовим для дослідження є побудова системи моделей на основі такого поєднання алгоритмів, що забезпечили би найбільш адекватні та якісні результати, з урахуванням можливої специфіки даних та об'єкта дослідження.

Перед тим як застосовувати алгоритми на відібраних даних необхідно провести їхню попередню обробку. Як показує практика, моделювання на «сирих», необроблених даних часто може призвести до похибок та отримання неадекватних, недостовірних результатів, невірної інтерпретації економічного сенсу таких результатів.

Корисним буде також попередньо перевірити дані на предмет їх придатності до кластеризації, адже отриманий після відбору ознак набір показників може взагалі не містити чітких кластерів, дані можуть бути занадто однорідними, формувати один єдиний кластер. Все це в результаті призведе до низької якості кластеризації; зниження розмірності таких даних в результаті можливо сформує таку двовимірну карту, на якій можна буде виділити кластери, проте економічний сенс у результаті такої кластеризації буде мінімальний. Тобто, тут можемо виділити наступні кроки:

Крок 3-4. Вибір алгоритму зниження розмірності та алгоритму кластеризації.

Для вирішення як задачі зниження розмірності, так і задачі кластеризації може бути застосований дуже широкий спектр алгоритмів, що поділяються на різні типи. Кожен алгоритм, залежно від типу, характеризується своєю специфікою роботи та різними наборами гіперпараметрів. З метою вибору найбільш підходящого алгоритму для моделювання нашого об'єкта дослідження нами були відібрані ті, які за принципом роботи та за особливостями взаємодії з даними

найбільше підходили для нашого набору даних, і які в ході численних експериментів показали найкращі результати. Алгоритми зниження розмірності та кластеризації будуть детальніше розглянуті в підрозділах 2.2 та 2.3. Їхнє застосування та отримані результати будуть представлені у Розділі 3.

Крок 5. Тестування даних на предмет їх придатності до кластеризації.

Для цього нами було виділено декілька підходів, які є найбільш поширеними та вживаними для тестування даних. Серед них основним є тест Хопкінса, який дозволяє провести перевірку відмінності внутрішнього розподілу набору даних від рівномірного розподілу. Більш детально дане питання буде розглянуте в наступному розділі.

Крок 6. Тестування та відбір методу первинної обробки даних. Наразі існує досить велика кількість підходів до попередньої обробки даних. Серед класичних та найбільш вживаних залишаються такі методи як стандартизація та нормалізація. Менш вживаними, але такими що заслуговують окремої уваги є стандартизація, чутлива до викидів (robust scaler) та вінзорування (winsorization). Більш детально дане питання буде розглянуте в Розділі 3.

Далі проводиться експериментальне дослідження, з виконанням програмного коду, який реалізовано в певному середовищі програмування за допомогою будь-якої з мов програмування. Ми використовували хмарне середовище Google Collab та мову програмування Python для підтримки реалізації всіх кроків процесу вирішення задачі аналізу фінансового стану комерційних банків. Більш детально інформаційне забезпечення процесу моделювання буде розглянуте в підрозділі 3.1. Тут можемо виділити наступні кроки:

Крок 7. Первинна обробка даних. Застосування відібраних на кроці 6 методів до первинної обробки даних з метою усунення викидів або зменшення їх впливу на результати моделювання, підвищення загальної якості набору даних.

Крок 8-9. Застосування алгоритмів зниження розмірності та кластеризації для аналізу фінансового стану комерційних банків України. На останніх кроках власне застосовується побудована програма на основі відібраних алгоритмів для вирішення поставленої у даному дослідженні задачі. Процес

моделювання, опис результатів та їхня інтерпретація будуть розглянуті у Розділі 3, в підрозділах 3.2 та 3.3.

Перші кроки в процесі моделювання є одними з ключових і полягають вони у формуванні такого набору фінансових коефіцієнтів, який з одного боку був би позбавлений надлишковості, та елементи якого були б несуперечними, неконфліктними, а з іншого боку – такого, який забезпечував би найвищу адекватність та ефективність роботи алгоритмів та найліпшу якість економічної інтерпретації результатів моделювання. Усі коефіцієнти мають бути співрозмірними, а весь набір коефіцієнтів має всебічно характеризувати діяльність банківської установи.

Для того, щоб сформувати такий набір показників, які найбільше б відповідали вищезазначеним критеріям, було опрацьовано низку джерел, серед яких: монографія Долінського Л. Б. [149], Методики рейтингового оцінювання агентств «Рюрик» [150] та «ІВІ – Рейтинг» [151], Методика визначення рейтингу стійкості банку, розглянута автором фінансового порталу «Мінфін» [152], роботи Заруцької О. П. [113, 114, 115] в області структурно-функціонального аналізу та інші дослідження [153, 154].

В результаті огляду наведених джерел, нами було сформовано набір, що містить 143 коефіцієнта, які характеризують діяльність комерційних банків з різних сторін. Проте, в даному дослідженні ми обмежились лише агрегованими за всіма банками статистичними даними, наявними у вільному доступі на офіційному сайті НБУ, а саме, даними за згрупованими балансовими залишками комерційних банків. Зважаючи на те, що відкрита статистика, представлена по банківській системі, містить обмежений набір даних, ми не змогли розрахувати всі з відібраних коефіцієнтів. Іншою умовою була доступність даних за досліджуваний період, адже для застосування алгоритму машинного навчання важливою є повнота вхідних змінних. НБУ з кожним роком все більше розширює перелік показників та статистики, що викладаються щоквартально або щомісячно у вільний доступ. Проте з тих даних, які цікавлять нас у рамках даного дослідження тільки інформація за згрупованими балансовими залишками банків надається НБУ з 2015

року. Таким чином, в результаті нами були відібрані тільки ті 63 з 143 показників, які піддаються розрахунку на основі відкритих статистичних даних НБУ.

Всі розраховані нами показники можна розділити на 9 груп:

- Аналіз активів: 14 показників.
- Аналіз пасивів: 8 показників.
- Аналіз достатності (адекватності) капіталу: 8 показників.
- Аналіз збалансованості активів та пасивів: 6 показників.
- Аналіз ліквідності: 3 показники.
- Аналіз фінансової ефективності: 16 показників.
- Аналіз податкового навантаження: 3 показники.
- Аналіз ефективності використання (залучення) персоналу: 4 показники.
- Загальний аналіз: 1 показник.

Ми не будемо розглядати всі показники детально, оскільки більша частина з них буде відкинута в процесі відбору ознак і не буде включена до моделі. Проте короткий огляд усіх показників можна знайти в таблиці в Додатку А.

Варто відзначити, що в Україні не існує єдиного універсального підходу до аналізу ефективності та надійності (ризикованості) діяльності банківських установ. Отже, незважаючи на попередньо проведений ґрунтовний аналіз в досліджуваній області, етап відбору коефіцієнтів все ж таки має суб'єктивний характер.

Наступним кроком є відбір ознак. Застосування алгоритмів на всіх відібраних показниках не є ефективним з ряду причин. Серед основних можна виділити наступні:

- Модель побудована на великій кількості вхідних змінних може бути надто складною з точки зору подальшої економічної інтерпретації отриманих результатів [155].
- Модель побудована на великій кількості вхідних змінних може бути надто ресурсоємкою, робота моделі може займати надто багато часу, що частіше за все не буде виправдане якістю отриманих результатів.

- При застосуванні алгоритмів зниження розмірності поширеним явищем є так зване «прокляття розмірності». Для адекватного та ефективного перетворення багатовимірних простору у простір меншого виміру, наприклад у 3-х або 2-х вимірний простір (якщо мова йде про візуалізацію даних), необхідний набір даних з певної кількості спостережень. Суть «прокляття» полягає в тому, що чим більшою є кількість початкових вимірів (кожна ознака є одним виміром), тим більша кількість даних необхідна, і з кожною додатковою ознакою необхідний обсяг даних росте експоненційно. Інакше дані будуть надто розрідженими та алгоритми не будуть в змозі відтворити локальну та глобальну структуру первинного багатовимірних набору. Саме поняття було вперше запропоноване та описане Річардом Белманом в його роботах [156, 157].
- Моделі, що відносяться до машинного навчання, можуть показувати нижчу здатність до узагальнення прикладів внаслідок ефекту перенавчання у випадку коли через модель пропускається надто велика кількість даних [158].

Для уникнення цих та інших можливих негативних факторів застосовуються підходи відбору ознак. Їх суть власне полягає у виділенні певного під-набору з початкового набору ознак, такого що не містив би надлишкових чи незначущих елементів. Це, в свою чергу, має забезпечити високу ефективність роботи алгоритмів та адекватність економічної інтерпретації отриманих результатів.

Відбір ознак нами був проведений в декілька етапів. На першому етапі нами було проаналізовано всі відібрані показники та висунуто декілька **припущень та спрощень** щодо розрахованого набору показників:

- Нами було відзначено такі показники, що за своєю економічною суттю подібні до інших. Наприклад, маємо два коефіцієнти:
 - К.1.2. Питома вага дохідних активів у сукупних активах. Порядок розрахунку: Працюючі активи / Чисті активи всього.

- К.1.13. Відношення кредитного портфелю до активів. Порядок розрахунку: Кредитно-інвестиційний портфель (КІП) / Чисті активи всього.

Розрахунок даних показників очевидно відрізняється чисельником, проте розглянувши складові працюючих активів та кредитно-інвестиційного портфелю можна відзначити, що вони мають незначні відмінності, а отже розраховані показники будуть подібними та мати високу позитивну кореляцію. Такого роду коефіцієнти, що схожі за економічною суттю, нами буде відсіяно на наступних кроках.

- Деякі показники виявилися ідентичними за своєю суттю, наприклад, Коефіцієнт ефективності використання залучених коштів по доходам та Коефіцієнт загальної дохідності активів відрізняються тільки тим, що в першому в знаменнику стоять Пасиви, а у другому – Активи. Відомо, що ці дві величини є рівними в балансі підприємства, а отже і показники є тотожними. В таких випадках нами вилучались з набору показники, що рідше зустрічались у розглянутій літературі.
- В статистичних матеріалах НБУ за 2015 рік категорії доходу та витрат не подаються окремими агрегованими показниками, тому їх розрахунок був проведений на основі наявних в статистиці даних, наступним чином: Доходи всього = Процентні доходи + Комісійні доходи + Торговий результат + Інші операційні доходи; Витрати всього = Процентні витрати + Комісійні витрати + Відрахування до резервів + Адміністративні витрати.
- В статистичних матеріалах НБУ за I-III квартали 2015 року відсутній показник «Торговий результат», проте подаються дані по Результату від операцій з цінними паперами в торговому портфелі банку, що був використаний замість нього за вказаний період.
- Коефіцієнт частки неякісних активів в портфелі банку був вилучений з вибірки, так як починаючи з I кварталу 2017 року, НБУ припинили публікацію додаткових показників, необхідних для розрахунку даного

коефіцієнта, а саме інформації про обсяги за кредитними операціями різних класів якості. Додаткові складнощі виникають і при аналізі показника за період, де дані доступні, так як банки можуть свідомо відносити певні кредитні операції до сумнівних, з метою зменшення прибутку на суму резервів, та в результаті зменшення податку на прибуток. Інша ситуація можлива, коли банк фінансує пов'язаних осіб, і, у випадку неможливості повернення позики, повторно кредитує через «треті» банки або через випуск облігацій, і таким чином підтримує високу якість виданих кредитів.

- Починаючи з IV кварталу 2017 року, НБУ подає дані по всім категоріям витрат як додатні значення, в той час як за усі періоди до IV кварталу 2017 року значення витрат подавались від'ємними. З метою уніфікації даних і уникнення помилок при розрахунку коефіцієнтів, всі витрати, що надаються з вказаного періоду та використовуються нами для розрахунку коефіцієнтів – були використані з оберненим знаком.

По завершенню першого кроку у нас залишилось 48 з 63 раніше відібраних показників, 15 показників було відкинуто як повністю ідентичні до деяких інших коефіцієнтів.

Другим кроком є дослідження мультиколінеарності набору ознак. Мультиколінеарність можна визначити як явище, за якого існує висока лінійна залежність між факторам, пояснюючими змінними моделі. Мультиколінеарність вважається негативним явищем, так як може призвести до нестійкості, невизначеності. Регресійні моделі втрачають ефективність, так як ускладнюється інтерпретація їх параметрів. Оцінки параметрів стають ненадійними, виникають великі стандартні похибки, що змінюються при зміні обсягів спостережень. Все це робить модель непридатною до використання для побудови прогнозів.

Нами застосовуються алгоритми зниження розмірності та кластеризації. Якщо розглядати вплив мультиколінеарності на роботу таких алгоритмів, то загалом вважається що вплив корельованих змінних на результати роботи неістотний, так як алгоритми кластеризації та зниження розмірності не

ґрунтуються на лінійних припущеннях. Проте, в основі даних алгоритмів лежить розрахунок відстаней між точками даних. Наявність мультиколінеарності у вхідній вибірці може вплинути на адекватність відображення багатовимірних даних на двовимірній площині через викривлення розрахунку відстаней між точками. Деякі точки можуть розташовуватись у двовимірному просторі значно ближче одна до одної за рахунок впливу корельованих змінних. Це не є критичним недоліком, так як такий результат буде закономірним і матиме адекватне економічне пояснення: долучаючи корельовані змінні до моделі ми встановлюємо схильність моделі надавати більшої ваги певній категорії коефіцієнтів. Проте, нами було вирішено вилучити більшу частину таких змінних для забезпечення об'єктивності отриманих результатів.

Для аналізу мультиколінеарності нами була побудована теплова карта на основі коефіцієнта кореляції Пірсона, який розраховується наступним чином:

$$r_{xy} = \frac{(\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^m (y_i - \bar{y})^2}} = \frac{cov(x, y)}{\sqrt{s_x^2 s_y^2}} \quad (2.1)$$

де $\bar{x}, \bar{y}$ – вибіркові середні x_m та y_m ,

s_x^2, s_y^2 – вибіркові дисперсії,

$r_{xy} \in [-1, 1]$.

Сама теплова карта парної кореляції ознак відображена на Рис. 2.2.

Як можна побачити на Рис. 2.2., досить багато пар ознак мають високу кореляцію між собою – чим вищою є кореляція, тим темнішими є ділянки на карті. Перелік всіх пар показників з істотною кореляцією та самі значення коефіцієнту кореляції цих пар можна переглянути у Додатку Б.

Відомо, що мультиколінеарність виключно складно і часто неможливо виправити змінюючи самі дані. Рішення даної проблеми однак лежить у відсіюванні частини ознак, що і є наступним та останнім етапом. Ми проводили відбір ознак опираючись аналітичний та експертний підходи.

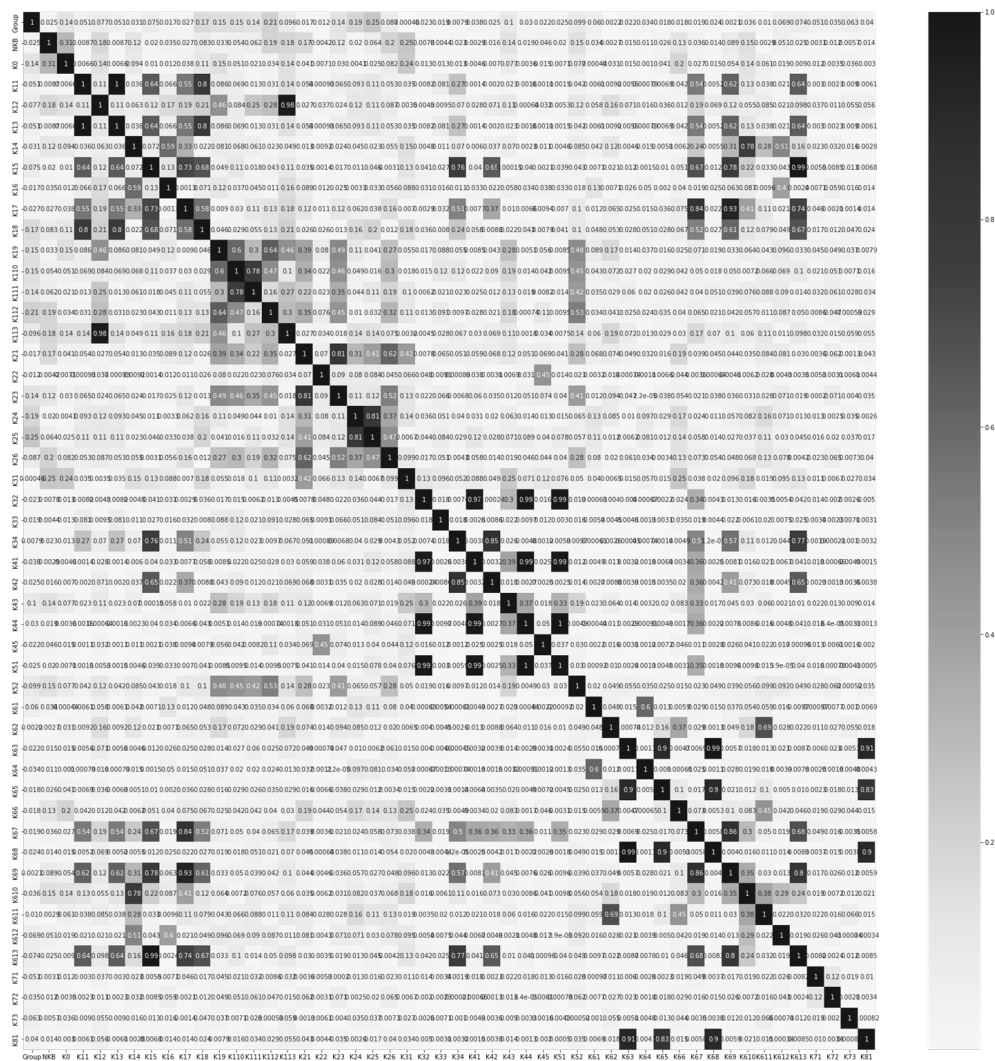


Рис. 2.2. Теплова карта парної кореляції ознак – показників банківської діяльності до проведення відбору ознак

Джерело: розроблено автором на основі статистичних даних НБУ

Аналітичний: за допомогою парної кореляції визначали залежні пари ознак, за граничне значення для коефіцієнту кореляції було взято 0.5, так як широко розповсюдженим є підхід, що визначає рівень кореляції як «високий» за показників вище цього порогу. На основі метрики фактору росту дисперсії (Variance Inflation Factor або VIF) обирається ознака з найменшим значенням.

VIF є одним з методів виміру колінеарності факторів, і розраховується наступним чином:

$$VIF = \frac{1}{1 - R^2} \quad (2.2)$$

де R^2 – коефіцієнт детермінації.

При відборі показників оцінки ми опирались на роботу Дормана та співавторів [159], в якій показники кореляції Пірсона та VIF розглядаються як інструменти визначення кореляції між факторами, та зазначається порогове значення для останнього як 10 та вище.

В якості додаткових перевірок використовувалися метрики середньої абсолютної різниці (Mean Absolute Difference – MAD) та спектрального відбору ознак (Spectral Feature Selection – SPEC), в тих випадках, коли VIF для двох ознак суттєво не відрізнявся (різниця < 1).

MAD детальніше розглядається в роботі Ферейри та Фігуїредо [160]. Розраховується наступним чином:

$$MAD_i = \frac{1}{n} \sum_{j=1}^n |X_{ij} - \bar{X}_i|, \quad (2.3)$$

де X_{ij} – окреме спостереження,

$\bar{X}_i$ – середнє значення по всіх спостереженнях,

n – кількість спостережень.

Чим вищим є значення метрики MAD для певної ознаки, тим вищою вважається роздільна здатність цієї ознаки.

Алгоритм SPEC проводить оцінку наскільки підходить ознака з точки зору її узгодженості зі структурою графа, отриманого на основі схожості між об'єктами. Алгоритм є відносно складним і має непросту математичну базу в своїй основі. Ми не будемо детально її розглядати в межах даної роботи, вся математична база подана в роботі авторів алгоритму [161]. Алгоритм виконується в три кроки: побудова матриці схожості об'єктів та її графове представлення; оцінка ознак з використанням системи власних значень графа через вимірювання узгодженості між кожною ознакою та нетривіальними власними векторами матриці Лапласа; ранжування ознак у спадаючому порядку з точки зору їх відповідності та узгодженості. Загалом, алгоритм SPEC є узагальненням більш відомого підходу Лапласової оцінки. Чим більшим є значення SPEC тим вищою є цінність ознаки для всього набору.

Розглянемо приклад відбору показників. Маємо два коефіцієнти K4.4. Відношення кредитного портфелю до зобов'язань та K.5.1. Коефіцієнт миттєвої ліквідності. Їх коефіцієнт кореляції складає 0,9965. Так як даний показник є вищим за встановлений нами поріг в 0,5 нам необхідно проаналізувати обидва показники і відсіяти один з них. Розглянемо значення наших метрик для обох коефіцієнтів.

K.4.4. $VIF = 77,71$, $MAD = 0,4535$, $SPEC = 0,37$.

K.5.1. $VIF = 11,51$, $MAD = 0,4537$, $SPEC = 0,56$.

Основною метрикою, на яку ми орієнтуємось, є VIF і він показує перевагу коефіцієнту K.4.4. Метрика MAD підтверджує його дискримінаційну здатність, хоч різниця і незначна. SPEC в даному випадку свідчить про значущість K.5.1, однак спираючись на загальну оцінку по всім метрикам ми вилучаємо коефіцієнт K.5.1 з вибірки. Усі інші пари показників, що мають кореляцію $> 0,5$ ми опрацьовуємо в рамках ітеративного процесу відбору ознак.

Дані по всім метрикам по ознакам з початкового набору показників представлені в таблиці в Додатку А.

В результаті, за використання даних метрик був отриманий базовий набір показників для моделі. Він складається з 22 ознак, 27 ознак було відкинуто в процесі відбору.

Експертний підхід: на основі базового набору ми додатково провели ряд експериментів, додаючи нові ознаки, які, на нашу думку, є важливими для включення в модель оцінки банківської ефективності, і виключали ознаки, які, на нашу думку, не є ключовими для оцінки ефективності роботи банку. Після кожного включення або виключення ознаки ми проводили зниження розмірності, кластеризацію набору даних та експертний візуальний аналіз адекватності розміщення банків на двовимірній карті. Ми перевіряли чи мають банки, що потрапляють в одну групу схожі характеристики, чи схожі в банків що знаходяться поруч оцінки, що надані іншими експертами (ми спирались на кредитні рейтинги національних рейтингових агентств), також до уваги приймалась оцінка якості кластеризації. В результаті, деякі ознаки було вирішено залишити в моделі,

незважаючи на значення метрик відбору ознак, а деякі ознаки було виключено з моделі так як вони не мали суттєвого впливу на результати моделювання.

Метою застосування експертного підходу для відбору показників в модель була максимізація адекватності розміщення точок-банків на двовимірній проекції та їх групування в кластери. На жаль, доступні аналітичні методи, такі як наприклад розглянуті вище метрики, не можуть надати нам оцінки рівня якості розподілу банків на кластери з точки зору економічної інтерпретації результатів такого розподілу. Отримавши набір даних, що забезпечував би найвищу якість розподілу, важливо перевірити та переконатись, що такий набір даних не погіршує загальну якість моделі з суто аналітичної точки зору.

Розглянемо Рис. 2.3, на якому зображено двовимірну проекцію, отримана після застосування алгоритму зниження розмірності UMAP на двох наборах ознак.

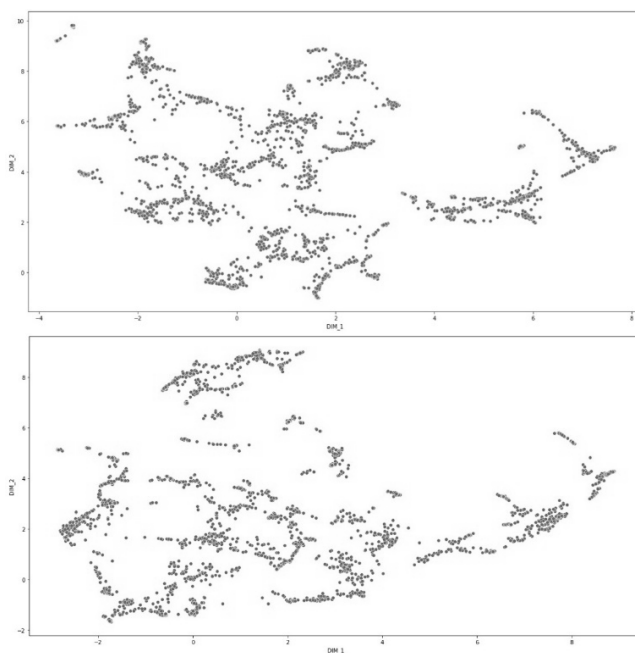


Рис. 2.3. Двовимірні проекції UMAP, при застосуванні на наборі ознак, відібраних суто аналітичним шляхом (зверху), та при застосуванні на наборі ознак відібраних аналітичним та експертним шляхом (знизу).

Джерело: розроблено автором на основі статистичних даних НБУ

Для порівняльного аналізу ми також використали алгоритм агломеративної ієрархічної кластеризації та метрику силуету, оцінки Девіса-Боулдіна та Калінського-Харабаза. Даний алгоритм кластеризації потребує задання бажаної кількості кластерів на вхід моделі як один із гіперпараметрів. В даному випадку

наша ціль - це проведення простого порівняльного експерименту, тому нами було розраховане середнє значення силуету для ряду кластеризацій з кількістю отримуваних кластерів від 8 до 15 (в такому проміжку, на нашу думку, має лежати оптимальна кількість кластерів, щоб якість інтерпретації отриманих результатів могла бути забезпечена на високому рівні). Результати проведеного експерименту наступні:

- Для набору ознак, відібраного тільки аналітичним методом, середнє значення силуету складає 0,4854, оцінок Девіса-Боулдіна та Калінського-Харабаза відповідно 0,72 та 3969.
- Для набору ознак, відібраного комплексно за допомогою аналітичного та експертного підходів, середнє значення силуету складає 0,4730, оцінок Девіса-Боулдіна та Калінського-Харабаза відповідно 0,69 та 4564.

Метрики Девіса-Боулдіна та Калінські-Харабаза вказують на перевагу комплексного підходу, метрика силуету незначно вища для аналітичного. В цілому, на основі метрик якості кластеризації, так і візуального аналізу отриманих проекцій, можна зробити висновок, що застосування комплексного підходу до відбору ознак не має суттєвого впливу на загальну аналітичну якість моделювання, і навіть трохи підвищує її. При цьому, набір даних, отриманий із застосуванням експертної компоненти, гарантує значно адекватніші результати з точки зору розташування банків у кластерах та, відповідно, з точки зору економічної інтерпретації отриманих результатів проектування та кластеризації.

Більш детально алгоритми кластеризації та метрики оцінки якості кластеризації будуть розглянуті в підрозділі 2.3.

Всього після експертного підходу в моделі залишилось 17 ознак, їх повний перелік подано в Додатку В.

В результаті відбору ознак рівень кореляції було знижено. На Рис. 2.4 зображено теплову карту кореляції пар тих ознак, що залишились після відсіювання.

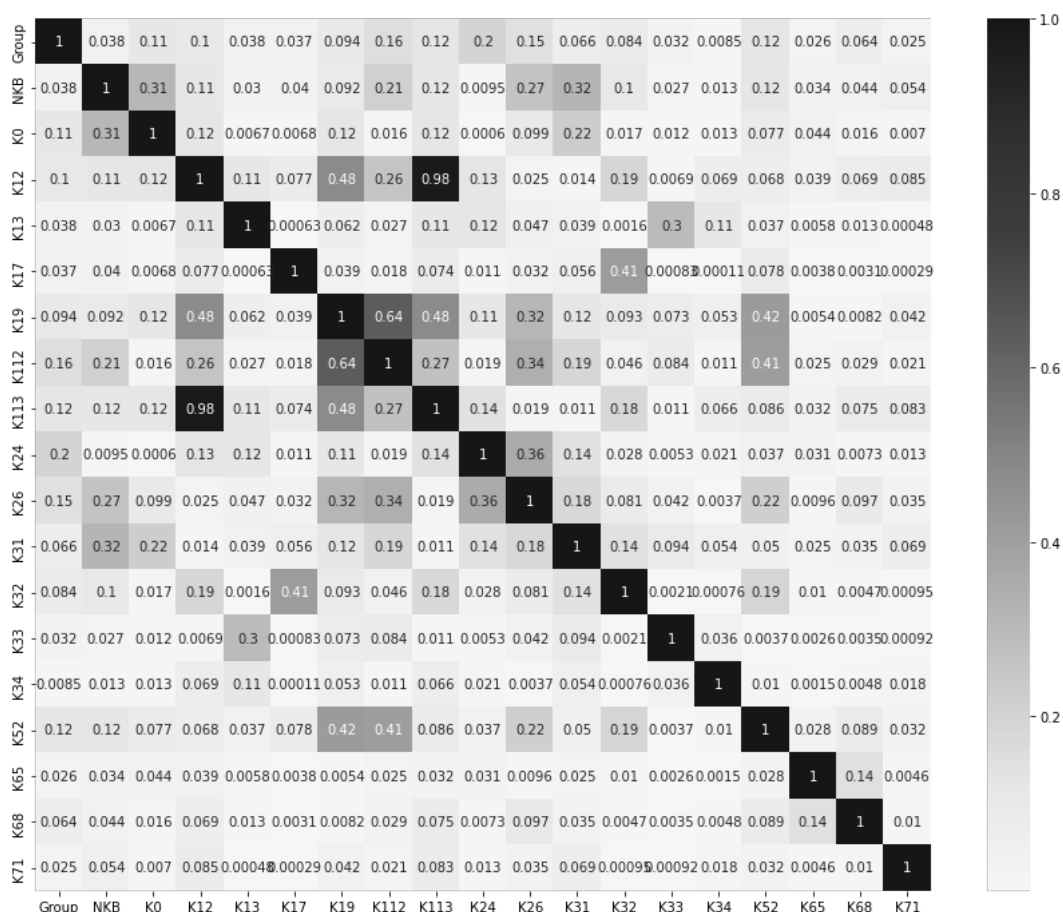


Рис. 2.4. Теплова карта парної кореляції ознак – показників банківської діяльності після проведення відбору ознак

Джерело: розроблено автором на основі статистичних даних НБУ

На даному етапі нами було зроблено додаткові **припущення та спрощення**. По-перше, як видно з Рис. 2.4, ми зберегли декілька пар ознак, незважаючи на їх кореляцію $> 0,5$. Було прийнято рішення залишити їх в моделі, так як дані ознаки позитивно впливають на адекватність подальшої інтерпретації результатів моделювання та якість розділення даних у двовимірному просторі. В даному випадку ці пари ознак можемо розглядати як одну ознаку з підсиленою вагою в моделі. В результаті відбору ознак середнє значення VIF по всім даним зменшилось з 1137,35 до 61,18. Подальше зменшення кількості ознак в кращому випадку не вплинуло на якість кластеризації і адекватність інтерпретації результатів, в гіршому – склало негативний вплив.

По-друге, для двох коефіцієнтів, а саме для K5.2 «Частка ліквідних активів» та K6.8 «Рентабельність витрат» було застосовано заміну від'ємних значень на -1. Від'ємні значення таких показників як прибуток та високоліквідні активи (що

застосовують при розрахунку вказаних коефіцієнтів) можуть бути індикаторами несприятливих процесів в рамках банківських установ. Така заміна значень була проведена з метою потенційного групування алгоритмами зниження розмірності та кластеризації банків, що є неліквідними та/або збитковими.

2.2. Модель аналізу фінансового стану комерційних банків на підґрунті алгоритмів зниження розмірності

Для проведення подальшого дослідження фінансового стану комерційних банків доцільно застосувати такі алгоритми машинного навчання як зниження розмірності, що обумовлене рядом причин. Виділимо ключові з них:

- Кластеризація багатовимірних даних важко інтерпретується з економічної точки зору, не може бути візуалізована.
- Зі зростанням кількості вимірів зростає кількість підпросторів даних, виникає явище «прокляття розмірності».
- Зі зростанням кількості вимірів викривлюється відстань між точками, що в свою чергу може призвести до викривлення самих результатів.
- Чим вище є розмірність даних, тим вищим є імовірність, що утворяться так звані «локальні кластери», що будуть релевантні певному підпростору даних проте з глобальної точки зору не будуть мати ніякого сенсу. Тобто такі кластери не можна буде обґрунтувати або пояснити [162].

Для вирішення проблеми кластеризації багатовимірних даних існують певні підходи, наприклад кластеризація під-просторів, кореляційна кластеризація та проекційна кластеризація. Зниження розмірності є одним з простих рішень даної проблеми, та на нашу думку, такий підхід найбільше підходить в рамках нашого дослідження.

Формально задачу зниження розмірності можна описати наступним чином. Нехай є деяка вибірка об'єктів $X = \{x_n\}_{n=1}^N, x_n \in R^D$. Задача зменшення розмірності полягає в тому, щоб отримати представлення даної вибірки в меншій

розмірності $T = \{t_n\}_{n=1}^N, t_n \in R^d$. Тут $d \ll D$, проте в деяких окремих випадках d може співпадати з D .

Загалом зниження розмірності можна визначити як перетворення даних з високо-розмірного простору у низько-розмірний простір таким чином, щоб низько-розмірне представлення зберігало деякі з суттєвих властивостей оригінальних даних. При цьому бажано, щоб розмірність проєкції була наближена до так званої внутрішньої розмірності даних – такої розмірності що досягається при мінімальній кількості змінних, що є достатньою для адекватного представлення оригінальних даних в низько-розмірному просторі.

На сьогоднішній день існує досить широкий спектр алгоритмів зниження розмірності. Основні та найбільш поширені у застосуванні були розглянуті нами в Розділі 1. Нами було вирішено обрати два алгоритми, кожен з яких матиме свою задачу в подальшому дослідженні: алгоритм стохастичного вкладення сусідів з t -розподілом (t -distributed Stochastic Neighbor Embedding, t -SNE) та алгоритм апроксимації та проєкції рівномірного різноманіття (Uniform Manifold Approximation and Projection, UMAP). Розглянемо їх більш детально.

Алгоритм t -SNE бере свій початок від свого попередника, алгоритму SNE, тому має сенс розпочати ознайомлення з нього. Задача яку має виконувати алгоритм полягає в наступному: маємо набір даних з точками, що описані багатовимірною змінною, розмірність якою значно вище 3-ох; необхідно отримати нову змінну, яка б найповніше зберігала структуру та логіку впорядкування вихідних даних. Першим кроком алгоритму SNE є перетворення евклідових відстаней між точками на умовні імовірності, що відображають властивості точок:

$$p_{j|i} = \frac{\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}\right)} \quad (2.4)$$

де x_i та x_j – дві точки даних,

x_k – інші точки даних в наборі даних,

σ – середньоквадратичне відхилення.

Формула 2.4 вказує наскільки точка x_i близька до точки x_j при гаусовому розподілі навколо x_i з заданим відхиленням σ , що є різним для кожної точки. Воно обирається таким чином, що точки в областях з більшою щільністю мають меншу дисперсію. Для цього використовується оцінка перплексії:

$$Perp(P_i) = 2^{H(P_i)} \quad (2.5)$$

де $H(P_i)$ – ентропія Шенона в бітах:

$$H(P_i) = - \sum_j p_{j|i} \log_2 p_{j|i} \quad (2.6)$$

Перплексія може бути визначена як згладжена оцінка ефективної кількості сусідів для точки X_i . Це один із гіперпараметрів алгоритму. Авторами рекомендоване задання значення перплексії в межах від 5 до 50 [164]. σ визначається для кожної пари точок за допомогою алгоритму бінарного пошуку.

Нехай y_i та y_j – це точки у просторі нижчої розмірності. Тоді імовірність по аналогії з формулою 2.4 розраховується наступним чином (авторами алгоритму встановлюється сігма, що рівна $\frac{1}{\sqrt{2}}$):

$$q_{i|j} = (\exp(-\|y_i - y_j\|^2) / (\sum_{k \neq i} \exp(-\|y_i - y_k\|^2))) \quad (2.7)$$

У випадку коли проекції точок y_i та y_j адекватно моделюють подібність між вихідними точками у просторі високої розмірності x_i та x_j то відповідні умовні імовірності $p_{i|j}$ та $q_{i|j}$ будуть рівними. Для оцінки якості відображення $p_{i|j}$ в $q_{i|j}$ застосовується дивергенція або відстань Кульбака-Лейблера. SNE мінімізує суму цих відстаней для всіх точок проекцій через застосування градієнтного спуску. Функція втрат визначається наступним чином:

$$Cost = \sum_i KL(P_i || Q_i) = \sum_i \sum_j p_{j|i} \log(p_{j|i} / q_{j|i}) \quad (2.8)$$

Градієнт виглядає наступним чином:

$$\frac{\partial Cost}{\partial y_i} = 2 \sum_j (p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j})(y_i - y_j) \quad (2.9)$$

Авторами алгоритму була запропонована фізична аналогія процесу оптимізації для кращого розуміння. Всі точки з'єднані пружинами. Жорсткість пружини, що з'єднує дві точки залежить від різності між подібністю двох точок в багатовимірному просторі та двох точок у проекційному просторі. Градієнт виступає результуючою силою, що діє на точку в проекційному просторі. Якщо систему «вивільнити», то через певний час вона дійде рівноваги, що і буде представляти шуканий розподіл. Алгоритмічно пошук рівноваги описується з урахуванням моментів:

$$y^t = y^{t-1} + \eta \left(\frac{\partial Cost}{\partial Y} \right) + \alpha(t)(y^{(t-1)} - y^{t-2}) \quad (2.10)$$

де η – швидкість навчання, або довжина кроку навчання,

α – коефіцієнт інерції.

Однак, використання класичного SNE хоча і демонструє непогані результати, проте має слабкі сторони пов'язані з складнощами в оптимізації функції втрат та з так званою «проблемою скупченості». Модифікація t-SNE дозволяє частково подолати ці слабкі сторони.

Функція втрат t-SNE має дві ключові відмінності від SNE. Перша: у t-SNE симетрична форма подібності в багатовимірному просторі та простіший варіант градієнту. Друга: замість гаусового розподілу для точок в проекційному просторі використовується t-розподіл Стюдента, його «важкі» хвости полегшують процес оптимізації і вирішують проблему скупченості. Графічне відображення відмінностей нормального розподілу та розподілу Стюдента показано на Рис. 2.5. Розглянемо вказані вище особливості детальніше.

В t-SNE пропонується мінімізувати одиничну дивергенцію між спільною імовірністю P в багатовимірному просторі з спільною імовірністю Q в проекційному просторі (як альтернатива мінімізації суми дивергенцій Кульбака-Лейблера):

$$Cost = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (2.11)$$

де p_{ii} та $q_{ii} = 0$, $p_{ij} = p_{ji}$, $q_{ij} = q_{ji}$ для будь-яких i та j , а p_{ij} визначаються так:

$$p_{ij} = \frac{(p_{j|i} + p_{i|j})}{2n} \quad (2.12)$$

де n – кількість точок в наборі даних.

Гradient для симетричного SNE описується простіше, і виглядає наступним чином:

$$\frac{\partial Cost}{\partial y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j) \quad (2.13)$$

Проблема скупченості полягає в тому, що відстань між двома точками в проекційному просторі, такими що відповідають двом середньо-віддаленим точкам у багатовимірному просторі, має бути суттєво більшою ніж відстань, що отримується гаусовим розподілом. Дану проблему вирішують хвости розподілу Стюдента. У t-SNE використовується t-розподіл з одним ступенем свободи.

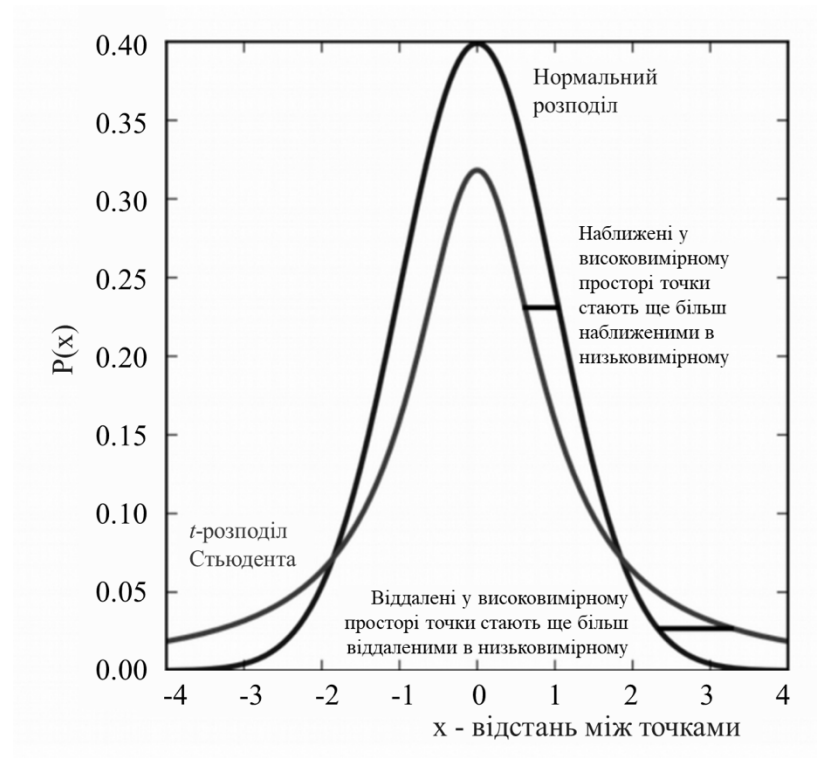


Рис. 2.5. Візуальне порівняння нормального розподілу та розподілу Стюдента

Джерело: [163]

Спільна імовірність для проекційного простору в такому випадку визначається наступним чином:

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}} \quad (2.14)$$

Відповідний градієнт буде виглядати наступним чином:

$$\frac{\partial Cost}{\partial y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j) (1 + \|y_i - y_j\|^2)^{-1} \quad (2.15)$$

В контексті фізичної аналогії, що згадувалась раніше, результуюча сила, що визначається формулою 2.15, буде суттєво притягувати точки проєкційного простору для близько-лежачих точок багатовимірного простору, і відштовхувати точки проєкційного простору для віддалених точок багатовимірного простору.

В алгоритмі t-SNE можна виділити 3 основних гіперпараметри: перплексія (*perplexity*), кількість ітерацій (*number of iterations*), перебільшення (*exaggeration* або *early exaggeration*).

Перплексія, параметр що вже згадувався вище, є ключовим в даному алгоритмі. Простими словами його можна описати як балансувальну силу, що визначає на що буде сфокусовано роботу алгоритму: на збереження локальної чи глобальної структури початкових даних. Інакше перплексію можна розглядати як параметр, що визначає кількість сусідів, для яких t-SNE зберігатиме відстані. Зміна перплексії в меншу сторону буде зміщувати фокус алгоритму на локальну структуру даних, а збільшення – на глобальну. Для утримання балансу між глобальною та локальною структурою даних, зі збільшенням обсягу даних перплексія також має збільшуватись. За замовчанням параметр є рівним 30.

Зміну двовимірної проєкції двох кластерів точок згенерованих випадковим чином при зміні параметру перплексії можна побачити на Рис. 2.6.

Як видно з Рис. 2.6 зі збільшенням параметру перплексії алгоритм більш якісно відтворює локальну структуру початкового набору даних, при значенні параметра 50 він досить точно повторює оригінальну структуру, однак при надто низьких чи надто високих значеннях параметра двовимірна проєкція не матиме сенсу.



Рис. 2.6. Зміна двовимірної проекції двох кластерів випадково згенерованих точок при зміні параметру перплексії. Зліва направо: оригінальні дані, оброблені алгоритмом дані при перплексії 2, 5, 30, 50 та 100. Кількість ітерацій для всіх тестів була встановлена на 50,000.

Джерело: [164]

Швидкість навчання визначає розмір кроків оновлення градієнту. Стандартне значення встановлене на рівні 200, зазвичай значення швидкості навчання може змінюватись від 100 до 1000. Надто високе значення параметру може призвести до того, що дані будуть схожі на «шар», усі точки будуть рівновіддалені від своїх найближчих сусідів. В оберненому випадку, якщо швидкість навчання встановлено дуже низькою, дані можуть виглядати як щільна «хмара», де всі точки знаходяться близько однак до одної і викидів майже немає. Збільшення швидкості навчання може допомогти у випадку, коли алгоритм потрапляє у небажаний локальний мінімум. Також, збільшення параметру рекомендоване у випадку зростання обсягу вхідної вибірки даних.

Перебільшення – параметр, що контролює відображення кластерів точок у проектованому просторі в порівнянні до багатовимірного простору, те як сильно будуть віддалені кластери в проектованому просторі в порівнянні з віддаленістю кластерів в багатовимірному просторі. Більші значення параметру будуть збільшувати відстань між кластерами у проектованому просторі відносно початкового простору. В більшості випадків даний параметр не оптимізується, та залишається за замовчуванням на рівні 12.

Алгоритм представлений і описаний в роботі автора ван дер Маатена [165].

t-SNE має також дві модифікації, що застосовувались нами при його застосуванні: це Barnes-Hut t-SNE та ініціалізація через PCA.

Відомо, що алгоритм t-SNE є досить повільним і зі зростанням обсягу вхідних даних складність алгоритму зростає квадратично, таким чином кажуть, що

алгоритм виконується за $O(N^2)$. Barnes-Hut t-SNE відзначається тим, що надає можливість алгоритму виконуватись за $O(N \log N)$, тобто за «логарифмічний час», що на порядок пришвидшує алгоритм при збільшенні обсягів даних. Дана модифікація побудована на варіації досить відомого алгоритму Барнса-Хата, що застосовується в області астрономії для проведення N-вимірних симуляцій, з метою апроксимації сил між відповідними точками в проекції. В її основі лежать так звані «дерева блукаючих точок», що застосовуються для обчислення розріджених попарних подібностей між об'єктами вхідних даних. Barnes-Hut t-SNE має в основі відносно складну та об'ємну математичну базу, тому не будемо зупинятись на її детальному викладенні, детальний огляд модифікації подано в роботі Лоуренса ван дер Маатена [166]. Результати проведених в роботі експериментів показали значний приріст обчислювальної потужності алгоритму, що дозволяє застосовувати його для проектування мільйонів об'єктів.

Ініціалізація через PCA в своїй основі є досить простою модифікацією. Стандартний t-SNE ініціалізується малими випадковими значеннями згенерованими за гаусовим розподілом. У випадку модифікації з PCA для ініціалізації використовуються перші два головних компонента, отриманих після застосування PCA на вхідних даних, стандартизованих таким чином, щоб мати таке саме стандартне відхилення, що і гаусовий розподіл у ініціалізації випадковими значеннями. Такий підхід дозволяє уникнути великих початкових відстаней, що може трапитись якщо дані не стандартизовані, що в свою чергу може призвести до передчасної збіжності градієнту.

Алгоритм UMAP є більш сучасним, та має більш глибоке математичне обґрунтування в порівнянні з t-SNE. Роботу самого алгоритму можна умовно розділити на декілька етапів.

На першому етапі на вхід алгоритму подається вибірка з n об'єктів: $X = \{x_1, \dots, x_n\}$. UMAP розраховує відстань між об'єктами за заданою метрикою (за замовчанням використовується класична евклідова відстань) і для кожного об'єкту x_i визначається перелік k найближчих до нього сусідів: $T = \{t_1, \dots, t_k\}$.

Крім цього, для кожного об'єкту розраховується відстань до його найближчого сусіда (забезпечуючи в результаті локальну зв'язність багатовиду):

$$\rho_i = \min_{t \in T} d(x_i, t). \quad (2.16)$$

Власне однією з ключових відмінностей UMAP від t-SNE є застосування підходу «найближчого сусідства», що є аналогом перплексії.

Далі розраховується величина σ_i , що задана наступним рівнянням:

$$\sum_{t \in T} \exp\left(-\frac{d(x_i, t) - \rho_i}{\sigma_i}\right) = \log_2 k. \quad (2.17)$$

Це експоненційний розподіл імовірностей у високовимірному просторі. Важливо відмітити, що на відміну від t-SNE імовірності не нормалізуються, однак специфіка алгоритму призводить до того, що імовірності високо- і низько-розмірних розподілів і так будуть знаходитись на проміжку від 0 до 1. Відсутність стандартизації також пришвидшує алгоритм.

На наступному кроці алгоритм виконує побудову зваженого орієнтованого графа, у якому ребра з'єднують кожний об'єкт з його сусідами. Вага ребра від x_i об'єкта до його t_j сусіда розраховується так:

$$\omega(x_i \rightarrow t_j) = \exp\left(-\frac{d(x_i, t) - \rho_i}{\sigma_i}\right). \quad (2.18)$$

Візуально у спрощеному двовимірному випадку отриманий граф мав би вигляд як зображено на Рис. 2.7.

На Рис. 2.7 в верхній частині зображено довільний двовимірний набір даних, сформованих виключно для демонстраційних цілей. В нижній частині зображено візуальне представлення сформованого алгоритмом UMAP зваженого орієнтованого графу.

Отримана раніше величина σ_i нормує суму ваг для кожного об'єкта до заданого числа $\log_2 k$.

Специфікою алгоритму є те, що для розрахунку відстаней між точками кожній з них призначається функція відстані, що не є сумісною з функціями інших точок. На практиці це означає, що відстані між точками є асиметричними, тобто відстань від графу А до графу Б не рівна відстані від графу Б до графу А. На графі

це мало б вигляд як на Рис. 2.7 (нижня частина рисунку): між двома точками існує два ребра різної довжини, так як довжина розраховується за різними метриками.

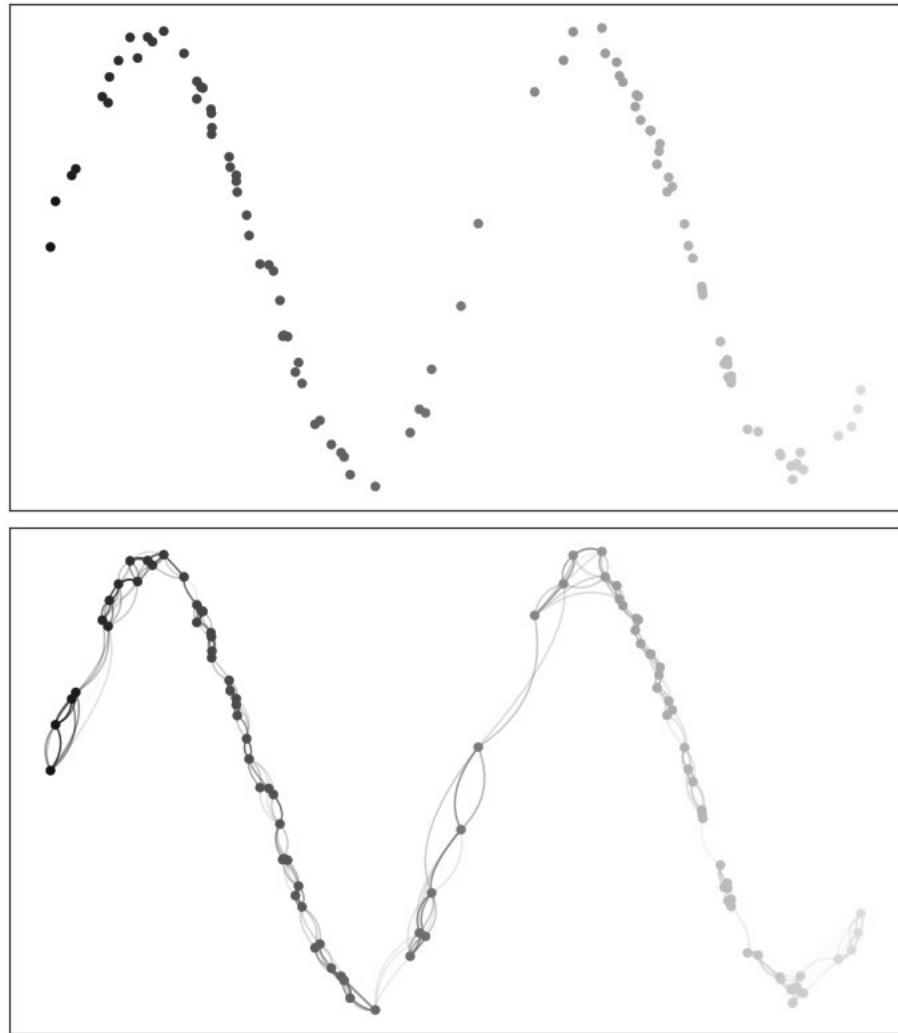


Рис. 2.7. Набір довільних двовимірних даних (знизу) та зважений орієнтований граф сформований UMAP (зверху)

Джерело: [167]

Вага ребра інтерпретується як імовірність існування даного ребра від одного об'єкта до іншого. Виходячи з цього, ребра між двома вершинами об'єднуються в одне з вагою рівною імовірності існування хоча б одного ребра (симетризація високо-розмірних імовірностей):

$$\omega(x_i, x_j) = \omega(x_i \rightarrow x_j) + \omega(x_j \rightarrow x_i) - \omega(x_i \rightarrow x_j) * \omega(x_j \rightarrow x_i) \quad (2.19)$$

Таким чином отримується зважений неорієнтовний граф, що на прикладі демонстраційних даних, представлених на Рис. 2.7, матиме вигляд, представлений

на Рис. 2.8. Як видно на Рис. 2.8 всі вершини тепер з'єднані між собою одним ребром.

Множина ребр E такого графа є нечіткою множиною з випадкових величин Бернуллі. Алгоритм створює новий граф в низько-розмірному просторі і проводить наближення множини ребр до вихідного простору. Для моделювання відстаней в проекційному низько-розмірному просторі застосовується сімейство прямих типу $1/(1 + a * y^{2b})$, що дуже схоже на t-розподіл в t-SNE, проте має відмінності:

$$q_{ij} = \left(1 + a(y_i - y_j)^{2b}\right)^{-1} \quad (2.20)$$

де a та b – параметри, що визначаються через нелінійну підгонку кускової функції за методом найменших квадратів, з урахуванням гіпер-параметру алгоритму найменшої відстані (min_dist):

$$\left(1 + a y_i - y_j\right)^{2b-1} \approx \begin{cases} 1 & \text{if } y_i - y_j \leq min_dist \\ e^{-(y_i - y_j) - min_dist} & \text{if } y_i - y_j > min_dist \end{cases} \quad (2.21)$$

В якості функції втрат використовується не дивергенція Кульбака-Лейблара а бінарна крос-ентропія, проводиться мінімізація даної функції:

$$\sum_{e \in E} \omega_h(e) * \log \frac{\omega_h(e)}{\omega_l(e)} + (1 - \omega_h(e)) * \log \left(\frac{1 - \omega_h(e)}{1 - \omega_l(e)} \right) \rightarrow \lim_{\omega_l} min \quad (2.22)$$

де $\omega_h(e)$ – функція належності нечіткої множини із ребр у багатовимірному просторі,

$\omega_l(e)$ – функція належності нечіткої множини із ребр у низько-вимірному просторі.

UMAP вирішує задачу мінімізації за допомогою стохастичного градієнтного спуску. Отримана множина з ребр визначає нове розміщення об'єктів і, відповідно, низько-розмірне відображення вихідного простору.

Алгоритм UMAP має ще дві ключові відмінності від t-SNE:

- UMAP ініціалізується не випадковим чином, а через лапласовий граф – проводиться побудова графу (*knn graph*, граф найближчих сусідів) і його формалізація через побудову матриці Лапласа за допомогою спеціальних матриць суміжності та ступневості (*adjacency & degree matrices*); матриця

факторизується і таким чином розв’язується задача розкладу власних значень. Це не має суттєвого впливу на результат роботи алгоритму, проте проєкції отримані за допомогою UMAP є більш стійкими, тобто не змінюються після повторного виконання.

- UMAP застосовує стохастичний градієнтний спуск, а не звичайний градієнтний спуск як t-SNE, що позитивно впливає на швидкість та ресурсо-ефективність алгоритму.

Більш глибокий порівняльний аналіз двох алгоритмів розглянуто у статті Осколкова Н. [163].

UMAP має 2 основні гіперпараметри: це кількість сусідів (*number of neighbors*) та найменша відстань (*minimal distance*).

Кількість сусідів – параметр, що визначає фокус алгоритму на локальну чи глобальну структуру даних, як перплексія для t-SNE. Це реалізується через регулювання та обмеження околу локального сусідства, в якому UMAP шукатиме точки з метою вивчення структури даних. Отже, встановлення менших значень призведуть до втрати глобального бачення структури даних, а встановлення надто високих значень призведуть до втрати деталей локальної структури. За замовчанням параметр має значення 15.

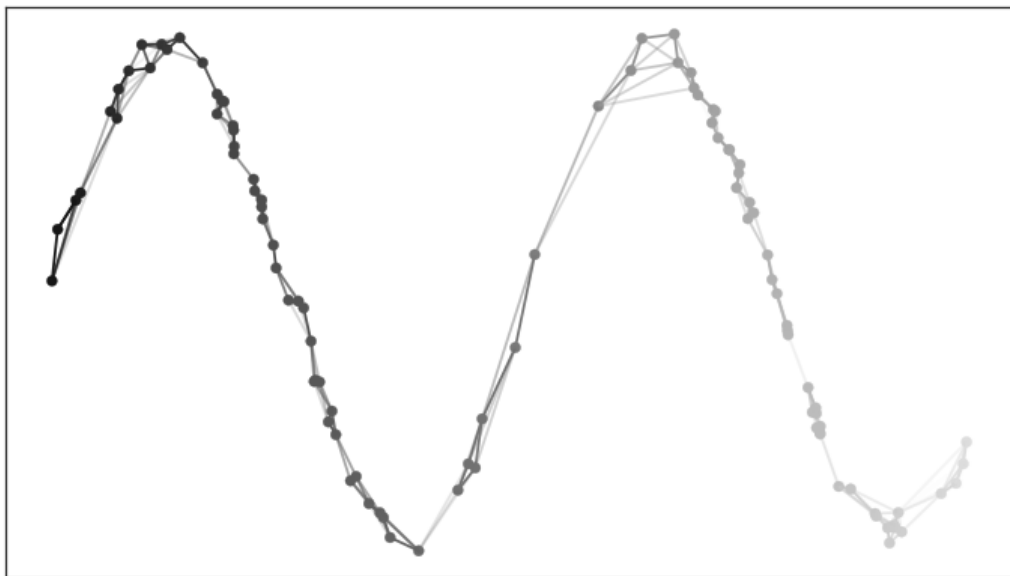


Рис. 2.8. Зважений неорієнтований граф з об’єднаними вершинами, побудований на прикладі демонстраційних даних

Джерело: [167]

Найменша відстань – це параметр, що визначає як щільно алгоритм може розміщувати точки на проекції. Він буквально визначає мінімальну відстань між точками, таку, що дозволена для розміщення точок у проектованому просторі. Нижчі значення параметру призведуть до отримання щільно згрупованих точок, а вищі значення – до більшого розпорошення точок, що в свою чергу дасть більше інформації саме про глобальну структуру початкових даних. За замовчанням параметр встановлено на рівні 0,1 та він може змінюватись в проміжку від 0,0 до 0,99.

Зміну двовимірної проекції довільно згенерованих даних при зміні розглянутих вище параметрів можна побачити на Рис. 2.9.

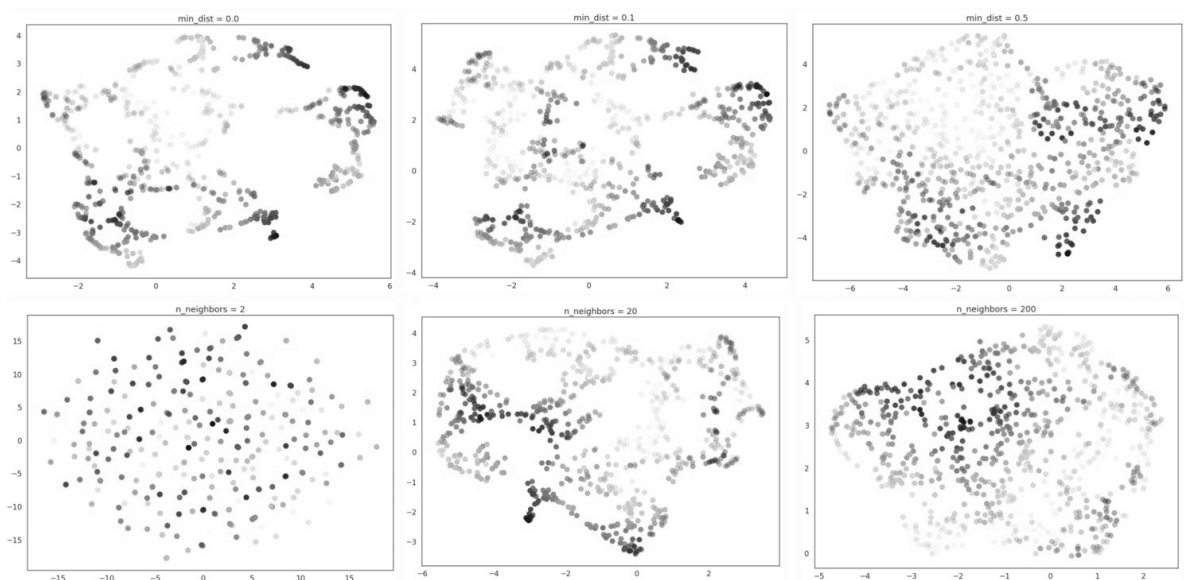


Рис. 2.9. Верхній ряд, зліва направо: зміна двовимірного відображення довільно згенерованих даних з параметрами мінімальної відстані = 0,0, 0,1 та 0,5; нижній ряд: з параметрами кількості сусідів = 2, 20 та 200.

Джерело: [167]

На Рис. 2.9 видно, що зі збільшенням параметрів *min_dist* та *n_neighbours* алгоритм втрачає здатність відтворювати локальні структури та проектує дані в певну однорідну глобальну структуру, що зберігатиме характеристики всього набору даних, проте не зможе відтворити певні локальні кластери та їх взаємозв'язки.

Більш детально сам алгоритм та його математичне обґрунтування описані в оригінальній роботі Леланда М., автора алгоритму [168] та на офіційному англomовному сайті UMAP [169].

Обидва алгоритми зниження розмірності мають такий параметр, як метрика відстані (distance metric). Класичний t-SNE використовує тільки евклідову відстань, однак sklearn реалізація дозволяє змінювати її. Цим параметром власне визначається підхід до обчислення відстані між точками. За замовчанням у всіх алгоритмах застосовується евклідова відстань:

$$d(p, q) = \sqrt{\sum_{k=1}^n (p_k - q_k)^2} \quad (2.23)$$

В науковій спільноті на сьогоднішній день розгорнута досить активна дискусія, щодо того чи є UMAP панацеєю для подолання всіх проблем та недоліків властивих алгоритму t-SNE, і чи доцільно взагалі використовувати останній. Так, наприклад деякі дослідження [158, 163] вказують на обмеженість t-SNE, такі як зав'язаність на локальній структурі даних, нездатність зберігати глобальну структуру, нездатність виділяти викиди даних, високі вимоги до обчислювальних потужностей.

Так, якщо алгоритм t-SNE не в змозі зберігати глобальну структуру, то відстані між кластерами точок не будуть мати якогось економічного сенсу. Віддаленість між окремими кластерами на проектованому просторі не буде означати віддаленість цих кластерів точок у початковому багатовимірному просторі. Також, t-SNE схильний «вирівнювати» відстані між точками в межах одного кластеру, що призводить до того, що кластери можуть виглядати більш щільними в проектованому просторі, аніж вони були в початковому. На деяких наборах даних t-SNE може помилково формувати певні форми з простого «шуму» в даних, що може призвести до непорозумінь та складнощів під час економічної інтерпретації результатів. Часто відзначається, що t-SNE підходить виключно для візуалізації даних, так як зниження розмірності до 4-х та більше вимірів є неефективним [163], проте це є як недоліком так і перевагою, враховуючи що даний

алгоритм демонструє дуже високу якість візуалізації даних і добре підходить для підтримки задачі візуального аналізу даних.

Серед інших сильних сторін можна виділити високу ефективність збереження локальної структури даних, здатність до перетворення нелінійних багатовимірних структур на відміну від деяких класичних методів зниження розмірності типу PCA.

UMAP в свою чергу позбавлений недоліків t-SNE, він є суттєво швидшим алгоритмом, його виконання займає приблизно в 5 разів менше часу; він показує дуже непогані результати щодо збереження глобальної структури даних та часто застосовується у поєднанні з іншими алгоритмами машинного навчання. Так, наприклад в роботі присвяченій дослідженню мутацій вірусу COVID-19 [170] UMAP успішно застосовується в поєднанні з алгоритмом кластеризації K-Means.

Однак, незважаючи на те, що в деяких працях вказується на переваги UMAP над t-SNE [171, 172], деякі дослідження вказують на високу ефективність t-SNE та можливість подолання його недоліків через застосування модифікацій алгоритму, оптимізації гіпер-параметрів та правильного підбору методу ініціалізації [173].

Таким чином, однозначно казати про універсальність та суцільну перевагу одного з алгоритмів складно. Зважаючи також на існуючу полярність думок в науковому середовищі, нами було обрано обидва алгоритми для подальших експериментів: t-SNE для підтримки візуального аналізу та UMAP для підтримки задачі кластеризації. Як буде більш детально описано в Розділі 3, обидва алгоритми досить ефективно впоралися зі своїми задачами.

Сам вибір даних алгоритмів для вирішення задачі аналізу фінансового стану комерційних банків обумовлений високою якістю візуалізації першого, та високою ефективністю у збереженні балансу між локальною та глобальною структурою, високою швидкістю та потужною математичною базою другого.

2.3. Модель аналізу фінансового стану комерційних банків на основі ієрархічного кластерного аналізу

Наступним кроком після отримання двовимірної проекції первинних даних є об'єднання точок отриманої проекції з метою виділення груп банків з певними спільними характеристиками. Для вирішення такого типу задач застосовуються такі алгоритми машинного навчання як алгоритми кластеризації. Розглянемо формальну постановку задачі кластеризації.

Нехай X – множина об'єктів, Y – множина міток (імен) кластерів. Маємо функції відстані між об'єктами $\rho(x, x')$. Існує скінченна навчальна вибірка об'єктів $X^m = \{x_1, \dots, x_m\} \subset X$. Необхідно розбити вибірку на підмножини, що не пересікаються (кластери) таким чином, щоб кожен з кластерів складався з об'єктів, близьких за метрикою ρ , а об'єкти різних кластерів суттєво відрізнялись один від одного. При цьому кожному з об'єктів $x_i \in X^m$ призначається номер кластеру y_i .

Власне алгоритм кластеризації – це функція $a: X \rightarrow Y$, яка кожному з об'єктів $x \in X$ ставить у відповідність номер кластеру $y \in Y$.

Кластеризація (навчання без учителя) відрізняється від класифікації (навчання з учителем) тим, що мітки об'єктів y_i не задані відразу, а в деяких випадках сама множина Y може бути невідома.

Вирішення задач кластеризації є досить специфічним з ряду причин. По-перше, універсального і найкращого критерію якості кластеризації не існує. Кластеризація може оцінюватись за допомогою деяких евристичних критеріїв, проте часто застосування різних підходів оцінки якості кластеризації може давати суперечливі результати. По-друге, кількість кластерів зазвичай наперед невідоме і встановлюється відповідно до суб'єктивних критеріїв дослідника. По-третє, результати кластеризації суттєво залежать від обраної метрики відстані, а її вибір зазвичай також повністю лежить на досліднику.

Так само як і алгоритмів зниження розмірності, алгоритмів кластеризації на сьогоднішній день існує досить багато. Нами було вирішено зупинитись на ієрархічній кластеризації, так як вона показала найкращі результати в процесі

тестування різних алгоритмів, незважаючи на свою простоту та лаконічність, і також продемонструвала високу якість кластеризації у поєднанні з обраними алгоритмами зниження розмірності.

Ієрархічна кластеризація – такий метод кластерного аналізу, що намагається побудувати певну ієрархію об'єктів. Групу цих алгоритмів ще часто називають алгоритмами таксономії. Вони відносяться до так званих «жадібних алгоритмів», тобто відтворюють ітеративний багатокроковий процес, на кожному кроці обираючи найкращий з можливих варіантів (покрокова оптимізація).

Важко стверджувати про те, хто саме є «батьком» даного алгоритму, багато дослідників були причасні до його розвитку. Найстаріші дослідження з використанням ієрархічної кластеризації зустрічаються і описані у Сібсона [174], Рохіфа [175] та Дефайса [176].

В результаті роботи ієрархічні алгоритми формують дендрограму – дерево залежностей між кластерами. Приклад дендрограми можна побачити на Рис. 2.10. Як видно з Рис. 2.10 точки даних поступово об'єднуються в більші групи, кластери. Задача дослідника визначити необхідний рівень деталізації, кількість кластерів що достатня для аналізу об'єкту дослідження.

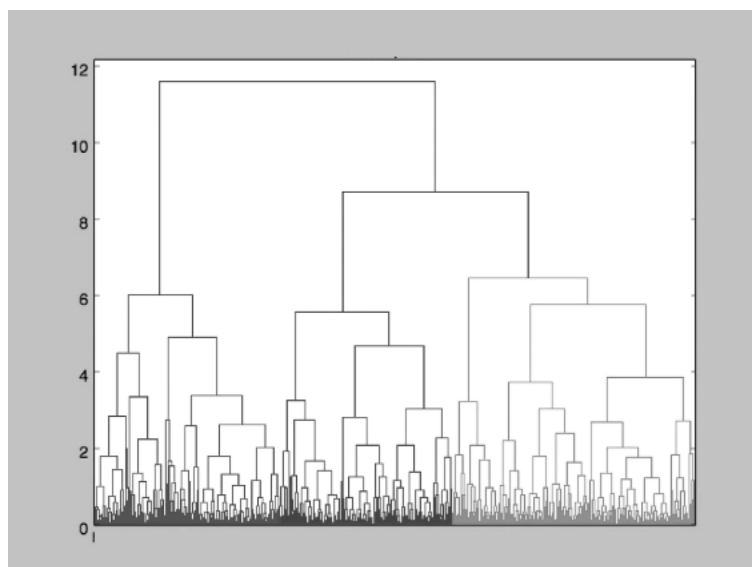


Рис. 2.10. Приклад дендрограми, побудованої алгоритмом ієрархічної кластеризації

Джерело: побудовано автором за допомогою пакету sklearn Python

Під дендрограмою зазвичай мається на увазі дерево, побудоване за матрицею близькості. Дендрограма дозволяє відобразити взаємозв'язки між об'єктами з заданої множини. Для створення дендрограми необхідна матриця подібності (або відмінності), що визначає рівень схожості між парами кластерів.

Загалом виділяють дві стратегії ієрархічної кластеризації:

- Розділювальна. За цієї стратегії спочатку всі точки знаходяться в єдиному кластері, в процесі роботи алгоритм проводить рекурсивне розбиття рухаючись вниз по ієрархії об'єктів.
- Агломеративна (або об'єднувальна). За цієї стратегії спочатку кожна точка є окремим кластером. В процесі роботи алгоритм об'єднує пари кластерів піднімаючись по ієрархії об'єктів.

Найбільш поширеним є застосування крайньої стратегії. Розглянемо її більш детально.

Перш за все, будується матриця подібностей (або матриця відмінностей) – квадратна симетрична матриця з нульовою діагоналлю, де кожна комірка містить відстань між парою точок. Метрика відстані є одним з ключових параметрів алгоритму, і має бути наперед обрана дослідником на свій розсуд, в залежності від поставленої задачі та об'єкту дослідження, або ж експериментальним шляхом. Зазвичай використовується класична евклідова відстань, формула якої розглядалась вище (2.23). Серед інших найбільш вживаних та поширених метрик саме для ієрархічної кластеризації виділяють наступні:

- Квадратична евклідова відстань:

$$\|a - b\|_2^2 = \sum_i (a_i - b_i)^2 \quad (2.24)$$

- Мангетенська відстань:

$$\|a - b\|_1 = \sum_i |a_i - b_i| \quad (2.25)$$

- Відстань Чебишева:

$$\|a - b\|_\infty = \max_i |a_i - b_i| \quad (2.26)$$

- Відстань Махаланобіса:

$$\sqrt{(a - b)^T S^{-1} (a - b)} \quad (2.27),$$

де S – матриця коваріації,

a та b – точки в просторі.

Алгоритм ієрархічної агломеративної кластеризації не описується складними математичними виразами та структурами та може бути представлений досить простим алгоритмом:

- Крок 1: всі точки даних призначаються в окремі кластери, тобто на початку кількість кластерів буде рівна кількості точок в наборі даних.
- Крок 2: визначається найменша відстань в матриці подібності, дві точки з найменшою відстанню одна від одної об'єднуються в один кластер. Матриця подібності оновлюється з урахуванням об'єднання двох точок. Розмірність матриці зменшується на одиницю, тобто матриця $N \times N$ стає розмірністю $(N - 1) \times (N - 1)$. Значення нового кластера встановлюється як максимальне значення з двох поєднаних точок.
- Крок 3: повторюємо крок 2 до тих пір, доки не залишиться один кластер.

На другому кроці алгоритму відбувається перерахунок значень в матриці подібності. Тут важливо відзначити, що метрика відстані застосовується виключно для розрахунку відстані між одно-елементними кластерами, тобто між просто двома точками. Для розрахунку ж відстаней між кластерами, що вже містять більше одного елементу застосовують один з методів розрахунку відстані між кластерами. Метод визначення відстаней між кластерами є другим ключовим параметром алгоритму, що має бути визначений та встановлений дослідником.

Серед найбільш поширених методів можна виділити наступні:

1. Метод одиночного зв'язку (single linkage), або метод найближчого сусіда. Відстань між двома кластерами розглядається як мінімальна відстань між двома елементами з різних кластерів:

$$\min\{d(a, b) : a \in A, b \in B\} \quad (2.28),$$

де $d(a, b)$ – відстань між елементами a і b , що належать кластерам A та B .

2. Метод повного зв'язку (complete linkage), або метод дальшого сусіда. Відстань між двома кластерами розглядається як максимальна відстань між двома елементами з різних кластерів:

$$\max\{d(a, b) : a \in A, b \in B\} \quad (2.29),$$

3. Метод середнього зв'язку (pair-group method using arithmetic mean) – виділяють зважений та незважений. За зваженого, відстань між двома кластерами приймається як середня відстань між елементами цих кластерів:

$$\frac{1}{|A| * |B|} \sum_{a \in A} \sum_{b \in B} d(a, b) \quad (2.30),$$

де $d(a, b)$ – відстань між елементами a і b , належних кластерам A та B ,
 $|A|$ та $|B|$ - потужності кластерів (кількість їх елементів).

4. Центроїдний метод (pair-group method using the centroid average) – виділяють зважений та незважений. За зваженого, відстань між двома кластерами приймається як відстань між їх центроїдами (центрами мас).

5. Метод Уорда (Ward's method). На відміну від інших методів кластерного аналізу, для оцінки відстаней між кластерами тут використовуються методи дисперсійного аналізу. В якості відстані між кластерами береться приріст суми квадратів відстаней об'єктів до центру кластеру, що отримується в результаті їх об'єднання:

$$\Delta(A, B) = \frac{n_A n_B}{n_A + n_B} \|\vec{m}_A - \vec{m}_B\|^2 \quad (2.31),$$

де m_j – центра кластеру j ,

n_j – кількість точок в кластері,

Δ називають вартістю злиття кластерів A та B .

На кожному кроці алгоритму об'єднуються два таких кластери, що призводять до мінімального збільшення дисперсії. Цей метод застосовується для задач з близько розташованими кластерами. Більш детально він описаний в роботі автора Дж. Уорда [177].

Кожен з розглянутих методів має свої сильні та слабкі сторони, вибір є суб'єктивним, проте зазвичай можна орієнтуватись на особливості вхідного набору

даних, так як кожен з методів краще або гірше підходить в залежності від його специфіки. Так, наприклад метод одиничного зв'язку є більш швидким з точки зору виконання алгоритму, та краще підходить для роботи з не кулеподібних по формі даних, проте показує погані результати при наявності в даних шуму. Метод повного та середнього зв'язку краще підходить для добре відділених один від одного кулеподібних кластерів, а в інших випадках може видавати суперечливі результати. Метод Уорда часто застосовується для роботи з зашумленими даними та у випадках, коли кластери в даних можуть бути наближені один до одного та мати різний розмір. Приклад роботи алгоритму ієрархічної кластеризації на тестових даних різної форми з використанням різних методів визначення відстані між кластерами можна побачити на Рис. 2.11.

Як видно з Рис. 2.11, метод одиничного зв'язку добре справляється з даними нестандартної форми, на відміну від інших методів (ряд 1 та 2 на рисунку).

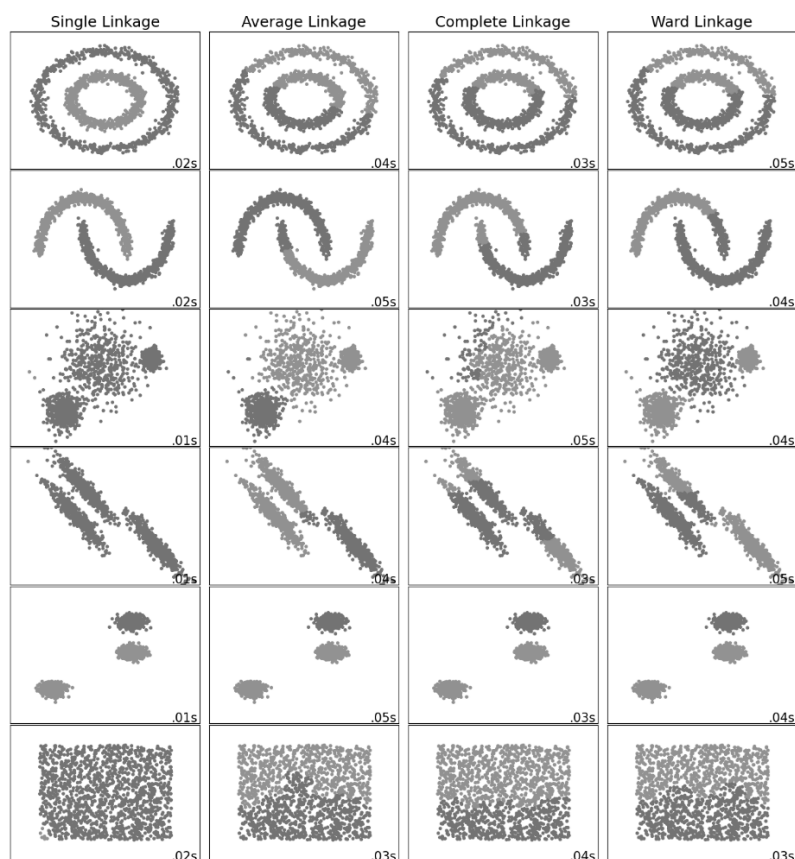


Рис. 2.11. Результати застосування алгоритму ієрархічної кластеризації з різними методами визначення відстані між кластерами на прикладі тестових даних різної форми.

Джерело: [178]

Найкращий результат для даних стандартної форми показує метод Уорда (ряд 3 та 5 на рисунку) в той час як інші методи демонструють незадовільний результат. Для даних витягнутої форми жоден з алгоритмів не демонструє прийняттого результату (ряд 4 на рисунку), тому для таких даних рекомендовано обрати інший алгоритм кластеризації. Ряд 6 на рисунку показує результати застосування різних методів для кластеризації рівномірного розподілу.

Серед усіх розглянутих методів нами було вирішено обрати метод Уорда, так як він краще підходить для роботи з даними і кластерами стандартної форми у вигляді скупчень точок (як видно з Рис. 2.11) та добре працює з зашумленими даними. Як ми можемо побачити з Рис. 2.3, двовимірні проекції наших даних, на які ми надалі проводитимемо кластеризації, дійсно мають стандартну форму, схожу на ті, що показані на рисунку в рядах 3 та 5. Також, подальший аналіз даних в Розділі 3 дозволяє нам припустити, що дані можуть бути зашумлені, в даних можуть бути присутні викиди, навіть після їх первинної обробки.

Розглянемо обраний нами метод Уорда трохи детальніше. Інші розглянуті методи використовують матрицю відмінності, що може бути отримана з використанням евклідової, мангетенської відстанями, відмінністю між категоріями та інших методів розрахунку відстаней. Метод Уорда, навпроти, базується на сумі квадратичних помилок, що обмежує вибір метрики відстані однією тільки евклідовою відстанню. Крім того, визначення суми квадратичних помилок потребує врахування так званих центроїдів кожного кластеру, тобто вектору середніх значень всіх спостережень, що належать до певного кластеру. Таким чином, метод приймає на вхід матрицю X розмірністю $n \times p$ (n спостережень по p ознакам).

Основною задачею методу Уорда є мінімізація погіршення загального показника внутрішньої суми квадратів, що є сумою квадратів відхилень між спостереженнями в кластері та центроїдом (середнім):

$$WSS = \sum_{i \in C} (x_i - \bar{x}_c)^2 \quad (2.32),$$

де WSS – це внутрішня сума квадратів,

$\bar{x}_c$ – центроїд кластеру С.

В процесі роботи алгоритм ієрархічної кластеризації покроково об'єднує по два існуючі кластери, що призводить до погіршення загального WSS. Метод Уорда розроблений таким чином, щоб пом'якшити це погіршення, а якщо точніше, то він мінімізує різницю між новим (більшим) WSS у об'єднаному кластері та суми WSS всіх компонентів, що були об'єднані, що по суті зводиться до мінімізації відстаней між центрами кластерів.

Для двох спостережень відстані визначаються наступним чином:

$$d_{AB}^2 = \frac{2n_A n_B}{n_A + n_B} \|\bar{x}_A - \bar{x}_B\|^2 \quad (2.33),$$

де $\|\bar{x}_A - \bar{x}_B\|$ – евклідова відстань між двома центрами кластерів:

n_A – координати точки А,

n_B – координати точки В.

Наступний крок подібний до всіх інших методів: обираються і об'єднуються два спостереження / кластери, що мають найменшу відстань один від одного.

Розрахунок відстані від кластеру чи спостереження Р до нового кластеру С, отриманого від об'єднання спостережень А і В є більш складним та проводиться за наступною формулою:

$$d_{PC}^2 = \frac{n_A + n_P}{n_C + n_P} d_{PB}^2 + \frac{n_B + n_P}{n_C + n_P} d_{PB}^2 - \frac{n_P}{n_C + n_P} d_{AB}^2 \quad (2.34),$$

де n_C – координати нового кластеру С,

n_P – координати кластеру чи спостереження Р.

Проте для спрощення розрахунків в алгоритмі беруться дані, отримані з матриці відмінності оновленої на попередньому кроці, що дозволяє уникнути постійного обчислення центроїдів. На етапі першого об'єднання розраховується проста квадратична евклідова відстань між двома спостереженнями, без центроїдів.

Як і будь-який інший алгоритм ієрархічна кластеризація має певні недоліки, такі як: низька швидкість та ефективність роботи з великими масивами даних, що обумовлена великою кількістю операцій перебору даних; неможливість алгоритму повертатись до попередніх кроків для повторної оптимізації, тобто відсутність

зворотного зв'язку в алгоритмі; сильний вплив на результати кластеризації має вибір методу визначення відстані між кластерами. Деякі методи погано взаємодіють з шумами та викидами в даних, інші можуть розбивати великі кластери на під-кластери, деякі погано справляються з кластерами різної форми або видають різні результати в залежності від порядку подачі даних в алгоритм.

Серед переваг алгоритму ієрархічної кластеризації можемо виділити наступні: результати кластеризації можуть бути досить легко інтерпретовані, зокрема через можливість побудови дендрограми, більшість програмних реалізацій алгоритму ефективно використовує додаткові ядра обчислювального пристрою, тобто є ресурсо-ефективними, алгоритм не потребує обов'язкового встановлення кількості кластерів апріорі для своєї роботи.

Загалом, визначення якості кластеризації є специфічною та дещо складною задачею, так як алгоритми машинного навчання без учителя частіше за все мають справу з задачами, де мітки класів (кластерів) апріорі невідомі, як у задачах класифікації. А отже отримані результати кластеризації не можна просто порівняти з певним еталоном, і визначити ефективність, якість роботи алгоритму. Проте існує ряд метрик, націлених на аналіз внутрішньої структури кластерів та їх розміщення між собою. Самі по собі ці метрики можуть не надати корисної інформації, проте порівняння метрик отриманих в результаті роботи різних алгоритмів та їх конфігурацій може показати, який з них є ефективнішим за інші. Розглянемо ці метрики детальніше.

Коефіцієнт силуету (Silhouette coefficient). Значення метрики варіюються від -1 до 1. Метрика розраховується наступним чином:

$$s = \frac{b - a}{\max(a, b)} \quad (2.35),$$

де a – це середня відстань між спостереженням та всіма іншими точками того самого кластеру,

b – середня відстань між спостереженням та всіма іншими точками найближчого сусіднього кластеру.

Чим вищим є значення метрики, тим краще [179]. Значення наближені до -1 говорять про високу щільність спостережень, до +1 – про ідеальну дискримінацію (розділення) спостережень, до 0 – про накладання сусідніх кластерів один на одного.

Оцінка Калінського-Харабаза (Calinski-Harabasz score) представляє з себе співвідношення між-кластерної та внутри-кластерної дисперсій:

$$S = \frac{tr(B_k)}{tr(W_k)} \times \frac{n_E - k}{k - 1} \quad (2.36),$$

де $tr(B_k)$ – слід матриці між-кластерних дисперсій,

$tr(W_k)$ – слід матриці інтра-кластерних дисперсій,

E – набір вхідних даних,

n_E – обсяг вхідного набору даних,

k – кількість кластерів.

Кращими є більш високі значення метрики [180].

Оцінка Девіса-Боулдіна (Davies-Bouldin score) – описує середню схожість між кластерами, що є мірою оцінки відстаней між кластерами і розмірами самих кластерів:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{ij}; \quad R_{ij} = \frac{s_i + s_j}{d_{ij}} \quad (2.37),$$

де s_i – середня відстань між кожною точкою кластеру i та центроїдом цього кластеру,

d_{ij} – відстань між центроїдами кластерів i та j .

Значення ближчі до 0 говорять про більш високу якість кластеризації [181].

Висновки до розділу 2

В другому розділі дисертаційної роботи розроблено та запропоновано загальний методологічний підхід до аналізу фінансового стану комерційних банків. Перші два кроки спрямовані на формування початкового набору показників та відбір ознак в модель. Наступні 4 кроки - це вибір алгоритмів машинного навчання,

що будуть застосовуватись, а саме алгоритмів зниження розмірності та кластеризації, тестування даних на предмет придатності до кластеризації, вибір методів первинної обробки даних. Останні 3 кроки - це первинна обробка даних, аналіз фінансової стійкості із застосування алгоритмів зниження розмірності та застосування алгоритмів кластеризації при алгоритмах зниження розмірності для аналізу фінансової стійкості.

Сформовано початковий набір показників діяльності комерційних банків на основі низки джерел. В результаті, отримано набір показників, що складається з 63 показників, таких що можуть бути розраховані на основі відкритої статистичної звітності НБУ. Всі показники поділені на такі групи: аналіз активів, аналіз пасивів, аналіз достатності (адекватності) капіталу, аналіз збалансованості активів і пасивів, аналіз ліквідності, аналіз фінансової ефективності, аналіз податкового навантаження, аналіз ефективності використання (залучення) персоналу, загальний аналіз.

На основі кореляційного аналізу визначено такі фактори вхідного набору даних, що мають парну кореляцію $> 0,5$. Проведено багатокроковий ітераційний процес відбору ознак: всі корельовані пари показників проаналізовано за допомогою таких метрик і показників як VIF, MAD та SPEC. В процес відбору ознак також включено експертну складову: на основі ряду експериментів з додаванням та виключенням показників в модель, таким чином, що (на основі досвіду, або за припущенням автора) можна було б підсилити роздільну здатність кластеризації та загальну якість моделі і адекватність інтерпретації результатів, було додано та виключено декілька показників в наборі даних, отриманого суто аналітичним шляхом. В результаті, сформовано набір показників, що складається з 17 ознак. Подано перелік відібраних показників, їх короткий опис та описові статистики.

Для вирішення задачі зниження розмірності вирішено обрати два алгоритми: t-SNE та UMAP. Висунуто припущення, що перший більше підходить для аналізу самої двовимірної проекції, як потужний алгоритм візуалізації, а другий – для подальшої кластеризації проекції. Дане припущення буде перевірене в наступному

розділі. Для вирішення задачі кластеризації обрано алгоритм ієрархічної агломеративної кластеризації. Розкрито логіку та етапи роботи алгоритмів та розглянуто математичний апарат, на якому вони побудовані. Для обраного алгоритму кластеризації розглянуто можливі варіанти розрахунку відстаней між спостереженнями та методи розрахунку відстаней між кластерами. Для розрахунку відстаней між кластерами обрано метод Уорда. Виділено та розглянуто основні метрики для оцінки якості кластеризації, порядок їх розрахунку та оптимальні значення.

Основні результати Розділу 2 опубліковано у наукових працях автора [182, 183, 184, 185, 186, 187, 188].

РОЗДІЛ 3

ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ФІНАНСОВОГО СТАНУ КОМЕРЦІЙНИХ БАНКІВ УКРАЇНИ НА ПІДҐРУНТІ СИСТЕМИ МОДЕЛЕЙ ЗНИЖЕННЯ РОЗМІРНОСТІ ТА КЛАСТЕРИЗАЦІЇ

3.1. Інформаційне забезпечення системи моделей аналізу фінансового стану комерційних банків

Перш за все, варто уточнити структуру та характер використовуваних даних. Ми використовували статистичні звіти НБУ по зведеним агрегованим балансовим залишкам в розрізі банків (дані по активам, пасивам, капіталу, фінансовим результатам), що надаються регулятором на щоквартальній основі. Статистика містить дані по працюючим банкам та деяким банкам, що знаходяться в стані ліквідації. Для дослідження банківської системи України було обрано інтервал часу з I кварталу 2015 року по IV квартал 2021 року включено (всього 24 періодів). Вибір саме такого періоду обумовлений зокрема тим, що він характеризується відносною стабільністю фінансового та банківського секторів, і відсутністю економічних шоків. Фінансова криза 2014-2015 років та війна 2022 року безсумнівно складають сильний вплив на функціонування та показники діяльності банківської системи в цілому та кожного комерційного банку зокрема. Включення даних періодів в модель з високою імовірністю негативно вплине на адекватність моделі, а в мету та задачі даного дослідження не входить врахування кризового та/або форс-мажорного фактору.

Всього вибірка містить 2529 спостережень. Спостереження, що містили пропуски даних або помилки розрахунків при діленні на нуль, було вирішено замінити на 0. Таких спостережень у вибірці налічується 116, що складає всього 4,5% вибірки. За такої частки значень механізм їх обробки не має суттєвого впливу на результати моделювання, проте нами було вирішено не видаляти такі значення з набору даних, так як пропущені або нульові значення у балансовій звітності можуть бути інтерпретовані як негативні індикатори роботи комерційного банку.

Вибірка банків містить дані по 147 банківським установами, з них 71 установа є платоспроможною та здійснює банківську діяльність станом на початок 2022 року, 76 банківських установ припинили банківську діяльність за період з початку 2015 року (52%).

Якщо детальніше проаналізувати розраховані показники та розглянути їх описові статистики, то легко побачити деякі особливості, що властиві даним по більшості з коефіцієнтів. Наочно ці особливості можна побачити на прикладі показника K1.3 в Таблиці 3.1 (перший рядок таблиці).

Таблиця 3.1.

Порівняльна характеристика даних до і після застосування методів попередньої (первинної) обробки на прикладі коефіцієнту покриття кредитного ризику K13.

Дані	Гістограма розподілу	Min	Max	Середнє	Стандартне відхилення
K1.3 до первинної обробки		-2,30	27691,87	18,01	570,83
K1.3 після вінзування та стандартизації		-0,67	3,27	0,000	1,000
K1.3 після вінзування та робустної стандартизації		-0,43	6,06	0,67	1,65

Джерело: розроблено автором на основі статистичних даних НБУ по агрегованим балансовим залишкам комерційних банків України за період з I кварталу 2015 року по IV квартал 2021 року.

Як видно з Таблиці 3.1, показник K1.3 має специфічний ненормальний розподіл, так як дані містять суттєві викиди, через що дані на гістограмі розподілу відображаються у вигляді одного стовпчика. Максимальне значення по даному показнику складає 27691,87, що, зважаючи на середнє значення рівне 18,01, є дуже суттєвим відхиленням. Застосування таких даних для подальшого моделювання негативно впливатиме на роботу алгоритмів, адекватність проектування багатовимірною простору у двовимірний та результати кластеризації.

Для того, щоб уникнути цього, в практиці моделювання та машинного навчання застосовуються методи первинної обробки даних, серед яких найбільш вживаними є стандартизація та нормалізація даних. На етапі роботи з даними нами були також розглянуті та протестовані деякі менш традиційні підходи до їх обробки.

Класична стандартизація даних проводиться наступним чином [189]:

$$z = \frac{X - \mu}{\sigma} \quad (3.1),$$

Де X – вхідні дані (за однією ознакою),

μ - середнє значення за однією ознакою,

σ – середньоквадратичне відхилення за однією ознакою.

По суті застосування стандартизації приводить дані до вигляду, в якому середнє значення вибірки є рівним 0, а середньоквадратичне відхилення рівне 1.

Менш відомими та вживаними різновидами класичної стандартизації є мінімаксна стандартизація (minmax scaler) та стандартизація через абсолютизацію максимального значення (maxabs scaler). За першого підходу дані трансформуються таким чином, щоб вони знаходились на заданому дослідником проміжку. За другого підходу дані трансформуються таким чином, що абсолютне максимальне значення по кожній ознаці стає рівним 1. У випадку, якщо проводиться трансформація даних для їх розподілу на проміжку від 0 до 1, такий підхід називається нормалізацією даних.

Цікавим підходом є так звана робустна стандартизація даних (robust scaler). Застосування робустної стандартизації забезпечує трансформування даних із використанням статистик, що є менш чутливими (робустними) до викидів. Даний метод усуває медіану з даних і розподіляє дані на проміжку між 1-им та 3-ім квантилем даних, тобто між 25-им і 75-им квантилем. Цей проміжок називається інтер-квартільний проміжок. У випадку, коли в даних наявні викиди, трансформації через медіану та інтер-квартільний проміжок показують кращі результати, ніж подібна трансформація через середнє значення та середньоквадратичне відхилення. Розрахунок проводиться наступним чином:

$$\frac{x_i - Q_1(x)}{Q_3(x) - Q_1(x)} \quad (3.2)$$

де x_i – вхідні дані (по одній ознаці),

$Q_1(x), Q_3(x)$ – відповідно значення першого та третього кuartилю.

Наостанок, розглянемо ще один підхід до обробки даних, що називається вінзорування, або вінзоризація (winsorization). Даний підхід застосовується для обмеження викидів даних, та вперше був представлений у 1960 році в роботі Уілфріда Діксона [190]. Його суть полягає в тому, щоб обмежити максимальні та мінімальні екстремальні значення для зменшення ефекту викидів, прирівнюючи такі значення до певних процентилів. Наприклад, 90%-ва вінзоризація встановлює всі значення спостережень, що вищі за 95-ий перцентиль рівними до значення 95-го перцентилля, а всі значення нижчі за 5-ий перцентиль рівними до значення 5-го перцентилля.

Відзначимо, що підходів до обробки даних ще існує досить багато, наприклад у деяких випадках, залежно від типу, структури, особливостей даних, може застосовуватись бінаризація, степенева або функціональна трансформація даних. Однак, ми не будемо розглядати всі з цих підходів, так як це виходить за межі нашого дослідження.

Розглянувши різні методи первинної обробки даних, необхідно визначити які з них є найбільш підходящим для застосування на нашому наборі даних. Водночас, важливим кроком є визначення загальної придатності даних до кластеризації, перед проведенням експериментального дослідження. Ми поєднали вирішення цих двох задач, розраховували та проаналізували деякі показники придатності даних до кластеризації в розрізі різних методів первинної обробки даних.

Розглянемо власне які методи можуть бути застосовані для аналізу даних з точки зору їх придатності до кластеризації.

Тест Хопкінса – статистичний тест, що дозволяє визначити чи є дані рівномірно розподіленими. Якщо тест позитивний (розраховане значення наближається до 0) – дані не є рівномірно розподіленими, отже можуть містити

окремі кластери. Вважається, що якщо тест $> 0,3$ – дані є рівномірно розподіленими і кластеризація даних навряд чи дасть адекватний результат [191].

Аналіз тенденції показує оптимальну кількість кластерів для заданого набору даних. Для цього використовується алгоритм кластеризації K-means та три методи оцінки його ефективності - silhouette, calinski та davies bouldin score. На основі оцінок, отриманих при розрахунку трьома метриками береться середнє значення. Вищі значення показника вказують на те, що дані краще піддаються розділенню; значення наближені до 2 скоріше за все вказують на те, що дані є досить однорідними, і алгоритм K-means здатен виділити обмежену кількість відокремлених скупчень точок.

Додатково нами був розрахований показник асиметрії (skew). Він не має відношення до придатності даних до кластеризації, проте може бути використаний як індикатор загальної якості розподілу даних. **Асиметрія** - оцінка рівня «скошеності» даних. З математичної точки зору є центральним моментом третього порядку. Чим більше асиметрія відрізняється від 0, тим менша подібність розподілу даних до нормального. Асиметрія розраховується наступним чином:

$$As = \frac{\mu_3}{\sigma^3} \quad (3.3)$$

де μ_3 – центральний момент третього порядку,

σ – середньоквадратичне відхилення.

Вище ми зазначали, що наш набір даних має деякі особливості, і лівостороння асиметрія, властива більшості коефіцієнтів, є однією з них. Такий «скіс» в цілому може бути нормальним явищем і пояснюється природою самих коефіцієнтів: більшість спостережень мають значення наближені до 0. Так, наприклад, за коефіцієнтом K0 більшість банків є малими за розміром активів, що є адекватним з економічної точки зору. Навпроти, за коефіцієнтом K1.3 середнє значення показника є рівним 18, що є високим (рівень кредитного резервування на рівні обсягів кредитів вже свідчить про можливо низьку якість кредитного портфелю), і при цьому показник ще й має сильну лівосторонню асиметрію (це можна зрозуміти з гістограми в Таблиці 3.1, вона має лише один стовпчик в лівій частині графіку). Така ситуація однозначно свідчить про наявність суттєвих викидів в даних. В таких

випадках, асиметрія може виступати індикатором того, чи є дані спотвореними викидами, додатними (висока додатна асиметрія) чи від’ємними (асиметрія < 0).

Розглянемо Таблицю 3.2, в якій представлені всі розраховані метрики для даних з різними підходами до їх первинної обробки.

Таблиця 3.2.

Метрики придатності даних до кластеризації та показники якості даних в розрізі різних методів їх первинної обробки

Метод обробки	Тест Хопкінса	Тенденція	Середня асиметрія	Стандартне відхилення
Початкові дані	0,00152	35,33333	7,20824	9806,93341
Стандартизація	0,01208	9,66667	7,20824	1,00020
Робустна стандартизація	0,00256	35,00000	7,20824	46269,30783
Нормалізація	0,07478	33,00000	1,27830	0,13831
Вінзорування	0,06434	33,33333	3,31239	0,38453
Мінімаксна стандартизація	0,04104	33,66667	7,20824	0,09605
Стандартизація <i>maxabs</i>	0,04304	34,00000	7,20824	0,10580
Вінзорування + стандартизація	0,06099	33,66667	3,31239	1,00020
Вінзорування + нормалізація	0,08970	34,66667	3,31239	0,13333
Вінзорування + робустна стандартизація	0,03726	2,66667	1,52929	2,02443

Джерело: розроблено автором на основі статистичних даних НБУ по агрегованим балансовим залишкам комерційних банків України за період з I кварталу 2015 року по IV квартал 2021 року

Відповідно до Таблиці 3.2, для наших даних, незалежно від методу попередньої обробки – тест Хопкінса показує значення наближене до 0, отже дані не є рівномірно розподілені, тобто міститимуть окремі кластери. Метрика тенденції

для більшості випадків має значення 33 або більше, що свідчить про високу схильність даних до кластероутворення. Відзначимо, що значення тенденції не обов'язково означає точну кількість кластерів в даних, а лише вказує на можливу орієнтовну їх кількість, і є індикатором придатності даних до кластеризації. З розрахунків асиметрії видно, що найбільший вплив на приведення даних до нормального розподілу та пом'якшення викидів мають нормалізація даних та поєднання вінзорування і робустної стандартизації. Суттєвий вплив мають такі підходи як вінзорування, поєднання вінзорування з нормалізацією та стандартизацією. Середньоквадратичне відхилення є адекватним для всіх методів окрім робустної стандартизації.

Порівнявши всі методи обробки даних, ми зупинились на виборі двох груп методів. Це вінзорування з подальшою стандартизацією та вінзорування з подальшою робустною стандартизацією. Комбінація вінзорування з класичною стандартизацією забезпечила кращі результати двовимірного проектування за допомогою алгоритму t-SNE, а поєднання вінзорування з робустною стандартизацією забезпечило значно адекватніші результати кластеризації двовимірної проекції, отриманої за допомогою UMAP. Ефективність даних двох підходів підтверджується і розрахованими метриками, представленими в Таблиці 3.2. та даними і гістограмами з Таблиці 3.1: спостерігається більш виражений нормальний розподіл даних з лівосторонньою асиметрією.

Для визначення верхнього та нижнього лімітів вінзорування ми слідували рекомендаціям дослідників Хоберга та Філіпса [192], що рекомендують встановлювати 1% в якості обох лімітів. Такий підхід добре себе показав при зниженні розмірності за допомогою алгоритму UMAP і подальшої кластеризації. Ми також провели ряд додаткових експериментів з лімітами, знижуючи розмірність різних наборів даних, і виявили підвищення якості візуалізації t-SNE при нижньому і верхньому лімітах відповідно у 2% та 5%.

Необхідно зазначити, що тенденція для даних, оброблених за допомогою вінзорування та робустної стандартизації має значення близьке до 2, що свідчить про низьку кластероутворювальну здатність даних. Однак ми проводимо

кластеризацію на двовимірній проекції а метрики вище розраховані на первинних даних, де тенденція буде нижчою, що пов'язано зі специфікою багато-вимірності даних. Перевіривши метрику тенденції для двовимірної проекції ми отримали результат 92, що свідчить про однозначну придатність даних до кластеризації.

Непогані результати показує нормалізація, проте стандартизація виявилась кращою з точки зору впливу на результат двовимірного проектування та подальшої кластеризації. Також, вважається, що стандартизацію краще застосовувати для даних що містять викиди, так як цей метод менш чутливий до них. Стандартизація частіше застосовується для методів машинного навчання без учителя. Якщо дані подібні за масштабом але не є нормально розподіленими також рекомендовано застосовувати стандартизацію. Зважаючи вищезазначене, нами було зроблено вибір на користь стандартизації як додаткового методу обробки даних.

Всі етапи моделювання, починаючи з первинної обробки даних і до етапу проведення кластеризації двовимірної проекції було виконано засобами мови програмування Python 3.6 у хмарному середовищі Google Collab на базі Google Computer Engine. Для реалізації всіх етапів моделювання нами застосовувався ряд сторонніх бібліотек:

1. Pyclusterend (1.7.0) – розрахунок та аналіз показника тенденції вибірки та тесту Хопкінса, аналіз даних та придатності до кластеризації.
2. Sklit-dimension (0.3) – статистичний аналіз даних, аналіз придатності даних до кластеризації, аналіз асиметрії.
3. Pandas (1.4.3) – робота з таблицями даних (датафреймами, dataframes). Датафрейми є загальноновживаним типом даних, що застосовується при роботі з великими масивами даних, бібліотека pandas значно спрощує роботу з даними та надає широкий спектр можливостей для маніпуляції ними.
4. Numpy (1.23.0) – проведення наукових розрахунків, робота з багатовимірними масивами, матрицями, додаткові методи для швидкої обробки даних та маніпуляцій над ними. Numpy.matlib – модуль для роботи з матрицями, що застосовуються в більшості алгоритмів як вхідний тип даних. Numpy.linalg – модуль з методами лінійної алгебри, необхідний для маніпуляції над даними та їх

підготовки для роботи з деякими алгоритмами, наприклад для розрахунку власних векторів, власних значень в алгоритмі SPEC.

5. `Matplotlib (3.5.2)` – візуалізація даних, побудова графіків різних типів та складних візуальних конструкцій на основі масивів даних. `Matplotlib.pyplot` – модуль, що забезпечує інтерактивність графіків, дозволяє виводити графіки на екран у вигляді картинок формату `.png` або `.jpeg`.

6. `Seaborn (0.11.2)` – візуалізація даних, має перевагу над `matplotlib` в тому, що надає більш зручний набір методів для тонкого налаштування параметрів та зовнішнього виду графіків, однак в цілому має надає менше можливостей для побудови складних графіків.

7. `Sklearn.preprocessing` на базі `scikit-learn (1.1.2)` – модуль, що містить набір методів призначений для попередньої обробки даних. Загалом `scikit-learn` це потужна бібліотека для моделювання даних, що містить широкий спектр моделей різного роду. Серед методів обробки даних модуля `preprocessing` ми застосовували `StandardScaler`, `RobustScaler`, `MinMaxScaler`, `MaxAbsScaler`, `Normalizer`.

8. `Scipy.stats.mstats` на базі `scipy (1.9.0)` – модуль для роботи зі статистичними функціями. Загалом бібліотека `scipy` містить колекцію готових математичних алгоритмів та готових функцій, що спрощують деякі рутинні операції, побудовані на основі `numpy`. Нами застосовувався метод `winsorize`.

9. `Scipy.sparse` на базі `scipy (1.9.0)` – модуль для роботи з розрідженими (`sparse`) матрицями, що застосовується в якості вхідного типу даних в деяких алгоритмах, наприклад у SPEC.

10. `Sklearn.feature_selection` на базі `scikit-learn (1.1.2)` – модуль, що містить набір методів для відбору ознак у модель. Нами застосовувався метод `VarianceThreshold`.

11. `Statsmodels.stats.outliers_influence` на базі `statsmodels (0.14.0)` – модуль для роботи з викидами та аналізу даних на предмет їх якості в процесі відбору ознак. Нами застосовувався метод `variance_inflation_factor`. `Statsmodel` в цілому – це бібліотека, що містить набір готових статистичних моделей різних видів.

12. `Umap-learn.umap` (0.5) – бібліотека, що містить готовий алгоритм UMAP та супровідні методи для аналізу даних з використанням даного алгоритму зниження розмірності.

13. `Sklearn.manifold` на базі `scikit-learn` (1.1.2) – модуль, що містить ряд моделей для багатовимірного проектування для роботи з даними типу багатовидів (manifolds). Нами застосовувалися алгоритм TSNE.

14. `Sklearn.metrics` на базі `scikit-learn` (1.1.2) – модуль, що містить метрики оцінки якості кластеризації. Нами застосовувались такі метрики як `silhouette_score`, `davies_bouldin_score`, `calinski_harabasz_score`.

Вибір Python для реалізації моделей даної роботи було зроблено опираючись на ряд переваг даної мови програмування, ключові серед яких, на нашу думку, наступні:

- Python має дуже велику кількість сторонніх бібліотек, призначених для підтримки роботи з даними, та для моделювання в цілому. Такі бібліотеки активно підтримуються та розширюються а їх перелік постійно росте.
- Python є дуже зручним, ефективним та водночас простим інструментом для вирішення завдання візуалізації даних. Такі бібліотека як `matplotlib` та `seaborn` вже пройшли перевірку часом та є дуже популярними на сьогоднішній день.
- Так як Python загалом призначений не тільки для роботи з даними, а є повноцінною мовою програмування як для веб-додатків так і програмного забезпечення, то він є перспективним інструментом для подальшої побудови інформаційної системи на основі моделей, представлених в даній роботі.
- Python є досить простою мовою програмування з точки зору синтаксису і має велику кількість вже вбудованих методів та бібліотек, що значно полегшує процес його вивчення та роботи з ним.

Зважаючи на вищезазначені переваги, на нашу думку, Python є більш гнучким аніж спеціалізовані програмні пакети, такі як STATISTICA або Deductor Studio і в

той же час простішим та відносно зручнішим ніж спеціалізовані мови програмування типу R або MATLAB.

3.2. Експериментальне дослідження фінансового стану комерційних банків України на підґрунті алгоритмів зниження розмірності

Для проведення нашого дослідження на попередніх етапах (див. Розділ 2.3) нами було відібрано два алгоритми зниження розмірності: t-SNE та UMAP. Як нами вже зазначалось раніше, за результатами проведених численних експериментів з даними алгоритмами у поєднанні з різними алгоритмами кластеризації ми дійшли висновку про доцільність застосування t-SNE для підтримки візуального аналізу і застосування UMAP у поєднанні з алгоритмом кластеризації (в рамках нашого дослідження найкращі результати показав алгоритм ієрархічної агломеративної кластеризації методом Уорда) для кластерного аналізу. На Рис. 3.1. можна побачити двовимірні проекції, отримані після застосування t-SNE та UMAP.

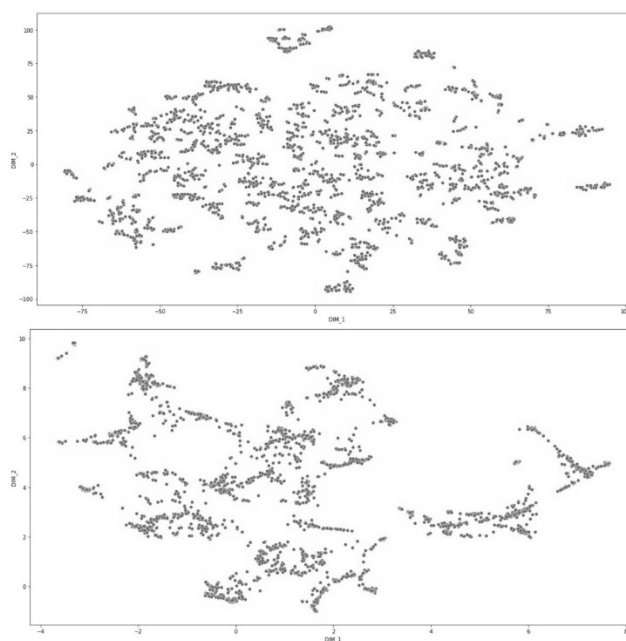


Рис. 3.1. Двовимірна карта проекції точок з 17-вимірного простору ознак, отримана за використання алгоритму t-SNE (зверху) та UMAP (знизу)

Джерело: розроблено автором на основі статистичних даних НБУ за I квартал 2015 року по IV квартал 2021 року

Як видно з Рис. 3.1., отримана за допомогою t-SNE карта представляє з себе досить однорідний масив точок, який алгоритм кластеризації не зможе адекватно

розбити на групи, а якщо бути точнішим, то в залежності від типу алгоритму – точки будуть або розбиті на велику кількість мікро-кластерів (така поведінка буде спостерігатись, наприклад, при застосуванні алгоритмів заснованих на розділенні (partitioning-based)), або об'єднані в один або пару макро-кластерів (така поведінка буде спостерігатись, наприклад, при використанні алгоритмів заснованих на щільності (density-based)). Такі особливості проекції мають забезпечити ефективність, адекватність застосування алгоритму кластеризації. Метрики якості кластеризації підтверджують перевагу UMAP над t-SNE: кластеризація проекції, отриманої за оптимальної конфігурації UMAP має в середньому показник силуету 0,4697; кластеризація проекції, отриманої за оптимальної конфігурації t-SNE має показник силуету 0,3607.

Незважаючи на непридатність результатів двовимірного проектування алгоритму t-SNE до кластеризації, він демонструє досить хороший результат у збереженні як локальної так і глобальної структури даних. Для того, щоб переконатись в цьому, ми відобразимо на двовимірній карті точки, що відносяться тільки до останнього періоду дослідження (IV квартал 2021 року). Отриманий результат зображено на Рис. 3.2. На карті-проекції на Рис. 3.2. видно, що банки вже є більш чітко згрупованими та групи можуть бути виділені наочно.

Загалом, основною задачею нашого дослідження є проведення кластерного аналізу за підтримки алгоритмів зниження розмірності. Проте візуальний аналіз двовимірних проекцій, отриманих в результаті застосування t-SNE також показав цікаві результати, тому нами було **вирішено виділити декілька другорядних задач**, першою з яких є **аналіз локальної та глобальної структури даних на їх двовимірній проекції**. В рамках даної задачі ми плануємо перевірити гіпотезу, що відібрані нами алгоритми зниження розмірності здатні на достатньому рівні зберігати локальну та глобальну структуру багатовимірних даних, для уможливлення подальшої економічної інтерпретації результатів та кластеризації двовимірних проекцій.

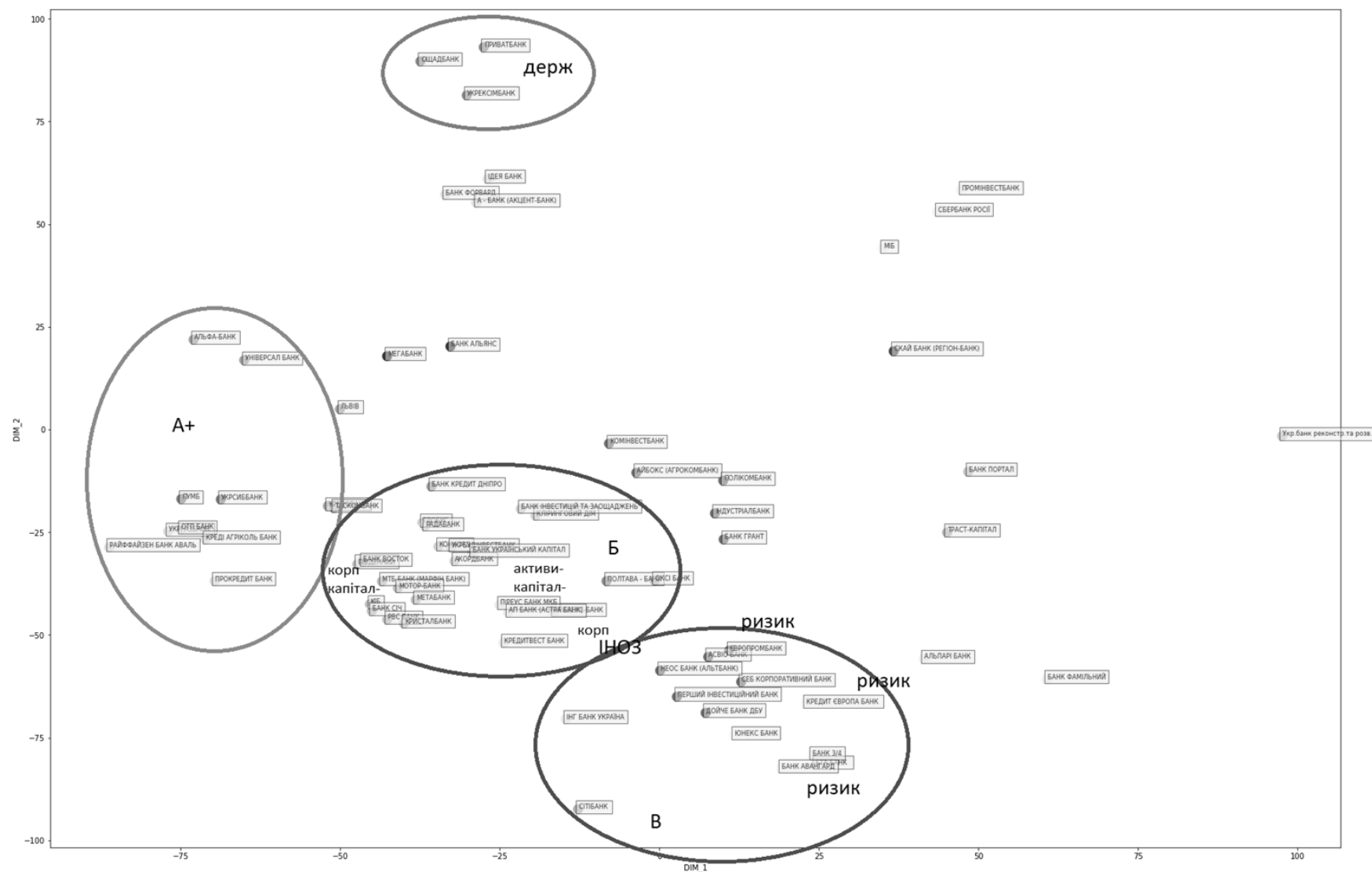


Рис. 3.2. Двовимірна карта проекції точок з 17-вимірного простору ознак, отримана за використання алгоритму t-SNE, з маркерами точок-банків

Джерело: розроблено автором на основі статистичних даних НБУ за I квартал 2015 року по IV квартал 2021 року

Вирішення даної задачі виконувалось через візуальний аналіз отриманої проекції, на якій відображені та промарковані точки-банки за останній період. На основі проекції t-SNE навіть наочно досить просто виділити групи точок. Нами були проаналізовані виокремлені групи, області та банки в них, спираючись на два підходи: розглядаючи рейтинги банків (кредитні рейтинги національних агенцій, рейтинги авторитетного українського фінансового видання Мінфін та офіційні національні та міжнародні кредитні рейтинги (якщо такі є) вказані на офіційному ресурсі НБУ); та аналізуючи ряд ключових на нашу думку коефіцієнтів по всім банкам, що входять в групу або область (K1.3 Коефіцієнт покриття кредитного ризику, K1.13 Відношення кредитного портфелю до активів, K3.1 Коефіцієнт автономності, K5.2 Частка високоліквідних активів, K6.8 Рентабельність витрат).

Більшість точок-банків на проекції розташовуються у лівому нижньому квадранті. Тут ми можемо виділити декілька областей.

В області **A+** згрупувались першокласні банки, що мають найвищі оцінки інших експертів і показують хороші результати за своїми показниками діяльності. Тут знаходяться ТОП-10 банків за рейтингом Мінфін, 8 з 11 банків мають найвищий рейтинг за національною шкалою uaAAA.

Сусідня область **B** сформована найбільшою групою банків, в ній розташовані невеликі банки, з середнім рівнем ефективності та інвестиційної привабливості, або на рівні вище середнього.

В **нижній частині** області знаходяться переважно корпоративні банки (сфокусовані на кредитування юридичних осіб), більшість з них можуть мати проблеми з капіталізацією; банки що знаходяться ближче до межі області є ризикованими, маючи високу частку неякісних активів або низьку рентабельність, ведучи при цьому досить ризиковану кредитно-інвестиційну політику.

Банки в **центральной та верхній частині** області (ближче до верхньої межі) так само є ризикованими, маючи проблеми з якістю активів та рівнем капіталізації, деякі з банків мають низьку рентабельність. Переважна більшість банків є універсальними банками, мають ризиковану структуру депозитів, низьку частку строкових вкладів. Банки в **нижній частині** області (ближче до межі з областю **B**)

– це переважно банки з іноземним капіталом, всі з банків орієнтовані на корпоративне кредитування. Всі установи мають низьку рентабельність, одна з них – суттєві збитки. Інші показники знаходяться на оптимальному рівні.

Банки в області **В** переважно мають показники ефективності та фінансової стійкості на середньому рівні, або нижче середнього, проте їх важко охарактеризувати як однозначно проблемні; переважно банки є іноземними, корпоративними або інвестиційними. 7 банків з 13 з даної області є банками з міжнародним капіталом, разом з банками з нижньої частини області **Б** ці банки складають половину всіх банків з іноземним капіталом. Банки, що наближаються до лівої або нижньої межі є ризикованими, маючи високу частку проблемних активів, низьку рентабельність, деякі є збитковими.

У верхній частині карти згруппувались державні банки, що за своїми показниками є скоріше проблемними, однак враховуючи урядову підтримку, їх не можна однозначно класифікувати як такі. Всі банки, які не потрапили до виділених областей можуть бути класифіковані як «аутсайтери», такі банки є проблемними з різних причин, і вони є згрупповані між собою під впливом різних наборів показників, відповідно до яких такі банки можна віднести до проблемних.

Відзначимо, що хоча даний аналіз і здається чисто суб'єктивним – це не зовсім так. При грамотному налаштуванні гіпер-параметрів, алгоритм буде завжди відтворювати певні закономірності: точки з екстремальними значеннями ознак будуть відштовхуватись до країв карти-проекції, інвестиційно привабливі та проблемні банки будуть знаходитись в різних частинах карти, банки з найкращими показниками будуть притягуватись до одного кута карти, банки з найгіршими показниками – до іншого. Опираючись на ці закономірності, досить просто провести аналіз двовимірної карти навіть за випадкової ініціалізації алгоритму.

Найбільш адекватний результат зниження розмірності був отриманий при наступних налаштуваннях t-SNE:

- *Perplexity* (регулює баланс між фокусом на локальні чи глобальні аспекти даних) = 15.
- *Learning rate* (швидкість навчання) = 500.

- *Iterations* (кількість ітерацій навчання) = 1000.
- Ініціалізація алгоритму через PCA.
- Застосування модифікації Barnes-Hut.

Дана конфігурація була визначена шляхом багаточисельних експериментів з різними гіпер-параметрами, і в цілому добре співвідноситься з рекомендаціями Кобака та Баренса в їх роботі [173]. При цьому, за результатами проведення візуального дослідження і виконання першої задачі, можна стверджувати про достатню здатність алгоритму зберігати як локальну структуру (банки зі схожими характеристиками розміщуються поруч у групах), так і глобальну структуру даних (групи інвестиційно привабливих і проблемних банків суттєво віддалені одні від одного).

Другою задачею є **аналіз розміщення активних банків на фоні банків, що вийшли з ринку**. В рамках даної задачі ми перевіряємо гіпотезу, що всі або більшість банків мають схожі показники, за якими можливо охарактеризувати проблемність того чи іншого банку. В такому випадку, ми можемо очікувати, що в результаті зниження розмірності даних, за збереження їх глобальної структури, проблемні банки будуть згруповані разом. При цьому причини, через які банки потрапляють в проблемні області можуть корелювати з причинами закриття неактивних банків, що потрапили в цю ж область. Для вирішення даної задачі відобразимо на двовимірній проекції окрім активних банків також банки, що вже вибули з ринку, взявши дані по ним за останній період, в якому вони були платоспроможні. Результат можна побачити на Рис. 3.3.

Як видно з Рис. 3.3, переважна більшість неактивних банків знаходиться у верхній правій частині проекції, тобто вони розміщені діаметрально протилежно групам активних, платоспроможних банків. Майже всі неактивні на момент дослідження банки знаходяться в області, куди ми віднесли банки «аутсайдери» (це 71 з 75 неплатоспроможних установи, тобто 95% всіх неактивних банків).

визнаний банкрутом та проводиться його ліквідація, одна з причин закриття – це проблеми з капіталом.

Р2: в даній області відсутні неактивні банки, 3 активні банки що потрапили сюди характеризуються низькою якістю активів та низькою ліквідністю. Дещо схожою є група держбанків, що розташовується поруч, однак вони є системними і мають значно більше активів.

Р3: з 10 неактивних банків, що потрапили в дану область, закриття 3-ох банків пов'язане з втратою ліквідності, 2-ох банків пов'язане з втратою капіталу, 3 установи вийшли з ринку через процедуру самоліквідації, 2 установи пройшли через процедуру злиття. Активні банки, що потрапили в область на досліджуваний період мають суттєві відхилення від норми майже по всім досліджуваним показникам: низька якість кредитно-інвестиційного портфелю, низька ліквідність, недокапіталізованість, пасивна кредитно-інвестиційна діяльність, низька рентабельність. Банки з такими характеристиками імовірно можуть бути кептивними (кишеньковими). Банки АТ «Сбербанк» та АТ «Промінвест» були вилучені з банківського ринку наказом НБУ після початку російського вторгнення в Україну.

Р4: в даній області знаходяться банки зі специфічними показниками діяльності, що пов'язане з особливістю їх функціонування. АТ «Розрахунковий центр» був призначений для обслуговування договорів на фінансових ринках, в липні 2021 року його задачі були визнані виконаними і установу вилучили з банківського ринку. АТ «Держзембанк» був стратегічним банком, що мав підтримувати підприємства АПК та впровадження відповідних державних програм. Однак в умовах відсутності відповідних змін до законодавства, банк з моменту отримання банківської ліцензії майже не здійснював банківську діяльність, передбачену статутом. У липні 2016 року його визнано неплатоспроможним. Виключенням тут є банк АТ «Гефест», так як він не є спеціалізованою установою, проте він має специфічні показники так як майже не вів ніякою діяльності з моменту свого відкриття. Банк було ліквідовано через непрозору структуру власності та через недостатню капіталізацію. Активний на

досліджуваний період АТ «Український банк реконструкції та розвитку» був створений як спеціалізована установа для підтримки розвитку інноваційної інфраструктури, а на даний момент є банком з 100% китайського капіталу. Він має низький рівень ліквідності, майже не має депозитів, збитковий.

P5: з 19 неактивних банків, що потрапили в дану область, 1 банк пройшов через процедуру злиття, 8 банків ліквідовано через непрозору структуру власності, шахрайство чи фінансові махінації, 7 установ вийшло з ринку через процедуру самоліквідації, 1 банк ліквідовано через погіршення фінансового стану і невиконання зобов'язань, 2 банки ліквідовано через низьку якість активів, пасивну діяльність та втрату ліквідності. Такі специфічні причини вибуття з ринку зазвичай супроводжуються попереднім погіршенням фінансових показників. З активних на досліджуваний період банків всі 4 установи, що потрапили в дану область характеризуються низькими обсягами депозитів, низькими обсягами прибутку (або збитком), 2 з 4 банків мають високу частку низькоякісних активів в портфелі, спостерігається зниження ліквідності та пасивність кредитно-інвестиційної політики. Всі ці ознаки з високою імовірністю можуть вказувати на кептивність банків з даної області. Деякі тенденції, такі як збільшення обсягів резервування при зменшенні частки низькоякісних активів, може свідчити про фінансові махінації.

На основі проведеного аналізу, ми можемо зробити висновок, що проблемні банки дійсно проявляють схильність потрапляти в області, в яких знаходились закриті на даний момент банки, в останній період перед їх ліквідацією. Знаходження банку в області, в якій сформувалось скупчення вже неактивних банків є тривожним знаком, а причини закриття неактивних банків, що знаходяться поруч, можуть вказувати можливі проблемні місця активних установ. Таким чином, ми підтверджуємо гіпотезу що висували в рамках другої задачі.

Також, виходячи з ідентифікованих областей банків-аутсайдерів, можна надати певні рекомендації, впровадження яких потенційно може поліпшити стан установ:

- Для ефективного управління ліквідністю банку рекомендовано в першу чергу регулярно проводити аналіз розривів ліквідності (гепів), а саме так

званого управлінського гепу, що враховує фактичні строки з урахуванням історичних строків руху залишків на рахунках, прогнозів та статистичного аналізу. На основі отриманих результатів аналізу комітет з управління активами та пасивами банку (КУАП) розробляє механізм лімітування розривів ліквідності, виходячи із фінансових можливостей банку та його стратегії розвитку. Обов'язковим елементом системи управління ліквідністю є проведення стрес-тестування на випадок виникнення кризового стану в межах банку або всієї банківської системи. На випадок виникнення кризового стану. Регулювання стану ліквідності може проводитись через операції на міжбанківському ринку. Ефективна організаційна структура банку, внутрішньої нормативної бази, постійний моніторинг ринкової ситуації також є ключовими елементами успішного процесу управління ліквідністю.

- Питання докапіталізації (або нарощування капіталу) банківських установ ніколи не втрачає своєї актуальності, так як інфляційні процеси постійно зменшують вартість власних коштів установ, а нестабільне економічне середовище передбачає можливість виникнення макроекономічних шоків. Для нарощення капіталу комерційні банки можуть провести повторний випуск акцій для розширення статутного капіталу, поповнити наявні резервні фонди (розширення резервного капіталу) за рахунок отриманого прибутку, або збільшити капітал за рахунок нерозподіленого прибутку наприкінці року. При цьому очевидно, що рівень капіталізації залежить від загальної ефективності функціонування банку. Базовою рекомендацією для підтримання достатнього рівня капіталізації буде застосування одного або декількох підходів до нарощення капіталу. Також, у випадку виявлення недокапіталізації, доцільним може бути зміна стратегії розвитку банку з активної експансивної на пасивно-стабілізаційну через суттєве скорочення використання власного капіталу на розширення бізнесу і формування додаткових резервів за рахунок отриманого прибутку, дивідендів та/або через розширення статутного

капіталу. Більш серйозні рекомендації стосовно капіталізації банків надає Базель III, вводячи необхідні нормативи додаткових буферів капіталу: контр-циклічний буфер (обмежує кредитну активність за рахунок підвищення вимог до капіталу, але не є доцільним тільки для фінансових та економічних систем, що активно розвиваються); буфер системної важливості (захищає великі системні банки від економічних шоків за рахунок попереднього формування додаткових резервів); буфер консервації (сприяє накопичуванню додаткового резерву вище мінімальних вимог регулятора у економічно стабільні періоди, для подальшого розконсервування в кризові періоди).

- Результатом значного погіршення якості кредитних активів за останні 10 років стали в основному зовнішні економічні шоки та загальна розбалансованість банківської системи, також перехід на нові міжнародні стандарти розрахунку проблемних кредитів. Іншим фактором погіршення якості активів є нехтування застосуванням адекватних аудиторських процедур для скринінгу та аналізу потенційного позичальника. Особливо по відношенню до великих позичальників-юридичних осіб, адже майже всі об'ємні позики надаються саме таким позичальникам. На сьогоднішній день вже можна спостерігати значне поліпшення ситуації в порівнянні з попередніми роками, для продовження цієї позитивної тенденції важливими є впровадження або вдосконалення існуючих методів та підходів до оцінки позичальників, як фізичних так і юридичних осіб, впровадження та модифікація існуючих методів скорингової оцінки позичальників та впровадження або адаптація зарубіжних успішних методик та підходів аудиторської оцінки великих позичальників – юридичних осіб.

Третьою, останньою задачею яку ми розглянемо це **аналіз стабільності комерційних банків на основі динаміки показників їх діяльності**. Наша гіпотеза в рамках даної задачі полягає в тому, що різкі або постійні зміни (коливання) в показниках роботи банку можуть свідчити про наявність проблем в їх операційній

діяльності. Такі зміни будуть супроводжуватись переміщенням банків по двовимірній карті-проекції.

Для перевірки даної гіпотези для початку розглянемо графік переміщення банку класу А (область А+ з Рис. 3.2), банку з показниками вище середнього (область Б з Рис. 3.2) та ризикованого банку (аутсайдера) на двовимірній проекції за останні 3 роки дослідження (після-кризовий стабільний період), поквартально і так само розглянемо неактивний банк за його даними за останні 3 роки перед його віднесенням до категорії неплатоспроможних (Рис. 3.4).

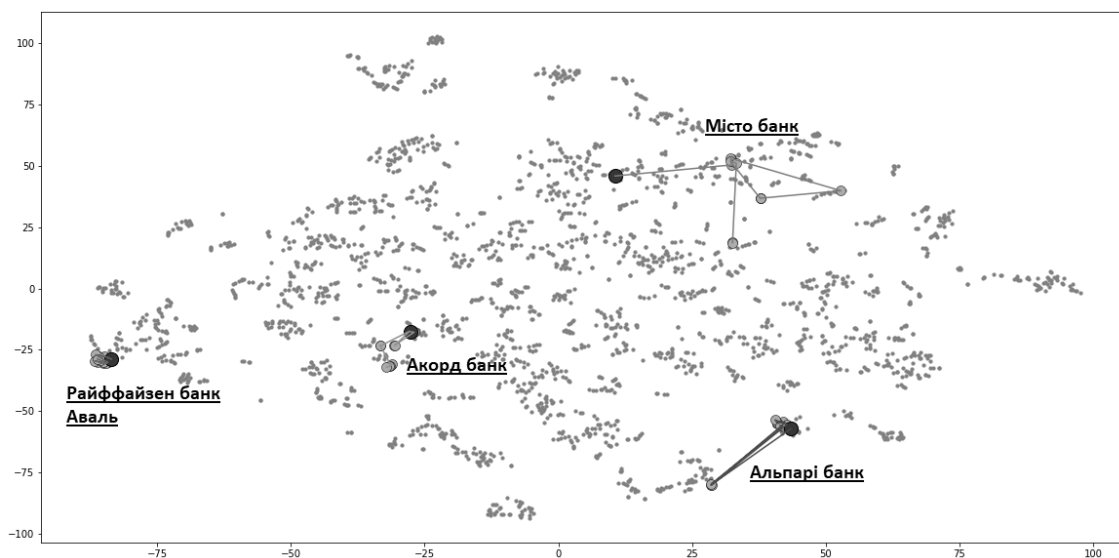


Рис. 3.4. Двовимірні проекції багатовимірного простору ознак, відображення динаміки банків Райффайзен банк Аваль, Акорд банк, Альпарі банк та Місто банк за період з I кварталу 2019 року по IV квартал 2021 року

Джерело: розроблено автором на основі статистичних даних НБУ за I квартал 2019 року по IV квартал 2021 року

Як видно з Рис. 3.4, першокласний банк (Райффайзен банк Аваль) практично не змінював свого положення протягом останніх трьох років, банк з групи Б (Акорд банк) демонструє незначні переміщення в рамках своєї групи, а неактивний банк (Місто банк) та потенційно ризиковані банки мають схожу особливість поведінки — вони часто змінюють своє положення на карті-проекції і при цьому можуть виходити далеко за межі своєї групи. Поведінку банків в динаміці можна також перевірити аналітично. Для цього ми можемо взяти координати точок двовимірної проекції, і розрахувати евклідову відстань між усіма парами точок, після чого

оцінити коливання відстаней за допомогою показника середньоквадратичного відхилення. Для вибраного для аналізу банку з групи А+ цей показник склав 3,82, для банку з групи Б – 76,10, для ризикованого банку – 386,65, а для неактивного банку – 421,07. Таким чином, ми можемо обґрунтовано стверджувати, що проблемні банки можуть характеризуватись суттєвою нестабільністю показників своєї діяльності у часі, що може виступати ознакою, відділяючою їх від більш надійних банків. Ми можемо говорити про підтвердження гіпотези, що висувалась в рамках даної задачі.

Варто відзначити, що даний підхід аналізу динаміки банківських показників не є альтернативою попереднім підходам візуального аналізу та не може повністю або частково замінити кластеризацію банківських установ. Це пояснюється тим, що за проведення кластеризації, нами для аналізу розглядається останній досліджуваний період, тобто проводиться статичний аналіз, що суттєво спрощує процес інтерпретації отриманих результатів. В представленому ж вище підході ми застосовуємо динамічний аналіз, тобто аналіз стану банківських установ у часі. При цьому цей підхід може бути ефективно застосований як доповнення до кластерного аналізу на етапі інтерпретації отриманих результатів, доповнюючи статичний аналіз динамічною складовою. Ми повернемось до нього і будемо застосовувати його для підтримки інтерпретації результатів кластерного аналізу в наступному підрозділі.

3.3. Експериментальне дослідження фінансового стану комерційних банків України на підґрунті алгоритму кластеризації

У попередньому підрозділі нами застосовувався алгоритм зниження розмірності t-SNE, що показав хороші результати в якості інструменту підтримки візуального аналізу даних. Однак, як було зазначено і доведено вище, результати роботи даного алгоритму погано підходять для подальшої кластеризації. Алгоритм UMAP в свою чергу видає результати, які складніше аналізувати візуально, однак отримані двовимірні проекції значно краще піддаються кластерному аналізу, а

отже даний алгоритм доцільно застосовувати у комплексі з алгоритмом кластеризації.

Для зниження розмірності в алгоритмі UMAP ми встановили наступні гіпер-параметри, що були визначені як оптимальні експериментальним шляхом, і спираючись на дослідження проведені автором алгоритму, візуальні результати якого зображені на Рис. 3.5:

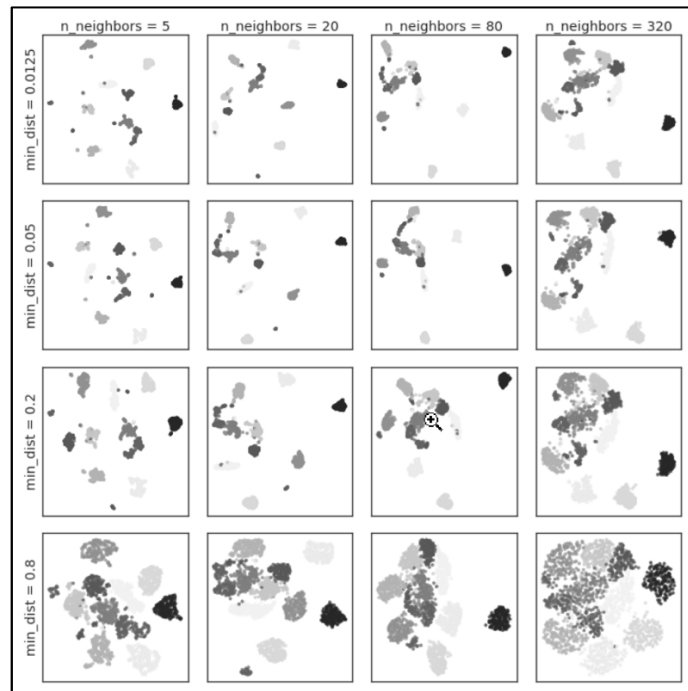


Рис. 3.5. Порівняння двовимірних проекцій, отриманих за різних гіпер-параметрів *min_dist* та *n_neighbors* алгоритму UMAP

Джерело: [168]

- *N_neighbors* (кількість сусідів, контролює акцент на збереження локальної чи глобальної структури) = 40.
- *Min_dist* (мінімальна відстань, на якій можуть знаходитись точки одна від одної у низько-розмірному представленні) = 0.
- *Distance* = correlation (кореляційна відстань).

Як видно з Рис. 3.5, і як вже було зазначено в підрозділі 2.3, за менших значень *min_dist* та *n_neighbors* точки на двовимірній проекції будуть розташовуватись більш щільно в межах своїх груп, а самі групи будуть сильніше відштовхуватись одна від одної, прагнучи найбільш точно унаслідувати глобальну структуру даних. Збільшення значень гіпер-параметрів матиме обернений ефект.

В результаті тестування алгоритму з використанням різних метрик відстаней, нами було обрано кореляційну відстань, що обчислюється наступним чином:

$$1 - \frac{(u - \bar{u}) \cdot (v - \bar{v})}{\|(u - \bar{u})\|_2 \|(v - \bar{v})\|_2} \quad (3.4)$$

де $\bar{u}$ – середнє значення з набору (масиву) даних u ,

$x \cdot y$ – сума добутку x на y ,

u та v – два набори (масиви) даних з двома координатами для кожної точки з масиву.

В результаті роботи алгоритму було отримано двовимірну проекцію, зображено на Рис. 3.1 (нижня частина).

Для проведення кластеризації нами застосовувався алгоритм агломеративної ієрархічної кластеризації методом Уорда. Серед параметрів алгоритму обов'язковим є тільки встановлення кількості кластерів, що мають бути отримані в результаті. Для того, щоб визначити оптимальну кількість кластерів, ми провели ряд експериментів з встановленням кількості кластерів від 8 до 15, спираючись при цьому на результати метрик силуету та оцінок Калінського-Харабаза і Девіса-Боулдіна, що розглядались нами в підрозділі 2.3. Кількість кластерів обрана для проведення експериментів на основі візуальної оцінки двовимірної проекції UMAP (Рис. 3.1, нижня частина) та керуючись міркуваннями щодо адекватності розподілу банків на групи: кількість кластерів менша за 8, на нашу думку, не дасть можливості достатньо ефективно розбити банки на кластери і може призвести до хибного віднесення деяких банків до однієї групи через нестачу «вільних» кластерів; а встановлення кількості кластерів більше 15 може призвести до виділення суттєвої кількості кластерів з одним чи декількома точками-банкам при аналізі за останній період, чи навіть до ситуації коли кожен кластер міститиме одну-дві точки в останньому періоді.

Результати прогонів алгоритму представлені в Табл. 3.3.

Таблиця 3.3.

**Значення метрик якості кластеризації, отриманих в результаті експерименту
з визначення оптимальної кількості кластерів**

Кількість кластерів	Силует	Оцінка Davies-Bouldin	Оцінка Calinski-Harabasz	Ранг
8	0,4679	0,6902	3629,43	3
9	0,4512	0,7796	3683,74	1,8
10	0,4503	0,7533	3851,12	1,6
11	0,4694	0,6809	4046,88	4,6
12	0,4846	0,6934	4294,83	4,8
13	0,4856	0,678	4624,13	6,4
14	0,4973	0,6841	4850,89	6,8
15	0,5148	0,6946	5384,49	7

Джерело: розраховано автором на основі статистичних даних НБУ по агрегованим балансовим залишкам комерційних банків України за період з I кварталу 2015 року по IV квартал 2021 року.

Аналізуючи Табл. 3.3, ми застосували метод ранжування, при цьому надавши показнику силуету вагу 0,6 (так як він є більш вживаним і вважається досить ефективною метрикою для оцінки якості кластеризації), а іншим двом – по 0,2. З отриманих значень рангів по якості кластеризації найкращим виявилось спостереження з 15 кластерами. Отриманий результат можна вважати таким, що описує якість кластеризації як високу, враховуючи, що ми не видаляли викиди з даних а згладжували їх для збереження їх впливу на процес розбиття банків на групи.

Отриманий візуальний результат кластеризації комерційних банків України візуально представлено на Рис. 3.6.

На Рис. 3.6 можна побачити сформовані 15 кластерів, які ми пронумерували в тому порядку, в якому ми проводили аналіз та економічну інтерпретацію отриманих результатів.

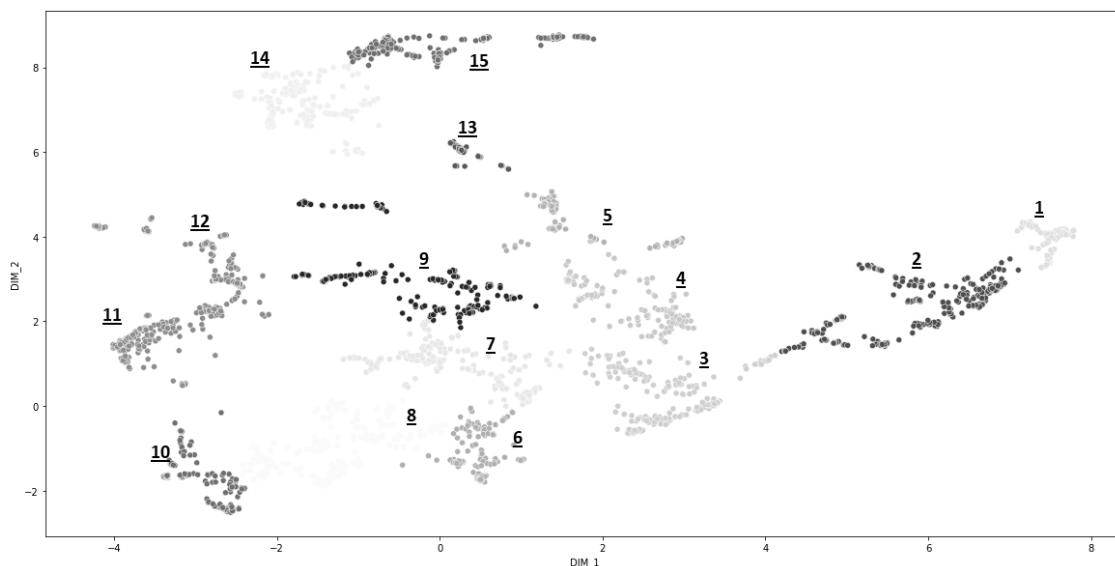


Рис. 3.6. Візуальна репрезентація результатів роботи алгоритму кластеризації на основі двовимірної проекції UMAP, 15 кластерів

Джерело: побудовано автором на основі статистичних даних НБУ за I квартал 2015 року по IV квартал 2021 року

Двовимірну карту активних банків станом на IV квартал 2021 року та неплатоспроможних і ліквідованих банків станом на останній період в якому вони були платоспроможними, отриману із застосуванням алгоритму UMAP, з нанесеними мітками банків можна знайти в Додатку Г.

Характеристика кластерів, отриманих в результаті кластеризації представлені в Табл. 3.4. Для характеристики кожного кластера ми вказали його номер; кількість активних банків, що потрапило до кластеру в останньому періоді дослідження (IV квартал 2021 року); кількість неактивних банків, що потрапили до кластеру в останній період дослідження, в якому ці банки були платоспроможними. Вказані деякі показники банківської діяльності, такі як:

- K0 Коефіцієнт системності. Оцінка ваги банку в банківській системі. Нами розглядалась сума по всім банкам з кластеру.
- K1.3 Коефіцієнт резервів під покриття втрат. Характеризує якість кредитного портфелю банку. Чим вища частка першокласних активів у портфелі, тим нижчими мають бути обсяги резервування. Оптимальним вважається значення $<4\%$ [149, 193]. Нами розглядалось середнє значення по всім банкам.

Таблиця 3.4.

Характеристика кластерів, отриманих в результаті кластеризації українських комерційних банків за період з 2015 по 2021 роки за допомогою алгоритму агломеративної ієрархічної кластеризації методом Уорда.

# кластеру	Активні	Неактивні	K0	K13	K113	K21	K31	K52	K68	P1	P2	Динаміка	Середній рейтинг
1	4	0	49,5%	77%	66%	26,8%	9,11%	23%	2,60%	53%	13%	21,92	5,25 / 9,00
2	11	0	33,5%	6%	88%	20,8%	11,24%	33%	2,86%	12%	3%	102,91	7,90 / 9,00
3	2	8	0,7%	6%	83%	44%	10,51%	7%	1,07%	19%	4%	306,93	3,00 / 9,00
4	4	2	2,7%	35%	66%	40,9%	12,20%	13%	4,01%	23%	15%	889,80	2,20 / 9,00
5	3	14	3,9%	400%	39%	5%	13,37%	24%	2,65%	84%	44%	482,66	-0,67 / 9,00
6	18	2	5%	8,9%	86%	23,2%	10,46%	34%	1,20%	22%	2%	936,27	3,12 / 9,00
7	3	1	0,2%	21,4%	55%	22,8%	25,9%	23%	4,76%	38%	3%	375,83	3,67 / 9,00
8	12	1	1,2%	9,9%	81%	19,2%	24,90%	33%	1,65%	22%	3%	1165,39	5,16 / 9,00
9	5	7	0,3%	32%	76%	10,7%	26,08%	53%	-1,61%	23%	1%	687,55	3,00 / 9,00
10	3	0	2,1%	0,7%	90%	16,3%	18,94%	22%	2,70%	4%	0%	14,89	6,34 / 9,00
11	1	12	0,01%	37%	50%	0,1%	65,68%	11%	0,28%	23%	18%	84,42	1,00 / 9,00
12	1	4	0,01%	0,04%	88%	0,8%	88,98%	60%	-2,70%	83%	0%	386,65	0,00 / 9,00
13	1	2	0,02%	166%	83%	0,3%	46,62%	9%	3,02%	6%	4%	4,66	4,00 / 9,00
14	2	6	0,08%	7,5%	90%	12,2%	12,70%	55%	-2,68%	7%	1%	1379,27	4,00 / 9,00
15	0	15	Н/Д	Н/Д	Н/Д	Н/Д	Н/Д	Н/Д	Н/Д	Н/Д	Н/Д	Н/Д	Н/Д

Джерело: розраховано та систематизовано автором на основі статистичних даних НБУ по агрегованим балансовим залишкам комерційних банків України за період з I кварталу 2015 року по IV квартал 2021 року.

- K1.13 Відношення кредитного портфелю до активів. Характеристика ділової активності банку. Вважається, що показник $< 65\%$ свідчить про пасивну політику, $> 75\%$ - про ризиковану, показник на проміжку 65-75% - про активну політику. Нами розглядалось середнє значення по всіх банкам [149].
- K2.1 Ризик незбалансованої стійкості. Характеризує рівень стійкості за рахунок можливості управління пасивами і маневреності управління строками розміщення активів. Ріст показника може свідчити про збалансованість управління активними та пасивними операціями по строкам, обсягам залучення і розміщення ресурсів. Зниження чи відсутність росту може свідчити про вразливість банку до будь-яким ринковим змінам [149]. Нами розглядалось середнє значення по всіх банкам.
- K3.1 Коефіцієнт автономності. Характеризує забезпеченість вкладів банку під ризиком його капіталом. Оптимальним вважається значення між 8 та 15%, значення менше за 8% може свідчити про високий ризи банкрутства [149, 151, 193]. Нами розглядалось середнє значення по всіх банкам.
- K5.2 Частка високоліквідних активів. Індикатор ліквідності, що характеризує структуру ліквідності активів банку, а отже достатність грошових коштів. Індикатор дозволяє зрозуміти, яку частку активів банк тримає у найбільш ліквідній формі, що дозволяє швидко погашати вимоги кредиторів і вкладників. Оптимальним вважається значення $> 15\%$ [149, 151]. Нами розглядалось середнє значення по всіх банкам.
- K6.8 Рентабельність витрат. Характеризує ефективність витрат банку, обсяг прибутку на одиницю витрат [149, 193]. Нами розглядалось середнє значення по всіх банкам.
- Частка активів 4-5-ї категорій (для фізичних осіб) та 9-10-ї категорій (для юридичних осіб). Такі активи вважаються найменш якісними. Додатково розрахований на основі відкритих статистичних даних показник (P1). Нами розглядалось середнє значення по всіх банкам.

- Відношення суми ризику по активам 4-5-ї категорій (для фізичних осіб) та 9-10-ї категорій (для юридичних осіб) до загального обсягу активів. Додатково розрахований на основі відкритих статистичних даних показник (P2). Нами розглядалось середнє значення по всім банкам.
- Показник стабільності банку в динаміці як середнє значення середньоквадратичних відхилень відстаней (Динаміка) між усіма парами точок за останні 3 роки по кожному з банків у кластері. Для орієнтування: значення відхилень по всім банкам лежать на проміжку від 1,82 до 3523,88, в середньому ризиковані та потенційно проблемні банки можуть мати значення цього показника на рівні більше 350.

Також, спираючись на вищезазначений перелік показників нами була сформована проста авторська рейтингова система, на основі якої було проведене ранжування всіх активних на період дослідження банків. Отримані рейтинги в більшості співпадають з рейтингами ресурсу Мінфін та національною рейтинговою шкалою. В таблиці для кожного кластеру наведене середнє значення рейтингу по всім банкам з кожного кластеру. Рейтинг визначався відповідно до наступного набору правил:

- Якщо $K0 > 1\%$ - додати один бал (суттєві обсяги активів в системі).
- Якщо $K1.13 < 65\%$ - відняти один бал, якщо між 65% та 85% - додати один бал (відповідно до оптимуму, однак з трохи пом'якшеною верхньою межею, так як висока активність скоріше є позитивною рисою).
- Якщо $K5.2 > 15\%$ - додати один бал, інакше відняти один бал, якщо $> 20\%$ - додати ще один бал (відповідно до оптимуму, та додатковий бал за підвищену ліквідність).
- Якщо $P1 < 25\%$ - додати 1 бал, якщо $P1$ між 25% та 50% - забрати один бал, якщо більше 50% - відняти 2 бали (межі встановлені на розсуд автора).
- Якщо $K1.3$ більше 25% і $P1$ менший за 25% - відняти один бал (збільшення резервів за одночасного поліпшення якості кредитного портфелю може свідчити про фіктивні операції або фінансові махінації).

- Якщо $P2 > 35\%$ - відняти один бал, інакше – додати 1 бал (межі встановлені на розсуд автора).
- Якщо $K3.1 > 8\%$ - додати один бал, якщо менше – відняти 1 бал (відповідно до оптимуму).
- Якщо $K6.8 > 1\%$ - додати один бал (найменший адекватний рівень показника).
- Якщо $K2.1$ між 20 та 50% - додати 1 бал (достатній рівень показника для зниження ризиків залежності від вкладів фізичних осіб та водночас забезпечення відсоткових доходів).
- Якщо Динаміка > 350 – відняти один бал, інакше додати один бал (висока волатильність, що була визначена нами при аналізі даного показника для проблемних банків).

Відзначимо, що дана рейтингова система не претендує на повну об'єктивність і високу точність оцінки. Хоча рейтинги банків і базуються виключно на коефіцієнтах, що розраховані на основі актуальної відкритої статистики НБУ, самі показники для розрахунку рейтингів були відібрані суто суб'єктивно, виходячи з нашого бачення важливості тих чи інших показників для аналізу фінансового стану банків. Ціллю рейтингової системи не є оцінка конкретних банків як поганих чи хороших, вона призначена виключно для ранжування сформованих в процесі моделювання кластерів для їх подальшого порівняння між собою.

Значення результуючої шкали варіюються від -2,00 до 9,00 балів. Банки, з кількістю балів від -2,00 до 1,00 можна охарактеризувати як **однозначно проблемні** (функціонують на межі неплатоспроможності, більшість показників відхиляються від оптимальних, фінансово нестійкі установи), банки з кількістю балів від 1,01 до 4,00 – як **ризиковані** (один чи декілька показників суттєво відхиляються від оптимуму, банки знаходяться на межі їх визначення як проблемних, переважно фінансово нестійкі), банки з кількістю балів від 4,01 до 6,00 – як **банки класу Б** (один показник може мати суттєві відхилення від оптимуму, інші показники можуть мати незначні відхилення, переважно фінансово стійкі,

потенційно інвестиційно привабливі установи), банки з кількістю балів від 6,01 до 9,00 – як **банки класу А** (немає відхилень по всім показникам, або вони незначні, фінансово стійкі, інвестиційно привабливі установи).

Схожим чином ми можемо охарактеризувати і отримані кластери: **проблемні кластери** – такі, що містять проблемні, ризиковані банки, які не є інвестиційно привабливими, є фінансово нестійкими та ненадійними; **ризиковані кластери** – такі, що важко охарактеризувати однозначно як проблемні, проте вони містять переважно банки, що вразливі до різних видів ризику, їх інвестиційна привабливість та фінансова стійкість є низькою; **кластери класу Б** – такі, що містять банки, які не можна однозначно охарактеризувати як фінансово стійкі та ефективні, так як деякі їх показники мають значні відхилення, однак такі банки все ще мають більшість показників на рівні вище середнього і інвестиції в них є доцільними за умови вигідних умов щодо повернень на інвестиції (відсотків), банки з потенціалом росту; **кластери класу А** – такі що містять однозначно фінансово стійкі, ефективні та надійні банки за своїми показниками і є об'єктами рекомендованими для інвестування з метою отримання оптимального співвідношення дохід / ризик.

На основі Табл. 3.4 та проведеного аналізу кластерів та їх характеристик, можемо провести профілювання кожного з них. Жирним виділено ключові, на нашу думку, характеристики властиві кожному конкретному кластеру, що виділяють його від інших.

Кластер 1. Системно важливі банки з великими обсягами активів, що при цьому мають ***проблемні активи*** в своєму портфелі і можуть мати ***недостатній рівень капіталізації***. 3 з 4 банків є державними установами. Банки очікувано є стійкими в динаміці. Кластер **класу Б**. На користь таких банків може працювати принцип «надто великі, щоб впасти». Системність установ з даного кластеру зумовлює урядову підтримку, тому незважаючи на відхилення показників від оптимуму, дані банки можна вважати надійними.

Кластер 2. Першокласні банки, що мають ***відмінні показники фінансового стану***, і є стійкими з точки зору волатильності показників. **Кластер класу А**.

Кластер 3. Малі банки, що ведуть активну кредитно-інвестиційну політику і які мають *проблеми з ліквідністю*. Коливання показників з часом помірні. Більшість неактивних банків в кластері були визнані неплатоспроможними через втрату ліквідності. Кластер є **ризикованим**.

Кластер 4. Невеликі, *дохідні* банки, що ведуть *помірну кредитно-інвестиційну діяльність і обслуговують переважно юридичних осіб*. Спостерігаються *проблеми з ліквідністю та якістю кредитного портфелю*. Висока волатильність показників в часі. Кластер можна охарактеризувати як **ризикований**.

Кластер 5. Переважно великі структури, *низькодохідні банки*, які мають *відхилення практично за усіма показниками діяльності і ведуть пасивну кредитно-інвестиційну діяльність* («холодні банки»). Волатильність показників в часі висока. Неактивні банки з даного кластеру закривались з різних причин, серед яких втрата ліквідності та недостатня капіталізація. Кластер можна охарактеризувати як **проблемний**.

Кластер 6. Малі та дуже малі банки, *здебільшого корпоративні або інвестиційні банки*. 7 з 18 банків мають високу частку *проблемних активів в портфелі*, більше половини банків можуть мати *проблеми з капіталізацією*. Висока волатильність показників в часі. **Ризикований** кластер.

Кластер 7. Здебільшого дуже малі, *переважно дохідні банки*, за своїми характеристиками дещо схожі на банки з кластеру 6, проте ведуть менш активну, *переважно пасивну кредитно-інвестиційну політику*. Мають *проблемні активи в портфелі*. Відносно невисока волатильність показників в часі. **Ризикований** кластер.

Кластер 8. Малі та дуже малі, здебільшого *високоєфективні, відносно стійкі та надійні банки*. В цілому такі банки *потенційно могли б бути банками класу А*, проте мають ряд негативних аспектів: деякі з банків мають *високу частку проблемних активів в портфелі та ризиковану структуру ресурсної бази*. Кластер класу Б.

Кластер 9. Дуже малі банки, що сфокусовані на обслуговування юридичних осіб, характеризуються *низькою рентабельністю або збитковістю, слабкою ресурсною базою*, деякі банки мають *суттєву частку неякісних активів*. Суттєва волатильність показників. Сукупність показників може вказувати на кептивність або специфічність роботи банків з даного кластеру (як наприклад Український банк реконструкції та розвитку, що є технічним банком). **Ризикований** кластер.

Кластер 10. Банки з *іноземним капіталом, високоефективні та фінансово стійкі банки, банки класу А*, проте мають *незначні відхилення* в показниках, що характеризують *ресурсну базу, рівень капіталізації, рентабельності*, що може витікати зі специфіки діяльності банків як дочірніх структур. Банки стабільні в динаміці. **Кластер класу А.**

Кластер 11. Малі, *«холодні» банки. Слабка ресурсна база* (про що, крім іншого, може свідчити високий рівень коефіцієнта автономії), *низька рентабельність, висока частка проблемних активів, ліквідність менше за оптимум*. Даний кластер можна охарактеризувати як **проблемний**, може містити кептивні (кишенькові) установи. Більшість неактивних банків були ліквідовані з причин порушення вимог прозорості або порушення законодавства.

Кластер 12. Схожий на кластер 11, однак банки з кластеру *ведуть активну, навіть ризиковану кредитно-інвестиційну політику, будучи при цьому збитковими і маючи проблемні активи в портфелі*. Високий рівень коефіцієнта автономії може свідчити про слабку ресурсну базу. Даний кластер можна охарактеризувати як **проблемний**, може містити кептивні (кишенькові) установи. Серед неактивних банків в кластер так само як і у випадку з кластером 11 потрапляли такі, що були ліквідовані через фіктивні операції.

Кластер 13. Схожий на кластери 11 та 12, *однак банки є рентабельними і, з іншого боку, мають проблеми з ліквідністю*. Високий рівень коефіцієнта автономії може свідчити про слабку ресурсну базу. Даний кластер можна охарактеризувати як **ризикований**, може містити кептивні (кишенькові) установи.

Кластер 14. Малі банки, що можуть мати інвестиційний потенціал, однак є *збитковими, характеризуються ризикованою структурою ресурсної бази*.

Високий показник ліквідності може свідчити про низькі обсяги активних, дохідних операцій та активів. Висока волатильність показників в часі. Кластер можна віднести до **ризикованих**.

Кластер 15. Даний кластер не містить активних банків на досліджуваний період. Основними причинами закриття неактивних банків, що потрапили до даного кластеру є втрата ліквідності та недостатня капіталізація. Очевидно, що даний кластер є **проблемним**.

Аналізуючи всі отримані кластери в контексті їх класів (А, Б, ризиковані, проблемні) можна зробити наступні висновки по кожному з цих класів:

- Першокласні банки згруповані в кластерах №2 та №10, вони мають відмінні показники фінансового стану, надійності, фінансової ефективності. Такі банки мають найвищу інвестиційну привабливість і є надійними як для кредитування, так і для вкладення коштів (інвестування). Банки з кластеру №10 мають незначні відхилення деяких показників від оптимуму, однак це може бути спричинене специфікою діяльності даних банків як структур з іноземним капіталом. Банки з групи є стабільними з точки зору динаміки показників своєї діяльності. До групи входять 14 банків та жодного неплатоспроможного банку на досліджуваний період.
- Клас Б мають 2 кластери: №1, №8. Перший та восьмий кластери обидва характеризуються проблемами з кредитним портфелем, проте перший кластер містить системні банки з значно більшими обсягами загальних активів, другий кластер характеризується високою волатильністю показників в часі. До групи входять 16 банків та 1 неплатоспроможний банк на досліджуваний період.
- Клас «ризикований» мають 7 кластерів: №3 (проблеми з ліквідністю, активна кредитно-інвестиційна політика), №4 (проблеми з ліквідністю та якістю кредитного портфелю, помірна кредитно-інвестиційна політика), №6 (проблеми з якістю кредитного портфелю та капіталізацією), №7 (проблеми з якістю кредитного портфелю, пасивна кредитно-інвестиційна

політика), №9 (проблеми з якістю кредитного портфелю та рентабельністю), №13 (проблеми з ліквідністю та фінансові ризики: слабка ресурсна база, високий рівень резервів при низькій частці проблемних активів, пасивне нарощення активів), №14 (слабка ресурсна база, збиткові). Більшість кластерів є волатильними з точки зору динаміки показників. Кластери №3 та №13 мають стабільніші показники, однак зважаючи на те, що ці показники знаходяться не на оптимальному рівні, така стабільність може свідчити про відсутність поліпшення в функціонуванні установ протягом останніх 3-х років. Група містить 35 банків та 28 неплатоспроможних установ на досліджуваний період.

- Клас «проблемний» мають 4 кластери: №5 (здебільшого низька рентабельність, дуже велика частка проблемних кредитів в портфелі, слабка ресурсна база, «холодні банки»), №11 (те ж саме, що і в попередньому кластері, додатково характеризується низькою ліквідністю та ризикованою кредитно-інвестиційною політикою), №12 (те ж саме, що в кластері №11, додатково характеризується збитковістю установ), №15 (містить виключно неплатоспроможні банки). Група містить 5 банків та 45 неплатоспроможних установ на досліджуваний період. Волатильність показників в часі висока.

Якщо аналізувати саме розміщення отриманих кластерів на проекції UMAP, то можемо зробити висновок про досить високий рівень збереження глобальної структури даних. Так, банки, що можна віднести до «ризикованої» групи знаходяться в центральній частині карти, це найбільше скупчення кластерів на проекції. Проблемні установи тяжіють до верхнього лівого кута проекції. Першокласні банки та банки класу «Б» знаходяться в протилежних частинах проекції: це нижній лівий кут та права частина проекції. Ці дві групи банків не поєднані, тобто не знаходяться поруч, через відмінності в своїх показниках, власне через які вони й буди віднесені до різних кластерів. Так, наприклад, першокласні кластери №2 та №10 мають різні обсяги загальних активів: в той час як банки

кластеру №2 мають сукупно 33,5% активів, то банки кластеру №10 мають тільки 2,1%, а тому були спроектовані в область з банками зі схожими обсягами активів.

Отримавши достатньо адекватні результати в процесі інтерпретації кластерів, ми можемо дійти висновку, що розділення українських комерційних банків (їх кластеризація), на основі широкого спектру показників їх діяльності можлива, і ефективні, інвестиційно привабливі банки можуть бути відокремлені від проблемних.

Що стосується питання вибору алгоритму зниження розмірності, навіть в результаті проведених досліджень і тестів, складно однозначно зробити висновок стосовно переваги алгоритму UMAP чи t-SNE. Однак, ми скоріше схильні заперечити ствердження щодо того, що UMAP є найбільш підходящим алгоритмом для зниження розмірності даних в контексті дослідження фінансової стійкості вітчизняних комерційних банків, з точки зору економічної інтерпретації та якості кластероутворення. Ми обґрунтовуємо це тим, що алгоритм t-SNE, після оптимізації його гіпер-параметрів, в цілях візуалізації та подальшої економічної інтерпретації отриманої двовимірної проекції підходить навіть краще, ніж UMAP. Хоча саме для кластерного аналізу UMAP все ж таки виявився більш ефективним.

Лістинг коду для Python-реалізації всіх етапів дослідження, починаючи від завантаження та первинної обробки даних і до етапу проведення кластеризації даних можна знайти в Додатку Д.

Повний перелік банків з розподілом по кластерах можна знайти в Додатку Е.

Важливою частиною останнього кроку нашого дослідження є перевірка адекватності побудованої системи моделей та алгоритмів. Специфіка задачі кластеризації, на відміну від задачі класифікації, полягає в тому, що ми не маємо певних еталонних значень, з якими ми могли б порівнювати отриманні в процесі моделювання результати. Однак, методи для оцінки якості кластеризації існують, і широко застосовуються такі метрики як силует (застосовується частіше за інші), оцінки Калінського-Харабаза та Девіса-Болдуїна. Дані метрики ми вже розглядали детальніше та застосовували в попередніх розділах. Для метрики силуету, значення якої можуть бути на проміжку від -1 до 1, вважається, що додатні значення свідчать

про адекватність кластеризації, але при наближенні значення до нуля, вважається що кластери перетинаються між собою. Для двох інших метрик взагалі не визначено яких-небудь оптимальних інтервалів чи граничних значень. Тому, перевірка якості нашої системи моделей буде складатись з двох кроків. По-перше, ми визначимо як відрізняється якість при застосуванні обраного нами алгоритму кластеризації на нашому наборі даних по комерційних банках та на двох інших синтетичних наборах даних. Візуальне представлення результатів кластеризації при пошуку 2-х кластерів відображено на Рис. 3.7. По-друге, ми проведемо порівняльний аналіз запропонованої нами системи моделей і алгоритмів з іншими найбільш поширеними та вживаними алгоритмами на основі зазначених вище метрик.

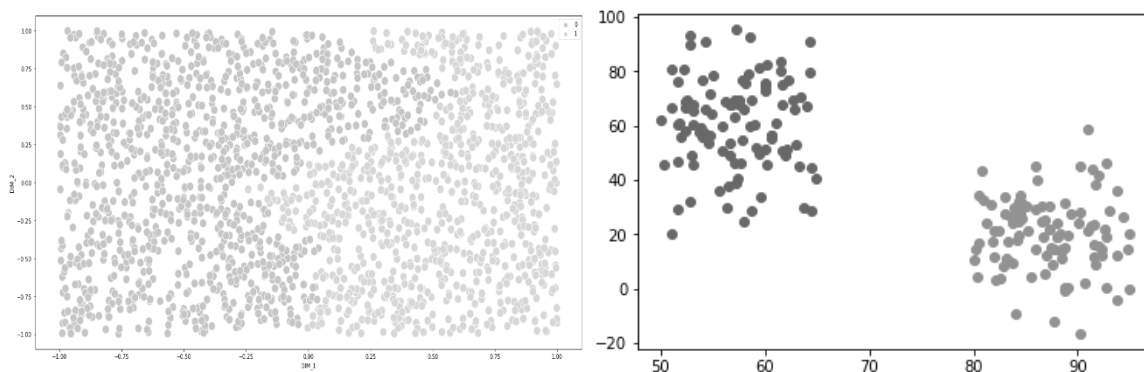


Рис. 3.7. Візуальна репрезентація результатів роботи алгоритму агломеративної ієрархічної кластеризації методом Уорда на основі двох синтетичних наборів даних: рівномірно розподілені дані (ліва частина) та два кластери, розділення яких наближене до ідеального (права частина)

Джерело: побудовано автором

На рис. 3.7. представлені 2 синтетичні набори даних: набір з 200 точок, що формують 2 кластери, розділення яких один від одного наближене до ідеального а щільність точок яких є високою; набір з 2000 точок, що рівномірно розподілені і не утворюють кластерів.

Результати тестування представлені в Табл. 3.5.

Відповідно до Табл. 3.5, якість кластеризації по даним з нашого дослідження є наближеною до оптимальних значень, відповідно до показників силуету та Девіса-Болдуїна, а оцінка Калінського-Харабаза є значно вищою, що може бути

спричинене більшою кількістю точок в наборі даних та більшою кількістю кластерів в даних. Таким чином, можемо стверджувати про достатньо високу якість кластеризації алгоритму агломеративної ієрархічної кластеризації методом Уорда, для роботи з даними по вітчизняним комерційним банкам.

Таблиця 3.5.

Оцінка якості кластеризації алгоритму агломеративної ієрархічної кластеризації методом Уорда для рівномірно розподілених даних, даних по комерційним банкам за 2015-2021 роки та по «ідеально» розділеним кластерам

	Силует	Девіс-Болдуїн	Калінський-Харабаз
Рівномірний розподіл	0,31	1,25	1031
Дані дослідження	0,51	0,69	5384
«Ідеальні» кластери	0,63	0,51	1107

Джерело: розрахунки автора

Маючи уявлення про певні оптимальні значення застосовуваних метрик, можемо провести порівняльний аналіз різних підходів до кластерного аналізу, різних алгоритмів зниження розмірності та кластеризації, їх комбінацій. Для цього нами було відібрано 6 широко вживаних алгоритмів зниження розмірності (ядерний метод головних компонент (Kernel PCA), метод багатовимірного шкалювання (MDS), ізометричного проектування (ISOMAP), спектрального проектування (Spectral Embedding), методи t-SNE та UMAP) та 7 алгоритмів кластеризації (K-середнє (K-means), Mean Shift, DBSCAN, спектральна кластеризація (Spectral Clustering), BIRCH, GMM та ієрархічна агломеративна кластеризація методом Уорда). Також, ми протестували всі зазначені алгоритми кластеризації напряду на наборі багатовимірних даних, щоб переконатись у необхідності застосування зниження розмірності. Результати аналізу представлені в Табл. 3.6.

Таблиця 3.6.

Порівняльна характеристика якості кластеризації для комбінацій алгоритмів зниження розмірності та кластеризації

	K-means			Mean Shift			DBSCAN			Spectral Clustering		
	silhouette_score	davies_bouldin	calinski_harabasz	silhouette_score	davies_bouldin	calinski_harabasz	silhouette_score	davies_bouldin	calinski_harabasz	silhouette_score	davies_bouldin	calinski_harabasz
Високорозмірні дані	0,17	1,22	250	0,11	0,71	45	0,55	0,34	46,2	0,35	0,53	176
Kernel PCA	0,36	0,8	3405	N/A	N/A	N/A	N/A	N/A	N/A	0,25	0,88	1567
MDS	0,32	0,81	1039	0,3	0,49	291	0,77	23	0,66	0,83	0,1	53
ISOMAP	0,38	0,72	1908	0,27	0,53	464	0,55	0,34	46	0,35	0,53	176
Spectral embedding	0,41	0,73	5782	N/A	N/A	N/A	N/A	N/A	N/A	0,1	0,72	489
t-SNE	0,42	0,73	2651	0,57	0,48	12077	0,27	1,1	1187	N/A	N/A	N/A
UMAP	0,5	0,69	4889	0,48	0,68	3798	0,39	2,05	141	0,49	0,59	4803
	BIRCH			GMM			Ward Hierarchical Clustering					
	silhouette_score	davies_bouldin	calinski_harabasz	silhouette_score	davies_bouldin	calinski_harabasz	silhouette_score	davies_bouldin	calinski_harabasz			
Високорозмірні дані	N/A	N/A	N/A	0,01	2,33	147	0,14	1,36	245			
Kernel PCA	N/A	N/A	N/A	0,35	0,84	3093	0,3	0,87	2765			
MDS	0,3	0,65	1481	0,3	2,28	668	0,27	0,94	873			
ISOMAP	0,33	0,64	2279	0,34	0,84	1467	0,34	0,77	1608			
Spectral embedding	N/A	N/A	N/A	0,52	0,56	1234	0,39	0,75	4901			
t-SNE	0,45	0,32	35475	0,38	0,72	2130	0,4	0,79	2329			
UMAP	0,49	0,67	4650	0,4	0,8	4057	0,51	0,69	5384			

Джерело: розрахунки автора

Як можна побачити з Табл. 3.6, алгоритми кластеризації при застосуванні на багатовимірних даних показують незадовільні результати по якості кластеризації. DBSCAN та спектральна кластеризація показують дещо кращі результати, однак з усіх алгоритмів кластеризації тільки спектральний метод є підходящим для роботи з багатовимірними даними, так як пристосований до специфіки таких даних [194, Додаток 22], в той час як DBSCAN скоріше за все дає хибно-позитивні результати. Однак очевидно, що результати роботи алгоритмів із застосуванням процедури зниження розмірності дають більш якісні результати.

При застосуванні алгоритмів зменшення розмірності, деякі алгоритми кластеризації не в змозі були визначити кластери в двовимірному наборі даних, тобто всі дані були віднесені до одного кластеру, або у випадку алгоритму DBSCAN – всі данні були визначені як шум. В таких випадках в таблиці вказується N/A.

З Табл. 3.6 можемо побачити, що алгоритми t-SNE та UMAP забезпечують найвищу якість, при кластеризації двовимірних проєкцій отриманих за їх використання. Для того, щоб переконатись, що саме система моделей UMAP + ієрархічна кластеризація є найефективнішою, нами був проведений більш точний та вибірковий порівняльний аналіз. Як ми вже зазначали і доводили раніше, двовимірні проєкції, отримані за допомогою алгоритму t-SNE погано піддаються кластеризації. Ми також можемо побачити на Рис.3.8, що двовимірні проєкції, отримані за допомогою алгоритмів Kernel PCA, MDS, ISOMAP, Spectral Embedding неможливо кластеризувати, отримавши при цьому достатню якість кластеризації, так як утворені проєкції не містять груп точок, що добре відділяються одна від одної.

Так, при кластеризації утворених проєкцій, відображених на Рис. 3.6, незалежно від кількості утворених кластерів, якість кластеризації буде низькою, а визначені кластери не будуть містити ніякого економічного сенсу.

Таким чином ми можемо спростити порівняльний аналіз до тестування зазначених вище алгоритмів кластеризації на двовимірній проєкції даних, отриманих за допомогою алгоритму зниження розмірності UMAP.

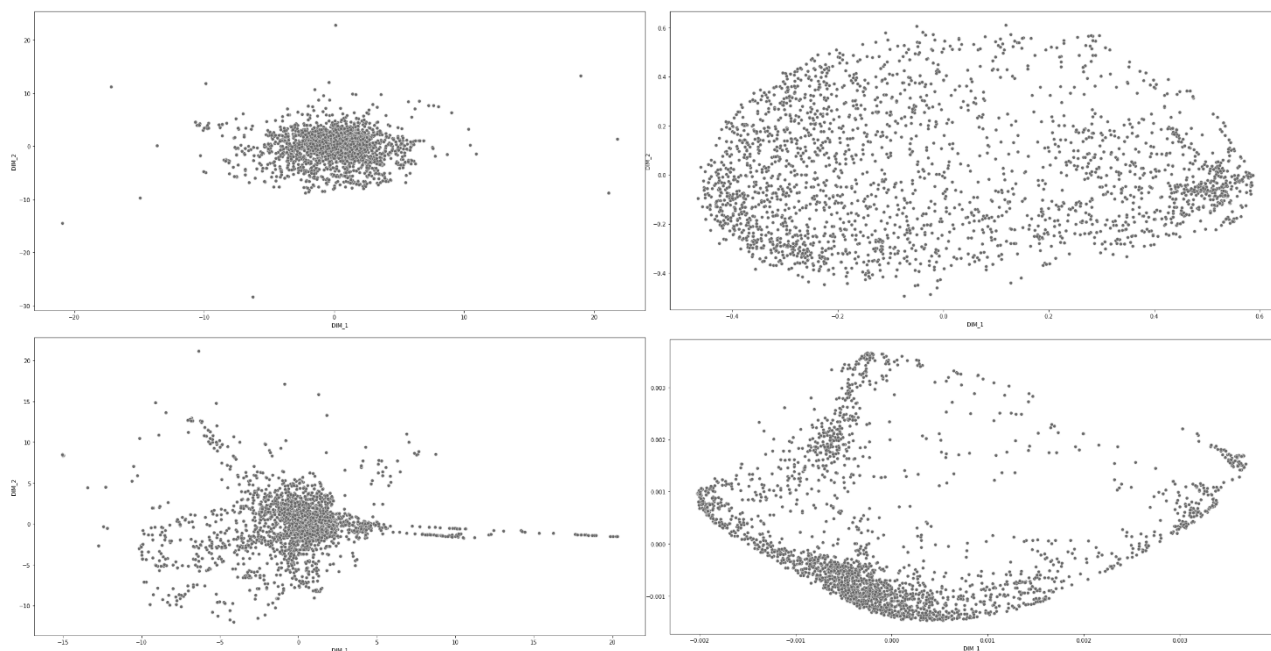


Рис. 3.8. Візуальна репрезентація результатів роботи алгоритмів MDS (зверху зліва), Kernel PCA (зверху справа), ISOMAP (знизу зліва), Spectral Embedding (знизу справа)

Джерело: побудовано автором

Тестування проводилося на основі тих самих трьох метрик силуету, Девіса-Болдуїна та Калінські-Харабаза. Для кожного з алгоритмів було проведено ряд експериментів з різними гіпер-параметрами, і визначено конфігурації наближені до оптимальних. Для цього проводилися послідовні тестування для кожного алгоритму, де кожен гіпер-параметр змінювався на певному числовому проміжку і робилися заміри значень всіх метрик якості кластеризації. В результаті для кожного алгоритму відбиралася така конфігурація. Що забезпечувала б найвищу якість кластеризації даних в рамках нашого дослідження. Для тих алгоритмів, що потребують встановлення кількості кластерів на вході, ми встановлювали таку кількість кластерів, що була визначена як оптимальна на етапі тестування конфігурацій алгоритму агломеративної ієрархічної кластеризації, тобто 15 кластерів. Ми вважаємо такий підхід цілком адекватним, адже алгоритми при порівнянні застосовуються до одного й того ж самого набору даних і кількість кластерів в цих даних ніяк не залежить від обраного алгоритму. Результати порівняльного аналізу представлені в Табл. 3.7.

Таблиця 3.7

**Порівняльна характеристика якості кластеризації двовимірної проекції
UMAP та їх зважене ранжування**

Алгоритм	Силует	Девіс-Болдуїн	Калінські-Харабаз	Ранг
Ієрархічна кластеризація	0,51	0,69	5384	6,2
K-means	0,5	0,69	4889	4,8
Spectral Clustering	0,49	0,59	4803	4,8
BIRCH	0,49	0,67	4650	4,4
Mean Shift	0,48	0,68	3798	3,2
GMM	0,4	0,8	4057	2,2
DBSCAN	0,39	2,05	141	1

Джерело: розрахунки автора

У Табл. 3.7 можемо побачити, що для простішого порівняння протестованих алгоритмів кластеризації, нами було застосовано рангову систему. Вона є ідентичною тій ранговій системі, що застосовувалась для визначення оптимального значення кількості кластерів для алгоритму агломеративної ієрархічної кластеризації. Ранжування є зваженим, тобто метрики мають різну вагу і ранг обчислюється на основі зважених результатів по кожній метриці. Метрика силуету отримала вагу 0,6 як найбільш широко вживана метрика для оцінки якості кластеризації, інші дві метрики отримали однакові ваги – по 0,2.

На основі проведеного порівняльного аналізу якості кластеризації різних алгоритмів, можемо стверджувати, що система моделей з алгоритму зниження розмірності UMAP та алгоритму ієрархічної агломеративної кластеризації методом Уорда є найбільш ефективною, маючи зважений ранг 6,2.

Наостанок відзначимо також деякі перспективи подальшого розвитку в даному напрямку досліджень. Наразі, спроектована та побудована система моделей та алгоритмів зниження розмірності та кластеризації працює на основі вхідних файлів з статистичними даними, і хоча результат за незмінних встановлених гіпер-параметрів буде видавати подібні результати і їх інтерпретація не буде складною, однак залучення людського фактору все ще є необхідним. Тому ключовим напрямком розвитку та вдосконалення ми вважаємо саме проектування та

реалізацію інформаційної системи з наступними характеристиками та властивостями:

- Повноцінний користувацький інтерфейс, що дозволяє користувачам вводити певні дані до програми, на основі яких інформаційна система буде проводити моделювання, або вносити коригування до існуючої моделі;
- Додаткові елементи користувацького інтерфейсу, що полегшують роботу з інформаційною системою, наприклад: додавання певних банків у «збережені», збереження та відтворення результатів останнього моделювання, експорт результатів моделювання в найбільш поширених форматах типу .PDF чи .CSV;
- Внутрішня база банківських даних та її автоматичне оновлення: інформаційна система зберігає показники банківської діяльності та супровідну банківську інформацію у внутрішніх базах даних, та регулярно оновлює ці дані через зовнішній інтерфейс програмування застосунку (Application Programming Interface – API);
- Модульна система інформаційної системи, що сприяє легкості розробки, підтримки та застосування додаткових під-програм (під-систем, модулів). Наприклад, це можуть бути модулі для визначення курсів валют, відсоткових ставок за депозитами та кредитами, модулі доступу до нормативної документації, новин банківського сектору чи окремих банків. Це також можуть бути модулі побудовані на основі додаткових моделей, наприклад модуль стрес-тестування тощо;
- Вбудована рекомендаційна система на базі засобів машинного навчання, експертних систем чи/та нейронних мереж. Така система матиме змогу аналізувати отримуванні результати кластерного аналізу та/або результати багатовимірної проектування в реальному часі, та проводити економічну інтерпретацію і надавати рекомендації щодо банків з кожного з отриманих кластерів.

Побудова інформаційної системи такого рівня в цілому є складним процесом, потребує попереднього детального, ретельного проектування, передбачає

створення відповідної архітектури, проектування баз даних, визначення та обґрунтування необхідних ресурсів тощо. Відзначимо, що реалізація проекту можлива повністю на базі мови програмування Python.

Побудовану та протестовану вище систему моделей можна з такою ж ефективністю застосовувати до інших часових періодів, наприклад до наборів статистичних даних до 2008 року (до початку потужної світової фінансової кризи) для аналізу стану банківського сектору в цілому та окремих комерційних банків зокрема (історичний аналіз); або до періодів стабільності після завершення війни та переходу до періоду відновлення економіки, фінансового та банківського секторів для аналізу актуальної на певний момент ситуації. Ще одним перспективним напрямком розвитку даного дослідження є тестування побудованої системи моделей на даних по кризовим періодам (або на даних, що включають як кризові періоди, так і періоди відновлення та стабільності) та її модифікація для врахування додаткових факторів, що притаманні саме таким періодам. Теоретично, це можна реалізувати через калібрування вже включених в систему моделей на нових даних та оптимізацію їх гіпер-параметрів та/або через включення в систему додаткової моделі (або моделей), що відповідатиме за аналіз саме факторів «форс-мажору» / «нестабільності».

Висновки до розділу 3

В третьому розділі дисертаційної роботи було проведене комплексне експериментальне дослідження. Проведено аналіз набору вхідних даних на предмет їх придатності до кластеризації. Одночасно з цим проведено тестування різних методів первинної обробки даних, таких як стандартизація, нормалізація, вінзорування та деяких інших, а також їх комбінацій. В результаті визначено, що алгоритм t-SNE ефективніше проектує дані, що трансформовані за допомогою вінзорування з верхнім та нижнім лімітами в 5% та 2% відповідно та класичною стандартизацією. UMAP генерує проєкцію, що забезпечує найбільш якісну кластеризація при застосуванні на даних, оброблених за допомогою вінзорування з

верхнім та нижнім лімітом в 1% та робустною стандартизацією. Аналіз даних показав, що дані, оброблені вказаними методами є неоднорідними та мають високу тенденцію, тобто містять достатню кількість внутрішніх груп (кластерів) для того, щоб застосування кластерного аналізу надало адекватні результати.

Визначено оптимальні гіпер-параметри для алгоритмів t-SNE, UMAP та для алгоритму агломеративної ієрархічної кластеризації, таких що забезпечували б найвищу роздільну здатність багатовимірних даних, баланс між глобальною та локальною структурою даних, високу якість кластеризації, адекватність економічної інтерпретації результатів

Проведено кластеризацію вітчизняних комерційних банків, при цьому успішно відділено інвестиційно привабливі установи від потенційно проблемних. Виділено 15 кластерів, що умовно можна розділити на 4 класи: клас А (першокласні банки, що характеризуються високою фінансовою стійкістю, надійністю, фінансовою ефективністю), що включає два кластери; клас Б (потенційно інвестиційно привабливі установи, достатньо стійкі та надійні однак мають ряд проблемних аспектів, що не дають їм потрапити до кластерів класу А) складається з 2 кластерів; клас «ризикований» (містить установи, що мають суттєві відхилення по декількох показниках, що не дає однозначно охарактеризувати їх як проблемні, однак робить їх вразливими до різних видів ризиків, а їх фінансова стійкість, ефективність та надійність є низькою), що складається з 7 кластерів; та клас «проблемний» (містить установи, що мають суттєві відхилення по більшості своїх показників), що містить 4 кластери.

Підтверджено, що банки які знаходяться у багатовимірному просторі різнопланових ознак, що описують їх роботу, знаходяться поруч з подібними собі. Саме з цієї причини вдалося отримати кластери з адекватним розподілом банків між ними на основі спроектованої двовимірної карти.

Досліджено поведінку комерційних банків в динаміці. Доведено, що якщо за останні декілька років банк демонструє активні переміщення на двовимірній проекції – це можна розглядати як тривожну ознаку і негативну характеристику фінансового стану та надійності банку.

Візуально продемонстровано, що проблемні банки формують цілу окрему область на двовимірній карті і добре дискримінуються від фінансово ефективних, надійних банків, так як розміщуються на двовимірній проекції у діаметрально протилежній частині проекції.

Проведено ряд експериментів різних моделей та алгоритмів та їх порівняльний аналіз на основі метрик силуету, Калінського-Харабаза та Девіса-Болдуїна. Визначено та доведено, що якість кластеризації алгоритму ієрархічної агломеративної кластеризації методом Уорда знаходиться на достатньо високому рівні. Доведено, що запропонована система алгоритмів є найбільш ефективним рішенням для вирішення задачі аналізу вітчизняних комерційних банків.

Таким чином була досягнута мета нашого дослідження та виконані відповідні задачі. Для її досягнення реалізовано комплекс етапів (pipeline) в середовищі Google Collab мовою програмування Python 3.6, що складається із збору даних, їх попередньої обробки, відбору ознак в модель, підбору гіпер-параметрів, зниження розмірності та кластеризації даних.

Основні результати Розділу 3 опубліковано у наукових працях автора [183, 184, 185, 186, 187, 188].

ВИСНОВКИ

У дисертаційній роботі представлено рішення науково-практичної проблеми аналізу фінансового стану комерційних банків України із застосуванням моделей та алгоритмів машинного навчання, а саме кластеризації та зниження розмірності, для підтримки управлінських рішень щодо вибору комерційного банку як об'єкту інвестування. Для цього запропоновано новий комплексний ґрунтовний методологічний підхід до аналізу фінансового стану вітчизняних банківських. В результаті проведеного дослідження можна сформулювати наступні висновки:

1. Розглянуто стан банківського сектору в досліджуваний період з 2015 по 2021 рік, а також поточний для визначення актуальності проведення дослідження щодо аналізу фінансового стану вітчизняних комерційних банків. Визначено, що зважаючи на суттєву мінливість, динамічність, нестійкість фінансового сектору в цілому та банківського зокрема, а також враховуючи поточний стан війни в державі та очікуваний період відновлення після її закінчення – мета дослідження зберігатиме актуальність протягом тривалого часу від моменту його проведення.

2. Сформоване теоретичне підґрунтя на основі детального аналізу предметної області та понятійного апарату. Систематизовано такі ключові поняття як фінансовий стан, ефективність діяльності, стійкість, надійність, платоспроможність, невизначеність та ризик, прибутковість. Запропоновано авторські визначення таких понять як фінансовий стан комерційного банку, ефективність банківської діяльності, фінансова стійкість банківської установи. Представлене авторське бачення взаємозв'язку розглянутих категорій, що є контекстом для дослідження, що проводиться.

3. Проведено глибокий якісний аналіз банківських ризиків, розглянуто основні класифікації ризиків на основі низки нормативних джерел, галузевих праць та інших джерел, побудовано авторську розширену класифікацію ризиків, досліджено взаємозв'язки між різними видами ризиків.

4. Розглянуто найбільш широко вживані вітчизняні та іноземні класичні моделі дослідження ризику підприємств в цілому та банківських установ зокрема,

також розглянуто комплексні моделі аналізу, окреслено їх основні недоліки. Запропоноване застосування алгоритмів та методів машинного навчання як більш універсальні, інноваційні, відносно прості в реалізації, такі що мають високу якість візуалізації та сприяють простоті подальшої економічної інтерпретації результатів моделювання.

5. Розроблено та запропоновано загальний методологічний підхід до аналізу фінансового стану комерційних банків, що складається з 9 кроків, виконання яких дало змогу сформувати та реалізувати необхідну систему моделей для досягнення поставленої мети дослідження.

6. Сформовано початковий набір показників діяльності комерційних банків на основі низки джерел. В результаті, отримано набір показників, що складається з 63 показників, таких що можуть бути розраховані на основі відкритої статистичної звітності НБУ. Сформовано набір показників з 17 ознак в процесі комплексного підходу відбору ознак, що складався з аналітичної складової (кореляційний аналіз всіх пар ознак з відбором через метрики VIF, MAD та SPEC) та експертної складової (на основі ряду експериментів з додаванням та виключенням показників в модель, таким чином, що можна було б підсилити роздільну здатність кластеризації та загальну якість моделі і адекватність інтерпретації результатів).

7. Запропоноване застосування двох алгоритмів машинного навчання для зниження розмірності: t-SNE, як потужний алгоритм візуалізації, для підтримки візуально-статистичного аналізу двовимірної проекції, UMAP для підтримки кластеризації двовимірної проекції і подальшого кластерного аналізу; та алгоритму агломеративної ієрархічної кластеризації для власне проведення кластеризації. Розкрито логіку та етапи роботи алгоритмів та розглянуто математичний апарат, на якому вони побудовані. Для обраного алгоритму кластеризації розглянуто можливі варіанти розрахунку відстаней між спостереженнями та методи розрахунку відстаней між кластерами. Для розрахунку відстаней між кластерами обрано метод Уорда. Виділено та розглянуто основні метрики для оцінки якості кластеризації, порядок їх розрахунку та оптимальні значення.

8. Спроектовано та реалізовано систему моделей, що відповідає за двовимірне проектування багатовимірних даних та кластеризацію двовимірних проекцій, проведено оптимізацію всіх алгоритмів на основі ряду експериментів з гіпер-параметрами. Випробувано різні методи первинної обробки даних, в результаті чого визначено, що застосування t-SNE ефективніше за використання вінзорування з верхнім та нижнім лімітами в 5% та 2% відповідно та класичної стандартизації; застосування UMAP при застосуванні вінзорування з верхнім та нижнім лімітом в 1% та робустної стандартизації.

9. Проведено аналіз набору вхідних даних на предмет їх придатності до кластеризації, доведено що дані, оброблені вказаними методами є неоднорідними та мають високу тенденцію, тобто містять достатню кількість внутрішніх груп (кластерів) для того, щоб застосування кластерного аналізу надало адекватні результати.

10. Проведено аналіз адекватності запропонованого методологічного підходу та побудованої системи моделей через експериментальне підтвердження висунутих у Вступі гіпотез:

- Підтверджено, що банки які знаходяться у багатовимірному просторі різнопланових ознак, що описують їх роботу, знаходяться поруч з подібними собі. Саме з цієї причини вдалося отримати кластери з адекватним розподілом банків між ними на основі спроектованої двовимірної карти. Отже, відібрані алгоритми здатні на достатньому рівні зберігати локальну та глобальну структуру даних.
- Візуально продемонстровано, що проблемні банки формують цілу окрему область на двовимірній карті і добре дискримінуються від фінансово ефективних, надійних банків, так як розміщуються на двовимірній проекції у діаметрально протилежній частині проекції.
- Доведено, в рамках аналізу динаміки банківських показників, що якщо за останні декілька років банк демонструє різкі або постійні зміни (коливання) на двовимірній проекції – це можна розглядати як тривожну ознаку і негативну характеристику фінансового стану та надійності банку.

11. Проведено комплексне експериментальне дослідження з метою аналізу фінансового стану вітчизняних комерційних банків. Проведена економічна інтерпретація отриманих результатів та здійснене профілювання комерційних банків згідно отриманим результатам алгоритму кластеризації. Виділено 15 кластерів в рамках кластерного аналізу, які умовно можна профілювати на 4 групи: клас А (високі фінансова ефективність, надійність, інвестиційна привабливість, низькі ризики, оптимальний фінансовий стан), клас Б (переважно фінансово ефективні, потенційно інвестиційно привабливі і мають помірний рівень ризику, фінансовий стан задовільний), ризиковані (вразливі до різних видів ризику, надійність та фінансова ефективність низька або нижче середнього, незадовільний фінансовий стан), проблемні (низькі надійність, фінансова ефективність, високі ризики, поганий фінансовий стан).

12. Проведено порівняльний аналіз ряду найбільш вживаних алгоритмів зниження розмірності та кластеризації на основі загальноприйнятих метрик оцінки якості кластеризації з метою визначення ефективності запропонованої системи алгоритмів та моделей. Експериментально доведено, що якість моделювання за використання алгоритму ієрархічної агломеративної кластеризації методом Уорда є високою а побудована система моделей та алгоритмів зниження розмірності та кластеризації є найбільш ефективним рішенням для досягнення мети дослідження: проведення аналізу фінансового стану вітчизняних комерційних банків. Порівняльний аналіз проводився на базі метрик силуету та оцінок Калінські-Харабаза та Девіса-Болдуїна. Алгоритми зниження розмірності Kernel PCA, Spectral Embedding, MDS та ISOMAP сформували двовимірні проекції непридатні до подальшої кластеризації, алгоритм t-SNE видає якісні результати для візуального аналізу двовимірної проекції, а за результатами зваженого ранжування на основі зазначених метрик, найкращий результат кластеризації досягається за використання алгоритмів UMAP та ієрархічної агломеративної кластеризації методом Уорда.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Основні показники діяльності банків України. *Національний банк України*. URL: <https://bank.gov.ua/ua/statistic/supervision-statist/data-supervision#1> (дата звернення: 18.05.2023).
2. Зрозуміти та відпустити: причини і результати "великого банківсько-го очищення". *Національний банк України*. URL: <https://badbanks.bank.gov.ua/> (дата звернення: 18.05.2023).
3. Про затвердження Комплексної програми розвитку фінансового сектору України до 2020 року : Постанова правління Нац. банку від 18.06.2015 р. № №391. URL: <https://bank.gov.ua/ua/news/all/kompleksna-programa-rozvitku-finansovogo-sektoru-ukrayini-do-2020-roku> (дата звернення: 18.05.2023).
4. Стратегія Національного банку України до 2025 року. *Національний банк України*. URL: <https://bank.gov.ua/ua/news/all/strategiya-natsionalnogo-banku-ukrayini> (дата звернення: 18.05.2023).
5. Ларіонова К. Л., Донченко Т. В. Управління ліквідністю банків України в сучасних умовах: нормативний аспект. *Вісник хмельницького національного університету*. 2020. Т. 280, № 2. С. 76–82. URL: <https://doi.org/10.31891/2307-5740-2020-280-2-14> (дата звернення: 18.05.2023).
6. Миненко Л. М. Трансформація підходів до оцінки ризику банківської ліквідності. *Ефективна економіка*. 2020. № 1. – URL: <http://www.economy.nayka.com.ua/?op=1&z=7573> (дата звернення: 18.05.2023). DOI: 10.32702/2307-2105-2020.1.54
7. Огляд банківського сектору, лютий 2023 року. *Національний банк України*. URL: <https://bank.gov.ua/ua/news/all/oglyad-bankivskogo-sektoru-lyutiy-2023-roku> (дата звернення: 18.05.2023).
8. Еркес О., Калита О., Сундук Т. Банківська система України в умовах війни. *SCIENTIA FRUCTUOSA (Вісник Київського національного торговельно-економічного університету)*. 2022. Т. 4, № 144. С. 122–133. URL: [https://doi.org/10.31617/1.2022\(144\)09](https://doi.org/10.31617/1.2022(144)09) (дата звернення: 18.05.2023).

9. Фінанси підприємства : підручник. 8-ме вид. / А. М. Поддєрьогін та ін. Київ : КНЕУ, 2013. 520 с.
10. Білик М.Д. Котов І. Л.. Фінансовий стан підприємства: теорія, методика. *Формування ринкових відносин в Україні*. 2013. № 5. С. 86—92.
11. Цал Цалко Ю. С. Фінансовий аналіз : підручник. Київ : Центр учб. літ., 2009. 566 с.
12. Чепка В. В., Свідерська І. М., Гавриленко Ю. О. Фінансовий стан підприємства: теоретичні основи. *Інвестиції: практика та досвід*. 2020. № 19-20. С. 96–102. DOI: 10.32702/2306-6814.2020.19-20.96
13. Загородній А. Г., Вознюк Г. Л. Фінансово-економічний словник. Львів : Львів. політехніка, 2005. 714 с.
14. Дзоблюк О. В., Михайлюк Р. В. Фінансова стійкість банків як основа ефективного функціонування кредитної системи : монографія. Тернопіль : Тернограф, 2009. 316 с.
15. Герасимович А. М., Алексеєнко М. Д., Парасій-Варгуненко І. М. Аналіз банківської діяльності : підручник. Київ : КНЕУ, 2004. 599 с.
16. Вінниченко О. В., Гудзь А. В. Фінансовий стан банку та методи його оцінки в Україні. *Вісник економіки транспорту і промисловості*. 2020. № 69. С. 217–228. URL: <https://doi.org/10.18664/338.47:338.45.v0i69.200560> (дата звернення: 18.05.2023).
17. Кузнецова С. А., Болгар Т. М., Пестовська З. С.. Банківська система : підручник. Київ, 2014. 400 с.
18. Шамрін Р. В. Розробка динамічної моделі діяльності комерційного банку. *Економіка та держава*. 2016. № 2. С. 42–45.
19. Суліма Н.М., Степасюк Л.М., Величко О.В. Економіка і фінанси підприємства: підручник. Київ : «ЦП «КОМПРИНТ»», 2015. 466 с.
20. Tomas C., Maurice S. C. Managerial Economics (The McGraw-hill Economics Series). McGraw Hill, 2004. 736 с.
21. Barr N. Economics of the welfare state. New York: Oxford University Press (USA), 2012. 412 p.

22. Sen A. Markets and freedom: Achievements and limitations of the market mechanism in promoting individual freedoms. *Oxford Economic Papers*. 1993. no. 45. P. 519 – 541.

23. Коваленко В. В., Вербицька Я. Є.. Оцінка ефективності функціонування банківської системи в умовах поглиблення глобалізаційних процесів. *Ефективна економіка*. 2012. №2. URL: <http://www.economy.nayka.com.ua/?op=1&z=953> (дата звернення: 18.05.2023).

24. Єріс Л. М. Ефективність банківської діяльності як результат якісного управління грошовими потоками. *Українська академія банківської справи НБУ*. 2014. №40. С. 109 – 115.

25. Аналіз банківської діяльності: підручник. Вид.-2-ге, без змін. / А. М. Герасимович та ін. Київ : КНЕУ, 2013. 600 с

26. Парасій-Вергуненко І. М. Аналіз банківської діяльності: навч.-метод. посібник для самост. вивч. дисц. Київ : КНЕУ, 2003. 347 с.

27. Примостка Л. О. Аналіз банківської діяльності: сучасні концепції, методи та моделі : монографія. Київ : КНЕУ, 2002. 316 с.

28. Криклій О.А., Маслак Н.Г.. Управління прибутком банку: монографія. Суми : ДВНЗ "УАБС НБУ", 2008. 136 с.

29. Катан Л. І., Плахотин М. С.. Прибутковість комерційного банку та її підвищення. *Агросвіт*. 2018. №21. С. 3–6. URL: http://nbuv.gov.ua/UJRN/agrosvit_2018_21_2 (дата звернення: 18.05.2023).

30. Гапак Н. М., Капштан С. А.. Особливості визначення фінансової стійкості підприємства. *Науковий вісник Ужгородського університету*. 2014. №1. С. 191–196. URL: http://nbuv.gov.ua/UJRN/Nvuues_2014_1_39 (дата звернення: 18.05.2023).

31. Савицька Г.В. Економічний аналіз діяльності підприємства : навч. Посібник. Київ : Знання, 2005. 662 с.

32. Банківська енциклопедія : енциклопедія / С. Г. Арбузов та ін. Київ : Центр наук. дослідж. Нац. банку України: Знання, 2011. 504 с.

33. Довгаль Ю.С., Чамара Р.О. Сутність фінансової стійкості комерційного банку та ефективні шляхи її забезпечення. *Фінансовий простір*. 2015. №1. С. 130-136. URL: http://nbuv.gov.ua/UJRN/Fin_pr_2015_1_18 (дата звернення: 18.05.2023).
34. Рисін В. В., Біда А. П.. Фінансова стійкість банку: чинники та особливості забезпечення. *Ефективна економіка*. 2021. №3. DOI: 10.32702/2307-2105-2021.3.1.
35. Дзюблюк О.В., Михайлюк Р.В. Фінансова стійкість банків як основа ефективного функціонування кредитної системи : монографія. Тернопіль: ТЗОВ «Терно-граф», 2009. 316 с.
36. Кочетков В. М. Забезпечення фінансової стійкості сучасного комерційного банку: теоретико-методологічні аспекти : монографія. Київ : КНЕУ, 2002. 238 с.
37. Регулювання діяльності банків в умовах глобальних викликів : монографія / Л. О. Примостка та ін. Київ : КНЕУ, 2012. 459 с.
38. Вітлінський В. В., Верченко П. І. Аналіз, моделювання та управління економічним : навч. метод. посіб.. Київ: КНЕУ, 2000. 292 с.
39. Вітлінський В. В., Великоіваненко Г. І. Ризикологія в економіці та підприємництві : монографія. Київ : КНЕУ, 2004. 480 с.
40. Про затвердження Положення про організацію системи управління ризиками в банках України та банківських групах : Постанова НБУ від 11.06.2018 р. № №64. URL: <https://zakon.rada.gov.ua/laws/show/v0064500-18#Text> (дата звернення: 18.05.2023).
41. Банківські ризики: теорія та практика управління : монографія. / Л.О. Примостка та ін.. Київ : КНЕУ, 2008. 456 с.
42. Бобиль В. В. Розробка елементів управління ринковим ризиком у банках. *Проблеми економіки транспорту*. 2013. №6. С. 83–93. URL: http://nbuv.gov.ua/UJRN/znpdnuzt_pet_2013_6_13 (дата звернення: 18.05.2023).
43. Вовк В. Я., Хмеленко О. В. Кредитування і контроль : навч. посіб.. Київ: Знання, 2008. 463 с.

44. Євтушенко Г. В., Бабошко А. І., Бушля Д. І.. Операційні ризики в системі банківської діяльності та нові шляхи їх попередження. *Глобальні та національні проблеми економіки*. 2015. №5. С. 396 – 400.
45. Набок Р. М. Окремі питання управління операційним ризиком у банках. *Вісник НБУ*. 2013. №1. С. 61 – 65.
46. Волкова Н. І., Мухіна А. С. Фінансові ризики банку: оцінка та механізм нейтралізації. *Modern Economics*. 2020. № 22. С. 6-12. URL: [https://doi.org/10.31521/modecon.V22\(2020\)-01](https://doi.org/10.31521/modecon.V22(2020)-01) (дата звернення: 18.05.2023).
47. Забчук Г., Іващук О.. Банківські ризики в Україні в умовах стратегічної невизначеності розвитку національної економіки. *Інститут бухгалтерського обліку, контроль та аналіз в умовах глобалізації, Міжнародний науковий журнал*. 2020. № 2. С. 102-117. DOI: <https://doi.org/10.35774/ibo2020.02.102> (дата звернення: 18.05.2023).
48. Банківські ризики легалізації доходів, отриманих злочинним шляхом: сутність, види та система управління / Мордань Є. Ю. та ін. *Економіка та держава*. 2021. № 1. С. 115–123. DOI: [10.32702/2306-6806.2021.1.115](https://doi.org/10.32702/2306-6806.2021.1.115) (дата звернення: 18.05.2023).
49. Кузніченко Я., Фролов С. Вплив банківських ризиків на фінансову безпеку банків. *Фінанси та банківська справа*. 2019. Т. 218, № 8. С. 105–112. DOI: [10.32752/1993-6788-2019-1-218-105-112](https://doi.org/10.32752/1993-6788-2019-1-218-105-112). (дата звернення: 18.05.2023).
50. Корупційні ризики в банківській діяльності / Шендрик В. В. та ін. *Financial and credit activity-problems of theory and practice*. 2019. Т. 4, № 31. С. 70-79.
51. Корнилюк Р. В. Системний ризик у банківській діяльності : дис. докт. ек. наук. Київ, 2020. 407 с.
52. Вітлінський В. В., Парнарівський О. В., Наконечний Я. С, Великоіваненко Г. І.. Кредитний ризик комерційного банку : навч. посіб. Київ : т-во "Знання", 2000. 251 с.
53. Онисько С. М., Шолудько О. В., Содома Р. І. Моніторинг кредитних ризиків банків-кредиторів. *Облік і фінанси*. 2015. №3. С. 108 – 115.

54. Мічинська О. Ю. Кредитні ризики банків: проблеми управління. *Ефективна економіка*. 2016. №3. URL: <http://www.economy.nayka.com.ua/?op=1&z=4857> (дата звернення: 18.05.2023).
55. Артем'єв О. О., Бестюк А. М. Вплив кредитних ризиків на розвиток банківської системи України в умовах трансформаційних процесів. *Науковий вісник Херсонського державного університету, серія "Економічні науки"*. 2020. №37. С. 62–67. URL: <https://doi.org/10.32999/ksu2307-8030/2020-37-11> (дата звернення: 18.05.2023).
56. Похилько С. В., Новіков В. М. Аналіз існуючих підходів щодо управління та мінімізації кредитного ризику банку. *Вісник СумДУ, серія "Економіка"*. 2019. №1 С. 53–63. DOI: 10.21272/1817-9215.2019.1-7 (дата звернення: 18.05.2023).
57. Ларіонова К. Л., Донченко Т. В. Аналіз та проблеми оцінки кредитного ризику банків України. *Вісник Хмельницького національного університету*. 2020. Т. 261, № 1. С. 233–240. URL: <https://www.doi.org/10.31891/2307-5740-2020-282-1-42> (дата звернення: 18.05.2023).
58. Стелмах В.С., Міщенко В.І. Ліквідність банку: окремі аспекти управління та світовий досвід регулювання та нагляду: науково-методичні матеріали. Київ : НБУ ЦНД, 2008. 286 с.
59. Міщенко В.І., Слав'янська Н. Г., Коренєва О. Г. Банківські операції: підручник. 2-е вид., переробл. і доп. Київ : Знання, 2007. 84 – 85 с.
60. Долінський Л. Б., Сащук Д. П. Оцінка надійності підприємств: теоретико-методологічний підхід. *Фінанси України*. 2008. №1. С. 108–117. URL: http://nbuv.gov.ua/UJRN/Fu_2008_1_11 (дата звернення: 18.05.2023).
61. Владичин У. В. Банківське кредитування : навч. посіб. Київ: Атіка, 2008. 648 с.
62. Basel II: Revised international capital framework. Bank for International Settlements. URL: <https://www.bis.org/publ/bcbsca.htm> (date of access: 18.05.2023).

63. The Basel Framework (Basel III edition). Bank for International Settlements. URL: https://www.bis.org/basel_framework/index.htm?export=pdf (date of access: 18.05.2023).
64. Про схвалення Методичних рекомендацій щодо організації та функціонування систем ризик-менеджменту в банках України : Постанова НБУ від 02.08.2004 р. № №361. URL: <https://zakon.rada.gov.ua/laws/show/v0361500-04#Text> (дата звернення: 18.05.2023).
65. Про затвердження Інструкції про порядок регулювання діяльності банків в Україні : Постанова НБУ від 28.08.2001 р. № №368. URL: <http://zakon2.rada.gov.ua/laws/show/z0841-01> (дата звернення: 18.05.2023).
66. Лавренюк В. В., Лук'янчук Д. Ю. Управління валютним ризиком банку в умовах турбулентності на світових фінансових ринках. *Бізнес Інформ*. 2020. №11. С. 288–295. URL: https://www.business-inform.net/article/?year=2020&abstract=2020_11_0_288_295&lang=en
67. Gray D. Interest Rate Risk Management at Community Banks. *Community Banking Connections*. URL: <https://www.communitybankingconnections.org/articles/2012/q3/interest-rate-risk-management> (date of access: 18.05.2023).
68. Resti A., Sironi A. Risk Management and Shareholders' Value in Banking : book. 1st edition. West Sussex : John Wiley & Sons Ltd, 2007. 782p.
69. Михайличенко М. С. Особливості управління активами та пасивами банку в умовах української економіки. *Наукові записки. Серія "Економічні науки"*. 2009. Т. 94. С. 65–74. URL: <https://ekmair.ukma.edu.ua/handle/123456789/4019> (дата звернення: 18.05.2023).
70. Бучко І. Є. Сучасні методи управління процентним ризиком. *Вісник Львівської комерційної академії. Серія економічна*. 2014. №. 45. С. 107-110. URL: http://nbuv.gov.ua/UJRN/Vlca_ekon_2014_45_24. (дата звернення: 18.05.2023).
71. Бобиль В. В. Розробка елементів управління ринковим ризиком у банках. *Проблеми економіки транспорту*. 2013. №6. С. 83–93. DOI: 10.15802/pte.v0i6.23037 (дата звернення: 18.05.2023).

72. Коваленко В. В. Управління ризиком ліквідності банку в умовах фінансової нестабільності. *Ефективна економіка*. 2020. № 7. URL: <http://www.economy.nayka.com.ua/?op=1&z=8057> (дата звернення: 18.05.2023).
73. Цай Р., Чжан М. Вплив кредитного ризику на ризик ліквідності: приклад українських банків. *Вісник Національного банку України*. 2017. С. 22–35. URL: <https://doi.org/10.26531/vnbu2017.241.021> (дата звернення: 18.05.2023).
74. Engle, Robert F. Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation. *Econometrica*. 1984. Vol. 4. no. 50 (4). P. 987–1008. URL: <https://doi.org/10.2307/1912773> (date of access: 18.05.2023).
75. Bollerslev T. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*. 1986. Vol. 31. no. 3. P. 307–327. URL: [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1) (date of access: 18.05.2023).
76. D.B. Nelson. Conditional heteroscedasticity in asset returns: a new approach. *Econometrica*. 1991. Vol. 59. no. 2. P. 347–370. URL: <https://doi.org/10.2307/2938260> (date of access: 18.05.2023).
77. Кузьмак О. М. Методи оцінки операційних ризиків банківських установ. *Економічні науки. Серія «Облік і фінанси»*. 2013. Т. 10, №1. URL: http://nbuv.gov.ua/UJRN/ecnof_2013_10%281%29__40 (дата звернення: 18.05.2023).
78. Bobyl V., Dron M., Taranenk, A. Operational risks: assessment and ways of mitigation. *Problemi ekonomiki transportu*. 2016. no. 16. URL: <https://doi.org/10.15802/pte.v0i16.151919> (date of access: 18.05.2023).
79. Євтушенко Г. В., Бабошко А. І., Бушля Д. І. Операційні ризики в системі банківської діяльності та нові шляхи їх попередження. *Глобальні та національні проблеми економіки*. 2015. № 5. С. 396–400.
80. Краснова І. В., Лавренюк В. В. Підходи до моделювання операційного ризику банку. *Бізнес Інформ*. 2022. №4. С. 110–119. URL: http://www.investplan.com.ua/pdf/4_2022/16.pdf
81. Фартушний І. Д., Лепеха К. А. Моделювання кредитного ризику комерційного банку. *Актуальні проблеми економіки та управління*. 2017. №11. URL: <http://ape.fmm.kpi.ua/article/view/102582> (дата звернення: 18.05.2023).

82. Про затвердження Положення про визначення банками України розміру кредит-ного ризику за активними банківськими операціями : Постанова НБУ від 30.06.2016 р. № №351. URL: <http://zakon2.rada.gov.ua/laws/show/v0351500-16> (дата звернення: 18.05.2023).

83. Ониськів Л. М. Актуальні проблеми оцінки кредитного ризику банку: нові вимоги та вплив на кредитування. *Глобальні та національні проблеми економіки. Серія «Гроші, фінанси, кредит»*. 2017. №15. С. 542 – 547. URL: <http://global-national.in.ua/archive/15-2017/108.pdf> (дата звернення: 18.05.2023).

84. Коцовська Р. Р. Оцінювання банками України кредитного ризику в сучасних економічних умовах. *Науковий вісник НЛТУ України*. 2016. Т. 26, № 6. С. 242 – 249. URL: http://nbuv.gov.ua/UJRN/nvnltu_2016_26.6_39 (дата звернення: 18.05.2023).

85. Про державне регулювання ринку цінних паперів в Україні : Закон України від 30.10.1996 р. № 448/96-ВР. URL: <http://zakon2.rada.gov.ua/laws/show/448/96-вр>. (дата звернення: 18.05.2023).

86. Притула Н. І., Малікова К. Л. Кредитно-рейтингова оцінка кредитоспроможності підприємств-позичальників. *Сайт видання Економічні науки*. URL: http://www.rusnauka.com/20_AND_2009/Economics/48664.doc.htm (дата звернення: 18.05.2023).

87. Матвійчук А. В., Чех І. О. Відбір значимих характеристик для побудови моделі апікаційного скорингу при споживчому кредитуванні. *Фінанси України*. 2016. №5. С. 91 – 111. URL: http://nbuv.gov.ua/UJRN/Fu_2016_5_8. (дата звернення: 18.05.2023).

88. Піщулін А. Кредитний скоринг від А до Я. Веб-портал "Bankchart". URL: http://www.bankchart.com.ua/individualniy_biznes/statti/kreditniy_skoring_vid_a_do_ya_chastina_1 (дата звернення: 18.05.2023).

89. Моделі оцінювання кредитоспроможності позичальників банку. *Студопедія*. URL: https://studopedia.com.ua/1_334145_rabota-gorizontallynoy-fermi.html (дата звернення: 18.05.2023).

90. Терещенко О. Дискримінантна модель інтегральної оцінки фінансового стану підприємства. *Економіка України*. 2003. № 8. С. 38 - 44.
91. Чайковський Я. І., Краснов С. О. Зарубіжні моделі аналізу кредитоспроможності позичальників-юридичних осіб. *Наука молода : зб. наук. праць молодих вчених THEU*. 2015. № 22. С. 151 – 162. URL: <http://dspace.wunu.edu.ua/handle/316497/3387> (дата звернення: 18.05.2023).
92. Смолева Т. М. Сучасні методи оцінки кредитоспроможності позичальників банками України. *Фінанси, учет, банки*. 2014. Т. 20, №1. С. 241 – 245. URL: http://nbuv.gov.ua/UJRN/Fub_2014_1_36 (дата звернення: 18.05.2023).
93. Mihalovič, M.. Performance Comparison of Multiple Discriminant Analysis and Logit Models in Bankruptcy Prediction. *Economics & Sociology*. 2016. Vol 4, no. 9. P. 101-118. URL: <https://doi.org/10.14254/2071-789x.2016/9-4/6> (date of access: 18.05.2023).
94. Affes, Z., Hentati-Kaffel, R. Predicting US Banks Bankruptcy: Logit Versus Canonical Discriminant Analysis. *Computational Economics*. 2017. Vol. 1, no 54. P. 199-244. URL: <https://doi.org/10.1007/s10614-017-9698-0> (date of access: 18.05.2023).
95. Performance Evaluation of Commercial Banks Based on Factor Analysis / Lv Q et. al. *E3S Web Of Conferences*. 2020. Vol. 214. P. 1 – 5. URL: <https://doi.org/10.1051/e3sconf/202021402017> (date of access: 18.05.2023).
96. Прімеров О. К., Кісельова Г. А. Застосування апарату нечіткої логіки для оцінки рівня банківського ризику. *Економіка і суспільство*. 2016. №7. С. 825 – 829. URL <https://ekmair.ukma.edu.ua/items/ab033166-57ee-4603-a2d9-538bb746cbde> (дата звернення: 18.05.2023).
97. Кожухівська О. А. Прогнозування ризиків кредитування фізичних осіб за математичними моделями. *Вісник Національного університету "Львівська політехніка". Інформаційні системи та мережі*. 2013. № 770. С. 177–185. URL: http://nbuv.gov.ua/UJRN/VNULPICM_2013_770_23 (дата звернення: 18.05.2023).
98. Писанець, К. К. Моделі оцінки кредитного ризику позичальників з часовим параметром для систем аплікаційного кредитного скорингу. *Бізнес-інформ*. 2013. № 7. С. 136–140. URL: <http://www.business->

inform.net/export_pdf/business-inform-2013-7_0-pages-136_140.pdf. (дата звернення: 18.05.2023).

99. Великоіваненко, Г. І., Трокоз, Л. О. Нейро-нечітка модель оцінювання прострочених позик комерційного банку. *Нейро-нечіткі технології моделювання в економіці*. 2014. № 3. С. 23–66. URL: <https://ir.kneu.edu.ua/handle/2010/5419> (дата звернення: 18.05.2023).

100. Кузнєцова, Н. В., Бідюк, П. І. Нейронні та мережі Байєса у задачі аналізу кредитних ризиків. *Реєстрація, зберігання і обробка даних*. 2015. Т. 17, № 2. С. 61–71. URL: <http://dspace.nbuv.gov.ua/handle/123456789/131568> (дата звернення: 18.05.2023).

101. Матвійчук, А. В. Штучний інтелект в економіці: нейронні мережі, нечітка логіка: монографія. Київ: КНЕУ, 2011. 439 с.

102. Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects / Dumitrescu E. et. al. *European Journal Of Operational Research*. 2022. Vol. 297, no. 3. P. 1178 – 1192. URL: <https://doi.org/10.1016/j.ejor.2021.06.053> (date of access: 18.05.2023).

103. Erdal, H., Karahanoğlu, İ. Bagging ensemble models for bank profitability: An empirical research on Turkish development and investment banks. *Applied Soft Computing*. 2016. Vol. 49. P. 861-867. URL: <https://doi.org/10.1016/j.asoc.2016.09.010> (date of access: 18.05.2023).

104. Le, H., Viviani, J. Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios. *Research In International Business And Finance*, 2018. Vol. 44. P. 16-25. URL: <https://doi.org/10.1016/j.ribaf.2017.07.104> (date of access: 18.05.2023).

105. Kohonen T. Self-Organizing Maps : book. 3rd Extended Edition. New York, 2001. p. 501.

106. An Overview of Principal Component Analysis / Karamizadeh S. et. al. *Journal Of Signal And Information Processing*. 2013. Vol. 4, no 3. P. 173-175. URL: https://www.scirp.org/pdf/JSIP_2013101711003963.pdf (date of access: 18.05.2023).

107. Schölkopf, B., Smola A., Müller K. R. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*. 1998. Vol. 10, no 5. P. 1299–1319. URL: <https://www.mlpack.org/papers/kpca.pdf> (date of access: 18.05.2023).
108. Genest, C., Nešlehová, J. G., Ramsay J. O. A Conversation with James O. Ramsay. *International Statistical Review / Revue Internationale de Statistique*. 2014. Vol 82, no 2. P. 161–183. URL: <https://ideas.repec.org/a/bla/istatr/v82y2014i2p161-183.html> (date of access: 18.05.2023).
109. Tenenbaum J. B., V. de Silva, Langford J. C.. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*. 2000. Vol. 290. P. 2319–2323. URL: <https://www.science.org/doi/10.1126/science.290.5500.2319> (date of access: 18.05.2023). DOI: 10.1126/science.290.5500.2319
110. Belkin M., Niyogi P. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. *Advances in Neural Information Processing Systems, MIT*. 2001. Vol. 14. P. 586–691. URL: <https://web.mit.edu/fustflum/documents/papers/niyogi.pdf> (date of access: 18.05.2023).
111. Belkin M., Niyogi P. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*. 2003. Vol 15, no. 6. P. 1373-1396. URL: <https://www2.imm.dtu.dk/projects/manifold/Papers/Laplacian.pdf> (date of access: 18.05.2023).
112. Ткаченко Н., Момот О. Науково-методичний підхід до кількісного оцінювання впливу функціонування банків з участю транснаціонального банківського капіталу на банківську систему України. *Вісник Тернопільського національного економічного університету. «Фінанси і кредит»*. 2018. № 3. С. 46-59. URL: <http://dspace.tneu.edu.ua/handle/316497/32914> (дата звернення: 18.05.2023).
113. Заруцька, О. Обґрунтування підходу до масштабного розподілу банків України на основі структурно-функціональних груп. *Вісник Національного банку України*. 2012. № 10. С. 20-24. URL: http://nbuv.gov.ua/UJRN/Vnbu_2012_10_13 (дата звернення: 18.05.2023).

114. Заруцька, О. П. Структурно-функціональний аналіз фінансової стійкості банківської системи з використанням карт Кохонена. *Нейро-нечіткі технології моделювання в економіці*. 2014. С. 67–93. URL: <https://ir.kneu.edu.ua:443/handle/2010/5434> (дата звернення: 18.05.2023).
115. Заруцька, О. П. Метод структурно-функціонального аналізу банківської системи в умовах скорочення ринку. *Нейро-нечіткі технології моделювання в економіці*. 2015. № 4. С. 18-43. URL: http://nbuv.gov.ua/UJRN/Nntm_2015_4_4 (дата звернення: 18.05.2023).
116. Orsenigo C., Vercellis C. Linear versus nonlinear dimensionality reduction for banks credit rating prediction. *Knowledge-Based Systems*. 2013. Vol. 47. P. 14-22, URL: <https://doi.org/10.1016/j.knosys.2013.03.001> (date of access: 18.05.2023).
117. Classification of Imbalanced Banking Dataset using Dimensionality Reduction / Valarmathi B. et. al. *International Conference on Intelligent Computing and Control Systems (ICCS)*. 2019. P. 1353-1357. URL: <https://ieeexplore.ieee.org/abstract/document/9065648> (date of access: 18.05.2023).
118. Steinhaus, H. Sur la division des corps matériels en parties. *Bulletin L'Académie Polonaise des Science*. 1957. Vol. 4. P. 801-804.
119. Lloyd S. Least square quantization in PCM's. *IEEE Transactions on Information Theory*. 1982. Vol. 28, no. 2. P. 129 – 137.
120. MacQueen J. Some methods for classification and analysis of multivariate observations. *In Proc. 5th Berkeley Symp. on Math. Statistics and Probability*. 1967. No. 1. P. 281—297.
121. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN / Schubert E. et. al. *ACM Transactions on Database Systems*. 2017. Vol. 42, no. 3. P. 1 – 21. URL: <https://doi.org/10.1145/3068335> (date of access: 18.05.2023).
122. OPTICS: Ordering Points To Identify the Clustering Structure / Ankerst M. et. al. *ACM SIGMOD International Conference on Management of Data*. 1999. Vol. 48, no. 2. P. 49–60. URL: <https://doi.org/10.1145/304181.304187> (date of access: 18.05.2023).

123. Campello R. J. G. B., Moulavi D., Sander J. Density-based clustering based on hierarchical density estimates. *PAKDD 2013: Advances in Knowledge Discovery and Data Mining*. 2013. Vol. 7819. P. 160 – 172. URL: https://doi.org/10.1007/978-3-642-37456-2_14 (date of access: 18.05.2023).
124. Cheng Y. Mean Shift, Mode Seeking, and Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1995. Vol. 17, no. 8. P. 790 – 799. URL: <https://ieeexplore.ieee.org/document/400568> (date of access: 18.05.2023).
125. Von Luxburg U. A tutorial on spectral clustering. *Stat Comput*. 2007. Vol. 17. P. 395 – 416. URL: <https://doi.org/10.1007/s11222-007-9033-z> (date of access: 18.05.2023).
126. Zhang T., Ramakrishnan R., Livny M. BIRCH: an efficient data clustering method for very large databases. *ACM SIGMOD Record*. 1996. Vol. 25, no. 2. P. 103 – 114. URL: <https://doi.org/10.1145/235968.233324> (date of access: 18.05.2023).
127. Biarnes A. Gaussian Mixture Models and Expectation-Maximization (A full explanation). *Towards Data Science*. URL: <https://towardsdatascience.com/gaussian-mixture-models-and-expectation-maximization-a-full-explanation-50fa94111ddd> (date of access: 18.05.2023).
128. Kaminskyi and M. Nehrey, "Information Technology Model for Customer Relationship Management of Nonbank Lenders: Coupling Profitability and Risk" 2021 11th International Conference on Advanced Computer Information Technologies (ACIT), Deggendorf, Germany, 2021, pp. 234-237. URL: <https://ieeexplore.ieee.org/document/9548581>
129. Kaminskyi A., Nehrey M. Clustering Stocks by ESG Score Values, Risks and Returns: Case of Expanded German Index DAX. *Advances in Artificial Systems for Medicine and Education VI*. 2023. Vol. 159. P. 264–276.
130. Соцька, Ю. І. Методологічні засади кластерного аналізу конкурентоспроможності банків України. *Фінансово-кредитна діяльність: проблеми теорії та практики*. 2015. № 2. С. 177–185. URL: http://nbuv.gov.ua/UJRN/Fkd_2015_2_22 (дата звернення: 18.05.2023).

131. Рашкован В., Поکیدін Д. Кластерний аналіз бізнес-моделей українських банків: застосування нейронних мереж Кохонена. *Вісник Національного банку України*. 2016. № 12. С. 13-40. URL: https://journal.bank.gov.ua/uploads/articles/238-2_ukr.pdf (дата звернення: 18.05.2023).
132. Примостка Л. О., Краснова І. В., Бричка С. М. Кластеризація українських банків в координатах "бізнес-модель - фінансова стійкість". *Science of Europe: Economics Sciences*. 2017. Т. 37, №1. С. 44–54. URL: [https://doi.org/10.18371/2221-755x1\(37\)2020208212](https://doi.org/10.18371/2221-755x1(37)2020208212) (дата звернення: 18.05.2023).
133. Мікуляк К. А. Застосування кластеризації для визначення конкурентних переваг банківських установ. *Modern Economics*. 2017. №6. С. 153–161. URL: <https://modecon.mnau.edu.ua/zastosuvannia-klasteryzatsii-dlia-vyznachennia-konkurentnykh-perevah/> (дата звернення: 18.05.2023).
134. Гур'янова Л. С., Панасенко О. В., Вознюк С. М. Системний аналіз і моделювання ринку освітніх послуг. Сучасний стан наукових досліджень і технологій в промисловості. 2021. № 1 (15). С. 32–42. URL: <http://journals.uran.ua/itssi/article/view/227961>
135. Klebanova, T., Rudachenko, O., Gvozdytskyi, V., Mozgovyi, I., & Guryanova, I. (2020). Classification of Regions of Ukraine by The Level of Social Tension. *WSEAS TRANSACTIONS ON SYSTEMS AND CONTROL*, 15, 576–581. URL: <https://wseas.org/wseas/cms.action?id=23195>
136. Humblot, T. Classification of European Banks According to their Business Model: an Objective Approach. *Eco Conjoncture*. 2020. No. 5. P. 1-10. URL: <https://economic-research.bnpparibas.com/pdf/en-US/Classification-European-banks-according-business-model-objective-approach-7/27/2020,39258> (дата звернення: 18.05.2023).
137. Negnevitsky M. Identification of failing banks using Clustering with self-organising neural networks. *Procedia Computer Science*. 2017. Vol. 108. P. 1327-1333. URL: <https://doi.org/10.1016/j.procs.2017.05.125> (date of access: 18.05.2023).
138. Kolodiziev O., Chmutova I., Biliaieva V. Selecting a kind of financial innovation according to the level of a bank's financial soundness and its life cycle stage.

Banks and Bank Systems. 2016. Vol. 11, no. 4, P. 40-49. DOI: 10.21511/bbs.11(4).2016.04 (date of access: 18.05.2023).

139. Dao B., Khanh P. Cluster Analysis of Vietnamese Banks. *SSRN Electronic Journal*. 2014. URL: <https://doi.org/10.2139/ssrn.2543094> (date of access: 18.05.2023).

140. Cluster approach to banking supervision with reference to bank risk profile / Kovalenko V. et. al. *Economic Annals-XXI*. 2019. Vol. 177, no. 5-6. P. 82-91. DOI: 10.21003/ea.V177-07. (date of access: 18.05.2023).

141. Корчинський В. В., Долінський Л. Б. Ідентифікація та кількісне оцінювання кредитного ризику комерційного банку. *Економічний аналіз: зб. наук. праць*. 2016. Т. 25, № 1. С. 180–189.

142. Великоіваненко Г. І., Корчинський В. В., Чернишова В. В. Дослідження ефекту перенавчання нейронних мереж на прикладі задачі аплікаційного скорингу. *Нейро-нечіткі технології моделювання в економіці*. 2016. № 5. С. 3–23.

143. Корчинський В. В., Черишова В. В. Побудова скорингової моделі на основі нейронної мережі типу Персептрон. *Український журнал прикладної економіки*. 2016. Т. 1, № 4. С. 46–55.

144. Лаговський В. В., Долінський Л. Б., Корчинський В. В. Особливості кредитно-інвестиційної діяльності на фінансовому ринку України. *Економічний вісник: науковий фаховий журнал*. 2020. Т. 25, № 4. С. 74–88.

145. Корчинський В. В. Побудова системи ризик-менеджменту в комерційному банку. *Інноваційний ресурс соціально-економічного прогресу України: зб. доп. 83-ї наук. конф. студ. КНЕУ, м. Київ, 25–27 квіт. 2016 р. Київ, 2016. С. 290–293.*

146. Корчинський В. В., Чергишова В. В. Скорингова модель оцінки кредитоспроможності клієнтів банків на основі інструментарію нечіткої логіки. *Інноваційний ресурс соціально-економічного прогресу України: зб. доп. 83-ї наук. конф. студ. КНЕУ, м. Київ, 25–27 квіт. 2016 р. Київ, 2016. С. 602–604.*

147. Матвійчук А. В., Корчинський В. В., Чернишова В. В. Алгоритм побудови скорингової моделі на основі нейронної мережі персептронного типу.

КМОСС: Матеріали ІІ Всеукр. науково-техн. конф., м. Дніпро, 1–3 листоп. 2016 р. Дніпро, 2016. С. 202–203.

148. Долінський Л. Б., Корчинський В. В. Аналіз особливостей кредитно-інвестиційної діяльності банків в Україні. *Економіка підприємства: сучасні проблеми теорії та практики*: Матеріали дев'ятої міжнар. наук.-практ. конф, м. Одеса, 11–12 верес. 2020 р. Одеса, 2020. С. 105–106.

149. Долінський Л. Б. Оцінювання та управління кредитним ризиком боргових зобов'язань : монографія. Київ : КНЕУ, 2017. 551 с.

150. Методологія рейтингової оцінки комерційного банку рейтингового агентства "ІВІ - Рейтинг". "ІВІ - Рейтинг". 2015. URL: <https://ibi.com.ua/UK/pravila-ta-reguljuvannja-upravlinnya-rejtinguvannyam-metodologiya/metodologiya.html> (дата звернення: 18.05.2023).

151. Корнілюк Р. Рейтинг устійчивости банков. Методика. *Фінансовий портал "Мінфін"*. 2015. URL: <http://minfin.com.ua/banks/rating/method/> (дата звернення: 18.05.2023).

152. Корнілюк Р. Рейтинг жизнеспособности украинских банков. Форбс. - 2015 URL: https://cxid.info/124110_zhiznesposobnye-banki-ukrainy-reiting.html (дата звернення: 18.05.2023).

153. Долінський Л. Б., Першин Д. О.. Оцінювання фінансової стійкості банківських установ в аспекті надійності депозитних вкладів. *БізнесІнформ*. 2015. №5. С. 233–239. URL: https://www.business-inform.net/annotated-catalogue/?year=2015&abstract=2015_05_0 (дата звернення: 18.05.2023).

154. Примостка Л. О., Краснова І. В., Бричка С. М. Кластеризація українських банків в координатах "бізнес-модель - фінансова стійкість. *Science of Europe: Economis Sciences*. 2017. № 21. С. 44–54. URL: <https://ir.kneu.edu.ua/handle/2010/36434> (дата звернення: 18.05.2023).

155. *An Introduction to Statistical Learning* / Gareth J. et. al. 1st ed. Berlin : Springer, 2013. 440 p.

156. Richard E. B. The Theory of Dynamic Programming. *RAND Corporation*. 1954. URL: <https://www.rand.org/pubs/papers/P550.html> (date of access: 18.05.2023).

157. Richard E. B. Adaptive control processes: a guided tour. Princeton : Princeton University Press, 1961. 274 p.
158. Application of high-dimensional feature selection: evaluation for genomic prediction in man / Mairead L. et. al. *Sci Rep.* 2015. Vol. 5. URL: [doi:10.1038/srep10312](https://doi.org/10.1038/srep10312) (date of access: 18.05.2023).
159. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance / Dormann, C. et. al. *Ecography*. 2012. Vol. 36, no. 1. P. 27-46. URL: <https://doi.org/10.1111/j.1600-0587.2012.07348.x> (date of access: 18.05.2023).
160. Ferreira A., Figueiredo M. Efficient feature selection filters for high-dimensional data. *Pattern Recognition Letters*. 2012. Vol. 33, no. 13. P. 1794-1804. URL: <https://doi.org/10.1016/j.patrec.2012.05.019> (date of access: 18.05.2023).
161. Zhao Z., Liu H. Spectral feature selection for supervised and unsupervised learning. *Proceedings Of The 24Th International Conference On Machine Learning*. 2007. P. 1151 – 1157. URL: <https://doi.org/10.1145/1273496.1273641> (date of access: 18.05.2023).
162. Kriegel H. P., Kröger P., Zimek, A. Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Transactions on Knowledge Discovery from Data*. 2009. Vol. 3, no. 1. P. 1 – 58. URL: <https://doi.org/10.1145/1497577.1497578> (date of access: 18.05.2023).
163. Oskolkov N. How exactly UMAP works. And why exactly it is better than t-SNE. *Towards Data Science*. 2019. URL: <https://towardsdatascience.com/how-exactly-umap-works-13e3040e1668> (date of access: 18.05.2023).
164. Wattenberg M., Viégas, F., Johnson I. How to use t-SNE effectively. *Distill*. 2016. URL: <http://distill.pub/2016/misread-tsne> (date of access: 18.05.2023).
165. van der Maaten L., Hinton G. Visualizing Data Using t-SNE. *Journal of Machine Learning Research*. 2008. Vol. 9. URL: https://www.researchgate.net/publication/228339739_Visualizing_data_using_t-SNE (date of access: 18.05.2023).

166. van der Maaten L. Barnes-hut-sne. *arXiv*. 2013. URL: <https://doi.org/10.48550/arXiv.1301.3342> (date of access: 18.05.2023).
167. McInnes L. How UMAP works. UMAP *Python implementation official documentation*. URL: https://umap-learn.readthedocs.io/en/latest/how_umap_works.html (date of access: 18.05.2023).
168. McInnes L., Healy, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv*. 2018. URL: <https://doi.org/10.48550/arXiv.1802.03426> (date of access: 18.05.2023).
169. Shubert E., Gertz M. Intrinsic t-Stochastic Neighbor Embedding for Visualization and Outliers Detection. A Remedy Against the Curse of Dimensionality?. *Heidelberg University*. URL: <https://pdfs.semanticscholar.org/d767/2e61c99ee7d6c3711ca2b78469e9e4b17e43.pdf> (date of access: 18.05.2023).
170. UMAP-assisted K-means clustering of large-scale SARS-CoV-2 mutation datasets / Hozumi Y. et. al. *Computers In Biology And Medicine*. 2021. Vol. 131. URL: doi: 10.1016/j.compbiomed.2021.104264 (date of access: 18.05.2023).
171. Application of Uniform Manifold Approximation and Projection (UMAP) in spectral imaging of artworks / Vermeulen M. et. al. *Spectrochimica Acta Part A: Molecular And Biomolecular Spectroscopy*. 2021. Vol. 252. URL: <https://doi.org/10.1016/j.saa.2021.119547> (date of access: 18.05.2023).
172. Dimensionality reduction by UMAP reinforces sample heterogeneity analysis in bulk transcriptomic data / Yang Y. et al. *Cell Reports*. 2021. Vol. 36, no. 4. URL: <https://doi.org/10.1016/j.celrep.2021.109442> (date of access: 18.05.2023).
173. Kobak D., Berens P. The art of using t-SNE for single-cell transcriptomics. *Nature Communications*. 2019. Vol. 10, no. 1. URL: <https://doi.org/10.1038/s41467-019-13056-x> (date of access: 18.05.2023).
174. Sibson R. SLINK: an optimally efficient algorithm for the single link cluster method. *The Computer Journal*. 1973. Vol. 16. P. 30 – 34. URL: <https://doi.org/10.1093/comjnl/16.1.30> (date of access: 18.05.2023).

175. Rohlf F.J. Single link clustering algorithms. *Handbook of Statistics*. 1982. Vol. 2. P. 267 – 284. URL: [https://doi.org/10.1016/S0169-7161\(82\)02015-X](https://doi.org/10.1016/S0169-7161(82)02015-X) (date of access: 18.05.2023).
176. Defays D. An efficient algorithm for a complete link method. *Computer Journal*. 1977. No. 20. P. 364 – 366. URL: <https://doi.org/10.1093/comjnl/20.4.364> (date of access: 18.05.2023).
177. Ward J. Hierarchical Grouping to Optimize an Objective Function. *Journal Of The American Statistical Association*. 1963. Vol. 58, no. 301. P. 236-244. URL: <https://doi.org/10.1080/01621459.1963.10500845> (date of access: 18.05.2023).
178. Comparing different hierarchical linkage methods on toy datasets. scikit-learn official Python documentation. URL: https://scikit-learn.org/stable/auto_examples/cluster/plot_linkage_comparison.html (date of access: 18.05.2023).
179. Rousseeuw P. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal Of Computational And Applied Mathematics*. 1987. No. 20. P. 53-65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7) (date of access: 18.05.2023).
180. Calinski T., Harabasz J. A dendrite method for cluster analysis. *Communications In Statistics - Theory And Methods*. 1974. Vol 1, no. 3. P. 1-27. URL: <https://doi.org/10.1080/03610927408827101> (date of access: 18.05.2023).
181. Davies D., Bouldin D. A Cluster Separation Measure. *IEEE Transactions On Pattern Analysis And Machine Intelligence*. 1979. №. 2. P. 224-227. URL: <https://doi.org/10.1109/tpami.1979.4766909> (date of access: 18.05.2023).
182. Корчинський В. В., Долінський Л. Б., Мірошніченко І. В. Кластерний аналіз надійності банківських установ в Україні. *Моделювання та інформаційні системи в економіці: зб. наук. праць*. 2017. № 93. С. 49–73.
183. Korchynskyi V., Velykoivanenko H. Application of Clustering in the Dimensionality Reduction Algorithms for Separation of Financial Status of Commercial Banks in Ukraine. *Universal Journal of Accounting and Finance*. 2022. Vol. 10, no. 1. P. 148–60. DOI: 10.13189/ujaf.2022.100116.

184. Долінський Л. Б., Корчинський В. В. Відбір кількісних показників для проведення кластеризації вітчизняних банків в аспекті фінансової стійкості. *Економіко-математичне моделювання*: зб. мат. Першої нац. наук.-метод. конф., м. Київ, 30 верес. – 1 жовт. 2016 р. Київ, 2016. С. 122–123.

185. 11. Долінський Л. Б., Корчинський В. В. Кластерний аналіз надійності вітчизняних банківських установ. *Інноваційний ресурс соціально-економічного розвитку України*: зб. доп. 84-ї наук. конф. студ. КНЕУ, м. Київ, 24 квіт. – 23 трав. 2017 р. Київ, 2017. С. 389–391.

186. Долінський Л. Б., Корчинський В. В. Застосування сучасних інформаційних технологій для кластеризації українських комерційних банків в аспекті фінансової стійкості. *Математичні методи, моделі та інформаційні технології в економіці*: Матеріали VI Міжнар. науково-метод. конф., м. Чернівці, 18–19 квіт. 2019 р. Чернівці, 2019. С. 88–92.

187. 14. Великоіваненко Г. І., Корчинський В. В. Моделювання фінансового стану комерційних банків України на основі інструментів та методів машинного навчання. *Актуальні проблеми економіки, фінансів, обліку та права в умовах сучасних викликів*: зб. тез доп. міжнар. науково-практ. конф., м. Полтава, 25 серп. 2021 р. Полтава, 2021. С. 53–54.

188. Великоіваненко Г. І., Корчинський В. В. Аналіз стабільності показників фінансового стану комерційних банків України на основі методів машинного навчання. *КМОСС*: Матеріали VII Міжнар. науково-техн. конф., м. Дніпро, 3–5 листоп. 2021 р. Дніпро, 2021. С. 93–94.

189. Urdan T. *Statistics in Plain English*. 4th ed. NY: Taylor and Francis Group, 2011. 287 p.

190. Dixon W. Simplified Estimation from Censored Normal Samples. *The Annals Of Mathematical Statistics*. 1960. Vol. 31, no. 2. P. 385-391. URL: <https://doi.org/10.1214/aoms/1177705900> (date of access: 18.05.2023).

191. Hopkins B., Skellam J. A New Method for determining the Type of Distribution of Plant Individuals. *Annals Of Botany*. 1954. Vol. 18, no. 2. P. 213-227. URL: <https://doi.org/10.1093/oxfordjournals.aob.a083391> (date of access: 18.05.2023).

192. Hoberg G., Phillips G. Real and Financial Industry Booms and Busts. *The Journal Of Finance*. 2010. Vol. 65, no. 1, P. 45-86. URL: <https://doi.org/10.1111/j.1540-6261.2009.01523.x> (date of access: 18.05.2023).
193. Методика рейтингового оцінювання для комерційних банків. НРА Рюрик. URL: http://www.rurik.com.ua/documents/methodology/metod_banks_Cut_final.pdf (дата звернення: 18.05.2023).
194. Xu D., Tian Y. A. Comprehensive Survey of Clustering Algorithms. *Annals of Data Science*. 2015. Vol. 2. P. 165–193. URL: <https://doi.org/10.1007/s40745-015-0040-1> (date of access: 18.05.2023).

ДОДАТКИ

ДОДАТОК А

Повний перелік коефіцієнтів, розрахованих на основі відкритої статистики НБУ до проведення відбору ознак з показниками метрик, отриманими в процесі відбору ознак

Таблиця А.1

Група коефіцієнтів 1: Аналіз активів та показник системності

ID	Назва показника	Порядок розрахунку	VIF	MAD	SPEC
1	2	3	4	5	6
K0	Коефіцієнт системності	<i>Активи банку / Активи всього в системі</i>	1,83	0,02	0,75
K1.1	Коефіцієнт кредитного ризику	<i>Кредитна заборгованість – розрахунковий резерв на можливі втрати по кредитах / Кредити всього</i>	33,52	0,23	0,57
K1.2	Питома вага дохідних активів	<i>Працюючі активи / Активи всього</i>	545,09	0,13	0,64
K1.3	Коефіцієнт резервів під покриття втрат	<i>Резерви під кредити / Кредитні вкладення всього</i>	7,87	0,41	0,29
K1.4	Загальна дохідність активів	<i>Доходи всього / Активи всього (чисті активи)</i>	18,28	0,04	0,77
K1.5	Рентабельність активів	<i>Чистий прибуток / Активи всього</i>	2,18	0,01	0,61
K1.6	Операційна рентабельність	<i>Прибуток до оподаткування + Відрахування до резервів під знецінення кредитів / Активи (чисті активи)</i>	5,87	0,02	0,60
K1.7	Коефіцієнт дохідності активів	<i>Процентний дохід / Дохідні активи</i>	47,85	0,05	0,78

Закінчення таблиці А.1

1	2	3	4	5	6
K1.8	Співвідношення резервів під активні операції до активів	<i>Резерви під активні операції</i> / <i>Активи (чисті активи)</i>	7,27	0,12	0,49
K1.9	Частка кредитів в активах	<i>Кредити всього</i> / <i>Активи всього (чисті активи)</i>	16684,71	0,18	0,49
K1.10	Частка споживчих кредитів	<i>Кредити клієнтів</i> / <i>Кредити всього</i>	113,87	0,13	0,56
K1.11	Частка міжбанківських кредитів в активах	<i>МБК</i> / <i>Активи всього (чисті активи)</i>	698,30	0,06	0,66
K1.12	Частка цінних паперів в активах	<i>Цінні папери</i> / <i>Активи всього (чисті активи)</i>	3532,43	0,14	0,59
K1.13	Відношення кредитного портфелю до активів	<i>КІП</i> / <i>Активи всього (чисті активи)</i>	33087,42	0,14	0,64

Таблиця А.2

Група коефіцієнтів 2: Аналіз пасивів

1	2	3	4	5	6
K2.1	Ризик незбалансованої стійкості	<i>Строкові депозити</i> / <i>Пасиви всього</i>	43,60	0,15	0,37
K2.2	Коефіцієнт структури зобов'язань	<i>Депозити на вимогу / Строкові депозити</i>	28,13	0,90	0,18

Закінчення таблиці А.2

1	2	3	4	5	6
K2.3	Коефіцієнт строкової структури депозитів	<i>Строкові та ощадні депозити</i> / <i>Депозити всього</i>	89,91	0,19	0,29
K2.4	Коефіцієнт використання міжбанківських кредитів	<i>МБК (в т. ч. кошти НБУ)</i> / <i>Зобов'язання</i>	33,79	0,13	0,62
K2.5	Частка вкладів у зобов'язаннях	<i>Вклади (депозити клієнтів)</i> / <i>Зобов'язання всього</i>	129,27	0,16	0,63
K2.6	Коефіцієнт залежності від вкладів фізичних осіб	<i>Вклади фізичних осіб</i> / <i>Зобов'язання всього</i>	9,64	0,17	0,66

Таблиця А.3

Група коефіцієнтів 3: Аналіз капіталу

K3.1	Коефіцієнт автономності	<i>Власний капітал / Чисті активи всього</i>	42,47	0,15	0,53
K3.2	Достатність капіталу за депозитами	<i>Власний капітал / Депозити всього</i>	27,80	0,61	0,22
K3.3	Коефіцієнт покриття позик	<i>Власний капітал</i> / <i>Позики всього</i>	12,76	0,52	0,35
K3.4	Рентабельність капіталу	<i>Чистий прибуток / Власний капітал</i>	1,92	0,11	0,48

Таблиця А.4

Група коефіцієнтів 4: Аналіз збалансованості активів і пасивів

К4.1	Коефіцієнт використання строкових депозитів	<i>Кредити всього</i> / <i>Депозити всього</i>	37,61	0,53	0,35
К4.2	Коефіцієнт клієнтської бази	<i>Кошти клієнтів всього</i> / <i>Активи всього</i>	37,92	0,18	0,51
К4.3	Відношення кредитних вкладів до зобов'язань	<i>Кредити всього</i> / <i>Зобов'язання всього</i>	28,00	0,33	0,38
К4.4	Відношення кредитного портфелю до зобов'язань	<i>Кредитно-інвестиційний портфель</i> / <i>Зобов'язання всього</i>	77,71	0,45	0,37
К4.5	Відношення дохідних ресурсів до платних ресурсів	<i>Працюючі активи</i> / <i>Летючі (онкольні) зобов'язання + строкові зобов'язання</i>	6,25	0,83	0,35

Таблиця А.5

Група коефіцієнтів 5: Аналіз ліквідності

К5.1	Коефіцієнт миттєвої ліквідності	<i>Високоліквідні активи / Зобов'язання до запитання</i>	11,51	0,45	0,56
К5.2	Частка високоліквідних активів	<i>Високоліквідні активи</i> / <i>Активи всього (чисті активи)</i>	14,28	0,13	0,55

Таблиця А.6

Група коефіцієнтів 6: Аналіз фінансової ефективності

1	2	3	4	5	6
K6.1	Коефіцієнт ефективності діяльності	<i>Чистий операційний дохід від основної діяльності</i> / <i>Адміністративні витрати + Інші операційні витрати</i>	6,05	0,51	0,44
K6.2	Коефіцієнт відношення комісійного та процентного доходів	<i>Комісійні доходи / Процентні доходи</i>	8,74	0,28	0,47
K6.3	Коефіцієнт ефективності затрат	<i>Доходи всього / Витрати всього</i>	27,86	0,42	0,44
K6.4	Відношення процентних доходів та витрат	<i>Процентні доходи / Процентні витрати</i>	26,21	0,72	0,29
K6.5	Коефіцієнт безризикового покриття доходів	<i>Комісійні доходи / Витрати всього</i>	8,99	0,25	0,52
K6.6	Коефіцієнт безризикового покриття адміністративних витрат	<i>Комісійні доходи / Адміністративні витрати</i>	5,74	0,20	0,68
K6.7	Спред	<i>(Процентні доходи / Працюючі активи) - (Процентні витрати / Зобов'язання)</i>	86,88	0,03	0,64
K6.8	Рентабельність витрат	<i>Прибуток (до оподаткування) / Витрати всього</i>	6,94	0,33	0,32
K6.9	Чиста процентна маржа	<i>Процентні доходи - процентні витрати</i> / <i>Процентні (дохідні) активи</i>	64,17	0,04	0,59
K6.10	Відношення адміністративних витрат до активів	<i>Адміністративні витрати</i> / <i>Активи всього</i>	9,32	0,03	0,75
K6.11	Відношення чистого комісійного доходу до активів	<i>Чистий комісійний дохід</i> / <i>Активи всього</i>	5,24	0,01	0,72

Закінчення таблиці А.6

1	2	3	4	5	6
K6.12	Відношення торгового результату до активів	<i>Торговий результат</i> / <i>Активи всього</i>	2,01	0,01	0,78
K6.13	Відношення загальних резервів до активів	<i>Загальні відрахування до резервів</i> / <i>Активи всього</i>	4,38	0,04	0,34

Таблиця А.7

Група коефіцієнтів 7: Аналіз податкового навантаження

K7.1	Коефіцієнт податкового навантаження на доходи	<i>Податки всього / Доходи всього</i>	1,31	0,04	0,46
K7.2	Коефіцієнт податкового навантаження на капітал	<i>Податки всього / Власний капітал</i>	1,73	0,01	0,66
K7.3	Коефіцієнт податкового навантаження на прибуток до оподаткування	<i>Податки всього / Прибуток до оподаткування</i>	1,65	0,12	0,84

Таблиця А.8

Група коефіцієнтів 8: Аналіз ефективності залучення персоналу

1	2	3	4	5	6
K8.1	Відношення адміністративно-господарських витрат до загальних витрат	<i>АГВ</i> / <i>Витрати всього</i>	13,26	0,30	0,58
K8.2	Коефіцієнт використання доходів на утримання апарату управління	<i>Фонд оплати праці + Інші витрати на утримання апарату управління</i> / <i>Доходи всього</i>	N/A	N/A	N/A

Закінчення таблиці А.8

1	2	3	4	5	6
K8.3	Відношення прибутку до формування резервів до витрат на утримання персоналу	<i>Прибуток до формування резервів та оподаткування</i> / <i>Витрати на утримання персоналу (заробітна плата + нарахування в фонд ЗП + інші витрати на персонал)</i>	N/A	N/A	N/A
K8.4	Відношення витрат на утримання персоналу до чистих активів	<i>Витрати на утримання персоналу (заробітна плата + нарахування в фонд ЗП + інші витрати на персонал)</i> / <i>Чисті активи</i>	N/A	N/A	N/A

ДОДАТОК Б

Показники кореляції Пірсона між парами ознак, між якими кореляція була визначена як суттєва (> 0.5)

Таблиця Б.1

Показники кореляції Пірсона між парами ознак

К1	К2	Кореляція	К1	К2	Кореляція	К1	К2	Кореляція	К1	К2	Кореляція
K1.1	K1.3	1,000000	K2.1	K2.3	0,809465	K1.3	K1.5	0,642507	K2.3	K2.6	0,521650
K4.4	K5.1	0,996543	K6.9	K6.13	0,801669	K1.1	K1.5	0,642507	K1.8	K6.7	0,519911
K6.3	K6.8	0,993584	K1.1	K1.8	0,795849	K1.3	K6.13	0,637021	K1.7	K3.4	0,511342
K3.2	K5.1	0,991626	K1.3	K1.8	0,795849	K1.1	K6.13	0,637021	K1.4	K6.12	0,505425
K1.5	K6.13	0,990728	K1.5	K6.9	0,783874	K1.9	K1.12	0,636967	K3.4	K6.7	0,501234
K4.1	K4.4	0,990029	K1.10	K1.11	0,775870	K1.3	K6.9	0,617631			
K3.2	K4.4	0,988771	K1.4	K6.10	0,775549	K1.1	K6.9	0,617631			
K4.1	K5.1	0,985147	K3.4	K6.13	0,766912	K2.1	K2.6	0,616280			
K1.2	K1.13	0,976862	K1.5	K3.4	0,764178	K1.8	K6.9	0,608909			
K3.2	K4.1	0,972347	K2.4	K2.5	0,743432	K6.1	K6.4	0,602632			
K1.7	K6.9	0,929152	K1.7	K6.13	0,738897	K1.9	K1.10	0,599594			
K6.3	K8.1	0,910233	K1.5	K1.7	0,727536	K1.4	K1.6	0,587416			
K6.5	K6.8	0,904204	K6.2	K6.11	0,685319	K1.7	K1.8	0,575578			
K6.8	K8.1	0,903552	K6.7	K6.13	0,682866	K3.4	K6.9	0,569578			
K6.3	K6.5	0,900853	K1.5	K1.8	0,675406	K1.1	K1.7	0,552931			
K6.7	K6.9	0,863374	K1.5	K6.7	0,673675	K1.3	K1.7	0,552931			
K3.4	K4.2	0,847103	K1.8	K6.13	0,673544	K1.3	K6.7	0,538928			
K1.7	K6.7	0,843108	K4.2	K6.13	0,647602	K1.1	K6.7	0,538928			
K6.5	K8.1	0,825229	K1.5	K4.2	0,647564	K1.12	K5.2	0,527211			

ДОДАТОК В

Перелік показників, відібраних в модель та їх характеристика (статистики вказані для показників до їх первинної обробки)

Таблиця В.1
Перелік показників

Назва коефіцієнту	Порядок розрахунку	Коротка характеристика	Min	Max	Avg	Std	Коментар щодо відбору ознаки в модель
1	2	3	4	5	6	7	8
Загальні показники							
K0. Коефіцієнт системності	Активи банку / Активи всього в системі	Характеризує системність банку та його розмір	0,00	0,28	0,01	0,03	Відібраний за допомогою підходу відбору ознак
Характеристика активів							
K1.2. Питома вага дохідних активів	Працюючі активи / Активи банку всього Працюючі активи = Кошти в інших банках + Кредити і заборгованість клієнтів + Цінні папери, що обліковуються за справедливою вартістю через інший сукупний дохід + Цінні папери, що обліковуються за амортизованою собівартістю + Інші фінансові активи	Характеризує якість активів: чим вище, тим більшою є ефективність використання ресурсів і ділова активність банку	0,00	1,00	0,71	0,19	Відібраний за допомогою підходу відбору ознак
K1.3. Коефіцієнт резервів під кредитні втрати	Резерви під кредити / Кредитні вкладення всього	Банки як правило формують більші обсяги резервів під активи нижчої якості, Чим вищим є показник, тим потенційно гіршим є якість кредитного портфелю	- 0,11	12236 ,18	12,88	299, 52	Відібраний за допомогою підходу відбору ознак

Продовження таблиці В.1

1	2	3	4	5	6	7	8
К1.7. Коефіцієнт дохідності активів	Процентний дохід / Дохідні активи всього	Показує скільки одиниць доходу приносить одиниця працюючих активів	0,00	20687 31,01	922,4 1	4368 0,68	Доданий в ході експериментів, так як покращує якість інтерпретації результатів та кластероутворення моделі
К1.9. Частка кредитів в активах	Кредити всього / Активи всього	Характеризує те, наскільки банк орієнтований на кредитування у своїх активних операціях (отримання процентного доходу)	0,00	1,15	0,48	0,23	Доданий в ході експериментів, так як покращує якість інтерпретації результатів та кластероутворення моделі
К1.12. Частка цінних паперів в активах	Цінні папери / Активи всього	Висока частка цінних паперів в портфелі може свідчити про приналежність банку до розряду інвестиційних та/або про наявність вільного доходу	-0,21	0,95	0,15	0,17	Відібраний за допомогою підходу відбору ознак
К1.13. Частка кредитно-інвестиційного портфелю в активах	Кредитно-інвестиційний портфель / Активи всього КІП = Кошти в інших банках + Кредити і заборгованість клієнтів + Цінні папери, що обліковуються за справедливою вартістю та амортизованою собівартістю	Характеризує ризикованість кредитної політики банку	0,00	0,99	0,70	0,19	Доданий в ході експериментів, так як покращує якість інтерпретації результатів та кластероутворення моделі

Продовження таблиці В. 1

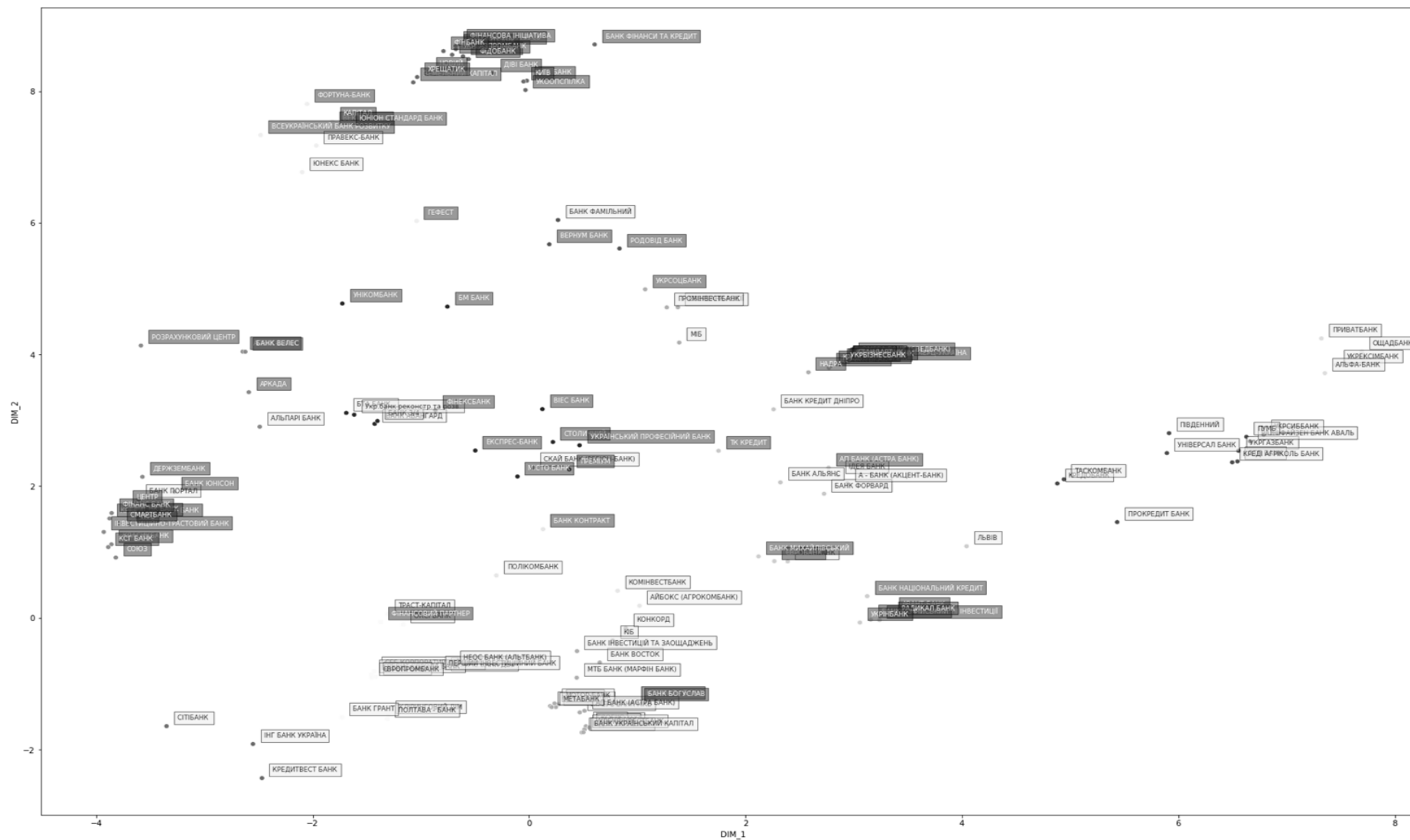
1	2	3	4	5	6	7	8
Характеристика пасивів							
К2.4. Коефіцієнт використання міжбанківських кредитів	Міжбанківські кредити / Депозити всього	Характеризує ступінь залежності ресурсної бази банку від МБК	-0,00	0,98	0,70	0,19	Відібраний за допомогою підходу відбору ознак
К2.6. Коефіцієнт залежності від вкладів фізичних осіб	Вклади фізичних осіб / Зобов'язання всього	Характеризує залежність банку від вкладів фізичних осіб та можливі ризики при паніці вкладників	0,00	0,96	0,36	0,22	Відібраний за допомогою підходу відбору ознак
Характеристика капіталу							
К3.1. Коефіцієнт автономності (адекватності капіталу)	Власний капітал / Чисті активи всього	Характеризує те, наскільки забезпечені ризиковані вклади банку його капіталом	-2,73	1,00	0,26	0,24	Відібраний за допомогою підходу відбору ознак
К3.2. Достатність капіталу за депозитами	Власний капітал / Депозити всього	Характеризує ступінь покриття коштів клієнтів капіталом	- 3079,3 5	26184 68,36	7141, 43	1326 06,6 4	Відібраний за допомогою підходу відбору ознак
К3.3. Коефіцієнт покриття позик	Власний капітал / Позики всього	Характеризує здатність повернути залучені кошти у випадку неповернення кредитів	- 1481,7 2	4322, 49	4,13	102, 30	Відібраний за допомогою підходу відбору ознак
К3.4. Рентабельність капіталу (ROC)	Чистий прибуток / Власний капітал	Характеризує ефективність використання коштів власників	-60,51	145,8 8	0,08	3,66	Відібраний за допомогою підходу відбору ознак
Характеристика ліквідності							
К3.3. Коефіцієнт покриття позик	Власний капітал / Позики всього	Характеризує здатність повернути залучені кошти у випадку неповернення кредитів	- 1481,7 2	4322, 49	4,13	102, 30	Відібраний за допомогою підходу відбору ознак
К3.4. Рентабельність капіталу (ROC)	Чистий прибуток / Власний капітал	Характеризує ефективність використання коштів власників	-60,51	145,8 8	0,08	3,66	Відібраний за допомогою підходу відбору ознак

Закінчення таблиці В. 1

1	2	3	4	5	6	7	8
Характеристика ліквідності							
К5.2. Частка високоліквідних активів	Високоліквідні активи / Активи всього ВЛА = Грошові кошти та їх еквіваленти + Кошти в інших банках + Цінні папери, що обліковуються за справедливою вартістю через інший сукупний дохід	Характеризує здатність банку швидко гасити вимоги кредиторів та вкладників	-1,00	0,99	-0,01	0,56	Відібраний за допомогою підходу відбору ознак
Характеристика фінансової ефективності							
К6.5. Коефіцієнт безризикового покриття доходів	Комісійні доходи / Витрати всього	Характеризує наскільки безризиковий дохід покриває сукупні витрати банку	-76,63	8,38	0,29	1,70	Доданий в ході експериментів, так як покращує якість інтерпретації результатів та кластероутворення моделі
К6.8. Рентабельність витрат	Прибуток (до оподаткування) / Витрати всього	Характеризує ефективність витрат банку, обсяг прибутку на одиницю витрат	-1,00	96,47	0,15	2,79	Доданий в ході експериментів, так як покращує якість інтерпретації результатів та кластероутворення моделі
Характеристика податкового навантаження							
К7.1. Коефіцієнт податкового навантаження на доходи	Податки всього / Доходи всього	Розмір податків на одиницю доходу	0,00	11,31	-0,06	3,71	Відібраний за допомогою підходу відбору ознак

ДОДАТОК Г

Двовимірна карта активних банків станом на IV квартал 2021 року та неплатоспроможних і ліквідованих банків станом на останній період в якому вони були платоспроможними, отримана із застосуванням алгоритму UMAP, з нанесеними мітками банків



ДОДАТОК Д

Лістинг Python-коду для реалізації всіх етапів (*pipeline*) моделювання

Підключення сторонніх бібліотек

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import RobustScaler
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import MaxAbsScaler
from sklearn.preprocessing import Normalizer
from scipy.stats.mstats import winsorize

from pyclustertend import hopkins
from pyclustertend import assess_tendency_by_mean_metric_score
import skdim

from sklearn.feature_selection import VarianceThreshold
from statsmodels.stats.outliers_influence import variance_inflation_factor

import umap
from sklearn.manifold import TSNE
from sklearn.cluster import AgglomerativeClustering
from sklearn import metrics
```

Завантаження та первинна обробка даних

```
banks_data = pd.read_excel('drive/MyDrive/Colab Notebooks/NEW DATA/NEW_COEFF_VIEW_
PERQUARTER_2015-2021.xlsx', sheet_name='TEST #6', index_col=0)

scaler = StandardScaler()
rscaler = RobustScaler()
minmax = MinMaxScaler()
maxabs = MaxAbsScaler()
normilizer = Normalizer()

winzor_rate_1 = 0.02
winzor_rate_2 = 0.05

coef_raw = banks_data.drop(columns = ['Period', 'Group', 'NKB', 'Bank'])

coeffs_noNA = banks_data.dropna().drop(columns = ['Period', 'Group', 'NKB', 'Bank'
])
coeffs_noNA.reset_index(drop=True, inplace=True)

bank_labels_noNA = banks_data.dropna()['Bank']
```

```

bank_labels_noNA.reset_index(drop=True, inplace=True)

bank_periods_noNA = banks_data.dropna()['Period']
bank_periods_noNA.reset_index(drop=True, inplace=True)

bank_group_noNA = banks_data.dropna()['Group']
bank_group_noNA.reset_index(drop=True, inplace=True)

bank_NKB_noNA = banks_data.dropna()['NKB']
bank_NKB_noNA.reset_index(drop=True, inplace=True)

df_scaled_noNA = pd.DataFrame(scaler.fit_transform(coeffs_noNA.values), columns=coeffs_noNA.columns)
df_rscaled_noNA = pd.DataFrame(rscaler.fit_transform(coeffs_noNA.values), columns=coeffs_noNA.columns)
df_normed_noNA = pd.DataFrame(normilizer.fit_transform(coeffs_noNA.values), columns=coeffs_noNA.columns)
df_winzored_noNA = pd.DataFrame(winsorize(coeffs_noNA.values, limits = [winzor_rate_1, winzor_rate_2]), columns=coeffs_noNA.columns)
df_minmaxed_noNA = pd.DataFrame(minmax.fit_transform(coeffs_noNA.values), columns=coeffs_noNA.columns)
df_maxabsed_noNA = pd.DataFrame(maxabs.fit_transform(coeffs_noNA.values), columns=coeffs_noNA.columns)

df_winzored_scaled = pd.DataFrame(scaler.fit_transform(winsorize(coeffs_noNA.values, limits=[winzor_rate_1, winzor_rate_2])), columns=coeffs_noNA.columns)

df_winzored_rscaled = pd.DataFrame(rscaler.fit_transform(winsorize(coeffs_noNA.values, limits=[winzor_rate_1, winzor_rate_2])), columns=coeffs_noNA.columns).drop(columns=drop_cols)

data_list_raw = [coeffs_noNA]

data_list_noNA = [df_scaled_noNA,
                  df_rscaled_noNA,
                  df_normed_noNA,
                  df_winzored_noNA,
                  df_minmaxed_noNA,
                  df_maxabsed_noNA,
                  df_winzored_scaled,
                  df_winzored_normed,
                  df_winzored_rscaled]

```

Оцінка асиметрії

```

for i in data_list_raw:

    skew_columns = (i.skew().sort_values(ascending=False))
    print(f'Average skew: {sum(skew_columns)/skew_columns.shape[0]}')

print()

```

```

for i in data_list_noNA:

    skew_columns = (i.skew().sort_values(ascending=False))
    print(f'Average skew: {sum(skew_columns)/skew_columns.shape[0]}')

print()

```

Оцінка стандартного відхилення

```

for i in data_list_raw:

    std_columns = (i.std().sort_values(ascending=False))
    print(f'Average std: {sum(std_columns)/std_columns.shape[0]:.5f}')

print()

```

```

for i in data_list_noNA:

    std_columns = (i.std().sort_values(ascending=False))
    print(f'Average std: {sum(std_columns)/std_columns.shape[0]:.5f}')

print()

```

Тест Хопкінса

```

hopkins_n = 1500

uniform = pd.DataFrame({'x' : np.random.uniform(-1, 1, 2000),
                        'y' : np.random.uniform(-1, 1, 2000)})

print(f'Hopkins test on uniform data: {hopkins(uniform, hopkins_n):.5f}')

for i in data_list_raw:

    print(f'Hopkins test: {hopkins(i, hopkins_n):.5f}')

print()

for i in data_list_noNA:

    print(f'Hopkins test: {hopkins(i, hopkins_n):.5f}')

print()

```

Оцінка тенденції

```

for i in data_list_raw:

    print(f'Tendency: {assess_tendency_by_mean_metric_score(i, 100):.5f}')

```

```

print()

for i in data_list_noNA:

    print(f'Tendency: {assess_tendency_by_mean_metric_score(i, 100):.5f}')

print()

```

Аналіз кореляції

```

def get_redundant_pairs(df):
    '''Get diagonal and lower triangular pairs of correlation matrix'''
    pairs_to_drop = set()
    cols = df.columns
    for i in range(0, df.shape[1]):
        for j in range(0, i+1):
            pairs_to_drop.add((cols[i], cols[j]))
    return pairs_to_drop

def get_top_abs_correlations(df, n=48):
    au_corr = df.corr().abs().unstack()
    labels_to_drop = get_redundant_pairs(df)
    au_corr = au_corr.drop(labels=labels_to_drop).sort_values(ascending=False)
    return au_corr[0:n]

print("Top Absolute Correlations")
print(get_top_abs_correlations(coeffs_noNA, 64))

```

Кореляційна теплова матриця

```

correl = abs(banks_data.corr())

plt.figure(figsize=(15,12))
sns.heatmap(correl, annot=True, cmap=plt.cm.Reds)
plt.show()

```

Mean Absolute Distance (MAD)

```

MAD = np.sum(np.abs(df_winzored_noNA.values - np.mean(df_winzored_noNA.values, axis = 0)), axis = 0) / df_winzored_noNA.values.shape[0]

fig, axes = plt.subplots(1, 1, figsize = (40, 10))

bar = sns.barplot(x=np.arange(MAD.shape[0]), y=MAD, color='grey')
axes.set_xticklabels(coeffs_noNA.columns)

for p in bar.patches:
    bar.annotate(format(p.get_height(), '.4f'),
                  (p.get_x() + p.get_width() / 2., p.get_height()),
                  ha = 'center', va = 'center',

```

```

        xytext = (0, 9),
        textcoords = 'offset points')

plt.show()

```

Variance Threshold (VT)

```

## VT CALCULATION

vt = VarianceThreshold()
vt.fit(df_winzored_noNA)

## PLOTTING

fig, axes = plt.subplots(1, 1, figsize = (40, 10))

bar = sns.barplot(x=np.arange(df_winzored_noNA.shape[1]), y=vt.variances_, color='
grey')
axes.set_xticklabels(df_winzored_noNA.columns)

for p in bar.patches:
    bar.annotate(format(p.get_height(), '.4f'),
                  (p.get_x() + p.get_width() / 2., p.get_height()),
                  ha = 'center', va = 'center',
                  xytext = (0, 9),
                  textcoords = 'offset points')

plt.show()

```

Variance Inflation Factor (VIF)

```

# MULTICOLLINEARITY ESTIMATION WITH VIF

## VIF estimation func

def calc_vif(X):
    vif = pd.DataFrame()
    vif['Variables'] = X.columns
    vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1]
)]

    return vif

## VIF estimation

VIF = calc_vif(df_winzored_noNA)

## PLOTTING

fig, axes = plt.subplots(1, 1, figsize = (40, 10))

```

```

bar = sns.barplot(x=np.arange(df_winzored_noNA.shape[1]), y=VIF['VIF'].values, color='grey')
axes.set_xticklabels(df_winzored_noNA.columns)

for p in bar.patches:
    bar.annotate(format(p.get_height(), '.4f'),
                  (p.get_x() + p.get_width() / 2., p.get_height()),
                  ha = 'center', va = 'center',
                  xytext = (0, 9),
                  textcoords = 'offset points')

```

SPEC

Повна реалізація розміщена у репозиторії:

<https://github.com/danilkolikov/fsfc/blob/master/fsfc/generic/SPEC.py>

```

SPEC = spec(df_winzored_noNA.values)

fig, axes = plt.subplots(1, 1, figsize = (40, 10))

bar = sns.barplot(x=np.arange(SPEC.shape[0]), y=SPEC, color='grey')
axes.set_xticklabels(df_winzored_noNA.columns)

for p in bar.patches:
    bar.annotate(format(p.get_height(), '.4f'),
                  (p.get_x() + p.get_width() / 2., p.get_height()),
                  ha = 'center', va = 'center',
                  xytext = (0, 9),
                  textcoords = 'offset points')

plt.show()

```

Зниження розмірності

```

test_data = df_winzored_scaled

tsne = TSNE(n_components=2,
            perplexity=15,
            learning_rate=500,
            n_iter=1000,
            init='pca',
            method='barnes_hut',
            verbose = 1,
            random_state=42,
            n_jobs = -1).fit_transform(test_data)

test_data = df_winzored_rscaled

umap_1 = umap.UMAP(n_components=2,
                  n_neighbors=20,
                  min_dist=0.0,

```

```
metric='correlation',
verbose=1).fit_transform(test_data)
```

Побудова двовимірної проєкції

```
reducer = umapp_1

df_reduce = pd.DataFrame(reducer, columns=['DIM_1', 'DIM_2'])

fig, axes = plt.subplots(1, 1, figsize = (20, 10))

sns.scatterplot(data=df_reduce, x='DIM_1', y='DIM_2', legend='full', palette='Spectral')

plt.show()
```

Кластеризація, оцінка якості та візуалізація

```
agglomerative = AgglomerativeClustering(n_clusters=15, distance_threshold=None).fit(reducer)
labels = agglomerative.labels_

print(metrics.silhouette_score(reducer, labels, metric='euclidean'))
print(metrics.davies_bouldin_score(reducer, labels))
print(metrics.calinski_harabasz_score(reducer, labels))

df_reduce = pd.DataFrame(reducer, columns=['DIM_1', 'DIM_2'])

fig, axes = plt.subplots(1, 1, figsize = (20, 10))

sns.scatterplot(data=df_reduce, x='DIM_1', y='DIM_2', legend='full', hue=labels, palette='Spectral')

plt.show()
```

Міграція банків, динамічний аналіз показників банківської діяльності, візуалізація

```
df = pd.DataFrame(bank_labels_noNA, columns=['Bank'])
df1 = df[df['Bank'] == 'РАЙФФАЙЗЕН БАНК АВАЛЬ']
df2 = df[df['Bank'] == 'АКОРДБАНК']
df3 = df[df['Bank'] == 'МІСТО БАНК']
df4 = df[df['Bank'] == 'АЛЬПАРИ БАНК']

n = 8

df_reduce = pd.DataFrame(reducer, columns=['DIM_1', 'DIM_2'])

fig, axes = plt.subplots(1, 1, figsize = (20, 10))

# sns.scatterplot(data=df_reduce, x='DIM_1', y='DIM_2', legend='full', palette=['grey'])
```

```

plt.scatter(df_reduce['DIM_1'], df_reduce['DIM_2'], c='grey', s=10)

plt.scatter(df_reduce.iloc[df1.index].iloc[-
n:,0], df_reduce.iloc[df1.index].iloc[-
n:,1], c='orange', s=100, zorder=1, edgecolors='k', linewidths=.5)
plt.plot(df_reduce.iloc[df1.index].iloc[-n:,0], df_reduce.iloc[df1.index].iloc[-
n:,1], zorder=2)
plt.scatter(df_reduce.iloc[df1.index].iloc[-n,0], df_reduce.iloc[df1.index].iloc[-
n,1], c='r', s=200, edgecolors='k')

plt.scatter(df_reduce.iloc[df2.index].iloc[-
n:,0], df_reduce.iloc[df2.index].iloc[-
n:,1], c='orange', s=100, zorder=1, edgecolors='k', linewidths=.5)
plt.plot(df_reduce.iloc[df2.index].iloc[-n:,0], df_reduce.iloc[df2.index].iloc[-
n:,1], zorder=2)
plt.scatter(df_reduce.iloc[df2.index].iloc[-n,0], df_reduce.iloc[df2.index].iloc[-
n,1], c='r', s=200, edgecolors='k')

plt.scatter(df_reduce.iloc[df3.index].iloc[-
n:,0], df_reduce.iloc[df3.index].iloc[-
n:,1], c='orange', s=100, zorder=1, edgecolors='k', linewidths=.5)
plt.plot(df_reduce.iloc[df3.index].iloc[-n:,0], df_reduce.iloc[df3.index].iloc[-
n:,1], zorder=2)
plt.scatter(df_reduce.iloc[df3.index].iloc[-n,0], df_reduce.iloc[df3.index].iloc[-
n,1], c='r', s=200, edgecolors='k')

plt.scatter(df_reduce.iloc[df4.index].iloc[-
n:,0], df_reduce.iloc[df4.index].iloc[-
n:,1], c='orange', s=100, zorder=1, edgecolors='k', linewidths=.5)
plt.plot(df_reduce.iloc[df4.index].iloc[-n:,0], df_reduce.iloc[df4.index].iloc[-
n:,1], zorder=2)
plt.scatter(df_reduce.iloc[df4.index].iloc[-n,0], df_reduce.iloc[df4.index].iloc[-
n,1], c='r', s=200, edgecolors='k')

plt.show()

```

Додаткова підготовка даних до візуалізації кластеризованих даних з маркерами

```

numClusters = set(labels)

df_test = test_data.copy(deep=True)

df_test['Bank'] = bank_labels_noNA
df_test['Cluster'] = labels
df_test['Period'] = bank_periods_noNA
df_test['Group'] = bank_group_noNA

df_test_now = df_test[df_test['Period'] == '2021_4']
df_test_now.reset_index(drop=True, inplace=True)

df_reduce_now = df_reduce.copy(deep=True)

```

```

df_reduce_now['Bank'] = bank_labels_noNA
df_reduce_now['Cluster'] = labels
df_reduce_now['Period'] = bank_periods_noNA

df_reduce_now_start = df_reduce_now[bank_periods_noNA == '2021_1']
df_reduce_now_end = df_reduce_now[bank_periods_noNA == '2021_4']

df_reduce_badonly = df_reduce_now[bank_group_noNA == 0]

df_reduce_badonly.sort_values(by=['Bank', 'Period'], inplace=True)
df_reduce_badonly.reset_index(drop=True, inplace=True)

df_reduce_badonly_end = df_reduce_badonly.copy(deep=True)

for i in range(1, df_reduce_badonly.shape[0]):

    if df_reduce_badonly.iloc[i, 2] == df_reduce_badonly.iloc[i-1, 2]:

        df_reduce_badonly_end.drop([i], inplace=True)

    else:

        pass

fig, axes = plt.subplots(1, 1, figsize = (30, 20))

sns.scatterplot(data=df_reduce_now_end, x='DIM_1', y='DIM_2', legend='full', hue=df_test['Cluster'], palette='Spectral', s=200)

for i in range(df_reduce_now_end.shape[0]):
    plt.text(x=df_reduce_now_end['DIM_1'].iloc[i]+0.1, y=df_reduce_now_end['DIM_2'].iloc[i]+0.1, s=df_reduce_now_end['Bank'].iloc[i],
            fontdict=dict(color='red',size=8),
            bbox=dict(facecolor='yellow',alpha=0.5))

plt.show()

```

Виведення переліку кластерів з банками, що вони вміщують + розрахунок кількості неактивних банків, що потрапило до кожного з кластерів

```

test = []

for i in numClusters:

    if df_test_now[df_test_now['Cluster'] == i]['Bank'].shape[0] > 0:

        test.append(df_test_now[df_test_now['Cluster'] == i]['Bank'].values)
        print(f'Cluster {i}:')
        print(df_test_now[df_test_now['Cluster'] == i]['Bank'].values)
        print('-----')

```

```

else:

    pass

test = []

df_reduce_badonly_end['Cluster'].value_counts()

```

Візуалізація банків з маркерами по активним банкам, та неактивним банка за останній період в якому вони були платоспроможні

```

fig, axes = plt.subplots(1, 1, figsize = (30, 20))

sns.scatterplot(data=df_reduce_badonly_end, x='DIM_1', y='DIM_2', hue='Cluster', legend='full', palette='Spectral')
sns.scatterplot(data=df_reduce_now_end, x='DIM_1', y='DIM_2', hue='Cluster', legend='full', palette='Spectral')

for i in range(df_reduce_now_end.shape[0]):
    plt.text(x=df_reduce_now_end['DIM_1'].iloc[i]+0.1, y=df_reduce_now_end['DIM_2'].iloc[i]+0.1, s=df_reduce_now_end['Bank'].iloc[i],
             fontdict=dict(color='red',size=8),
             bbox=dict(facecolor='yellow',alpha=0.5))

for i in range(df_reduce_badonly_end.shape[0]):
    plt.text(x=df_reduce_badonly_end['DIM_1'].iloc[i]+0.1, y=df_reduce_badonly_end['DIM_2'].iloc[i]+0.1, s=df_reduce_badonly_end['Bank'].iloc[i],
             fontdict=dict(color='white',size=8),
             bbox=dict(facecolor='red',alpha=0.5))

plt.show()

```

ДОДАТОК Е

Розподіл платоспроможних станом на IV квартал 2021 року банків по кластерам, відповідно до проведеного експериментального дослідження

Кластер 1: ['ПРИВАТБАНК', 'ОЩАДБАНК', 'УКРЕКСІМБАНК', 'АЛЬФА-БАНК']

Кластер 2: ['УКРГАЗБАНК', 'РАЙФФАЙЗЕН БАНК АВАЛЬ', 'УКРСИББАНК', 'ОТП БАНК', 'КРЕДІ АГРІКОЛЬ БАНК', 'КРЕДОБАНК', 'ПРОКРЕДИТ БАНК', 'ПУМБ', 'УНІВЕРСАЛ БАНК', 'ПІВДЕННИЙ', 'ТАСКОМБАНК']

Кластер 3: ['МЕГАБАНК', 'БАНК ЛЬВІВ']

Кластер 4: ['ІДЕЯ БАНК', 'БАНК ФОРВАРД', 'БАНК КРЕДИТ ДНІПРО', 'А - БАНК (АКЦЕНТ-БАНК)', 'БАНК АЛЪЯНС']

Кластер 5: ['СБЕРБАНК РОСІЇ', 'ПРОМІНВЕСТБАНК', 'МІБ']

Кластер 6: ['КОНКОРД', 'КІБ', 'ПІРЕУС БАНК МКБ', 'БАНК ВОСТОК', 'БАНК ІНВЕСТИЦІЙ ТА ЗАОЩАДЖЕНЬ', 'АКОРДБАНК', 'МТБ БАНК (МАРФІН БАНК)', 'ГЛОБУС', 'БАНК СІЧ', 'ІНДУСТРІАЛБАНК', 'РАДАБАНК', 'РВС БАНК', 'КРИСТАЛБАНК', 'УКРБУДІНВЕСТБАНК', 'МОТОР-БАНК', 'АП БАНК (АСТРА БАНК)', 'БАНК УКРАЇНСЬКИЙ КАПІТАЛ', 'МЕТАБАНК']

Кластер 7: ['ПОЛІКОМБАНК', 'АЙБОКС (АГРОКОМБАНК)', 'КОМІНВЕСТБАНК']

Кластер 8: ['ДОЙЧЕ БАНК ДБУ', 'СЕБ КОРПОРАТИВНИЙ БАНК', 'КРЕДИТ ЄВРОПА БАНК', 'КЛПРИНГОВИЙ ДІМ', 'ПОЛТАВА - БАНК', 'ПЕРШИЙ ІНВЕСТИЦІЙНИЙ БАНК', 'БАНК ГРАНТ', 'АСВІО БАНК', 'НЕОС БАНК (АЛЬТБАНК)', 'ЄВРОПРОМБАНК', 'ОКСІ БАНК', 'ТРАСТ-КАПІТАЛ']

Кластер 9: ['БТА БАНК', 'БАНК АВАНГАРД', 'БАНК 3/4', 'Укр.банк реконстр.та розв', 'СКАЙ БАНК']

Кластер 10: ['СІТІБАНК', 'ІНГ БАНК УКРАЇНА', 'КРЕДИТВЕСТ БАНК']

Кластер 11: ['БАНК ПОРТАЛ']

Кластер 12: ['АЛЬПАРІ']

Кластер 13: ['БАНК ФАМІЛЬНИЙ']

Кластер 14: ['ПРАВЕКС-БАНК', 'ЮНЕКС-БАНК']

Кластер 15: не містить платоспроможних банків на досліджуваний період.

ДОДАТОК Ж

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

У наукових фахових виданнях України:

1. Корчинський В. В., Долінський Л. Б. Ідентифікація та кількісне оцінювання кредитного ризику комерційного банку. Економічний аналіз: зб. наук. праць. 2016. Т. 25, № 1. С. 180–189. (1,12 друк. арк., з них 1 друк. арк. належить особисто автору: побудова класифікації ризиків, побудова моделі дискримінантного аналізу позичальників для оцінки кредитного ризику банків, побудовано карту кредитного ризику);
2. Великоіваненко Г. І., Корчинський В. В., Чернишова В. В. Дослідження ефекту перенавчання нейронних мереж на прикладі задачі аплікаційного скорингу. *Нейро-нечіткі технології моделювання в економіці*. 2016. № 5. С. 3–23. (1,22 друк. арк., з них 1,15 друк. арк. належить особисто автору: проведено моделювання кредитного ризику позичальників-фізичних осіб та оцінка адекватності моделей дослідження);
3. Корчинський В. В., Чернишова В. В. Побудова скорингової моделі на основі нейронної мережі типу Персептрон. *Український журнал прикладної економіки*. 2016. Т. 1, № 4. С. 46–55. (1,25 друк. арк., з них 1,12 друк. арк. належить особисто автору: проведення аналізу та нормалізації статистичних даних, побудова нейронної мережі та дослідження ефективності та якості її прогнозних властивостей);
4. Корчинський В. В., Долінський Л. Б., Мірошніченко І. В. Кластерний аналіз надійності банківських установ в Україні. *Моделювання та інформаційні системи в економіці: зб. наук. праць*. 2017. № 93. С. 49–73. (1,2 друк. арк., з них 1,08 друк. арк. належить особисто автору: побудовано нейронну мережу типу карт самоорганізації Кохонена для кластерного аналізу надійності комерційних банків);
5. Лаговський В. В., Долінський Л. Б., Корчинський В. В. Особливості кредитно-інвестиційної діяльності на фінансовому ринку України. *Економічний вісник: науковий фаховий журнал*. 2020. Т. 25, № 4. С. 74–88. (0,58 друк. арк., з них 0,52 друк. арк. належить особисто автору: проведено детальний статистичний

аналіз ряду показників, що характеризують динаміку економічного розвитку і розвитку фінансового сектору, проведено кореляційних аналіз цих показників і визначено істотну кореляційну залежність між ними);

У зарубіжному періодичному науковому виданні, що зареєстроване у міжнародній наукометричній базі Scopus:

6. Korchynskyi V., Velykoivanenko H. Application of Clustering in the Dimensionality Reduction Algorithms for Separation of Financial Status of Commercial Banks in Ukraine. *Universal Journal of Accounting and Finance*. 2022. Vol. 10, no. 1. P. 148–60. URL: <https://doi.org/10.13189/ujaf.2022.100116>. (1,18 друк. арк., з них 1,06 друк. арк. належить особисто автору: побудовано систему алгоритмів машинного навчання для проведення кластерного аналізу фінансового стану комерційних банків України, проведено аналіз фінансового стану комерційних банків на основі результатів зниження роозмірності);

В інших виданнях:

7. Корчинський В. В. Побудова системи ризик-менеджменту в комерційному банку. *Інноваційний ресурс соціально-економічного прогресу України*: зб. доп. 83-ї наук. конф. студ. КНЕУ, м. Київ, 25–27 квіт. 2016 р. Київ, 2016. С. 290–293. (0,17 друк. арк., з них 0,15 друк. арк. належить особисто автору: запропоновано дискримінантну модель для оцінки надійності позичальника і кредитного ризику) (форма участі: очна);

8. Корчинський В. В., Чергишова В. В. Скорингова модель оцінки кредитоспроможності клієнтів банків на основі інструментарію нечіткої логіки. *Інноваційний ресурс соціально-економічного прогресу України*: зб. доп. 83-ї наук. конф. студ. КНЕУ, м. Київ, 25–27 квіт. 2016 р. Київ, 2016. С. 602–604. (0,12 друк. арк., з них 0,1 друк. арк. належить особисто автору: запропоновано скорингову модель для оцінки кредитоспроможності клієнтів банків) (форма участі: очна);

9. Долінський Л. Б., Корчинський В. В. Відбір кількісних показників для проведення кластеризації вітчизняних банків в аспекті фінансової стійкості. *Економіко-математичне моделювання*: зб. мат. Першої нац. наук.-

матод. конф, м. Київ, 30 верес. – 1 жовт. 2016 р. Київ, 2016. С. 122–123. (0,12 друк. арк., з них 0,1 друк. арк. належить особисто автору: запропоновано та обґрунтовано перелік показників для подальшого проведення кластеризації вітчизняних банків) (форма участі: очна);

10. Матвійчук А. В., Корчинський В. В., Чернишова В. В. Алгоритм побудови скорингової моделі на основі нейронної мережі персептронного типу. *КМОСС: Матеріали II Всеукр. науково-техн. конф., м. Дніпро, 1–3 листоп. 2016 р. Дніпро, 2016. С. 202–203. (0,12 друк. арк., з них 0,1 друк. арк. належить особисто автору: розглянуто алгоритм побудови скорингової моделі для оцінки кредитоспроможності позичальників) (форма участі: заочна);*

11. Долінський Л. Б., Корчинський В. В. Кластерний аналіз надійності вітчизняних банківських установ. *Інноваційний ресурс соціально-економічного розвитку України: зб. доп. 84-ї наук. конф. студ. КНЕУ, м. Київ, 24 квіт. – 23 трав. 2017 р. Київ, 2017. С. 389–391. (0,11 друк. арк., з них 0,1 друк. арк. належить особисто автору: запропоновано та розглянуто концептуальний підхід до проведення кластерного аналізу надійності вітчизняних банків) (форма участі: очна);*

12. Долінський Л. Б., Корчинський В. В. Застосування сучасних інформаційних технологій для кластеризації українських комерційних банків в аспекті фінансової стійкості. *Математичні методи, моделі та інформаційні технології в економіці: Матеріали VI Міжнар. науково-метод. конф., м. Чернівці, 18–19 квіт. 2019 р. Чернівці, 2019. С. 88–92. (0,2 друк. арк., з них 0,18 друк. арк. належить особисто автору: обґрунтована доцільність застосування сучасних та інноваційних інформаційних технологій для проведення кластеризації комерційних банків) (форма участі: заочна);*

13. Долінський Л. Б., Корчинський В. В. Аналіз особливостей кредитно-інвестиційної діяльності банків в Україні. *Економіка підприємства: сучасні проблеми теорії та практики: Матеріали дев'ятої міжнар. наук.-практ. конф, м. Одеса, 11–12 верес. 2020 р. Одеса, 2020. С. 105–106. (0,12 друк. арк., з них 0,1*

друк. арк. належить особисто автору: проведений аналіз фінансового стану банківського сектору України в динаміці) (форма участі: заочна);

14. Великоіваненко Г. І., Корчинський В. В. Моделювання фінансового стану комерційних банків України на основі інструментів та методів машинного навчання. *Актуальні проблеми економіки, фінансів, обліку та права в умовах сучасних викликів*: зб. тез доп. міжнар. науково-практ. конф., м. Полтава, 25 серп. 2021 р. Полтава, 2021. С. 53–54. (0,2 друк. арк., з них 0,18 друк. арк. належить особисто автору: запропоновано та розглянуто підхід до аналізу фінансового стану комерційних банків за допомогою інструментів машинного навчання) (форма участі: заочна);

15. Великоіваненко Г. І., Корчинський В. В. Аналіз стабільності показників фінансового стану комерційних банків України на основі методів машинного навчання. *КМОСС: Матеріали VII Міжнар. науково-техн. конф.*, м. Дніпро, 3–5 листоп. 2021 р. Дніпро, 2021. С. 93–94. (0,12 друк. арк., з них 0,1 друк. арк. належить особисто автору: проведено аналіз стабільності показників комерційних банків в часі і розглянуто залежність між динамікою показників та фінансовим станом банків) (форма участі: заочна).

«ЗАТВЕРДЖУЮ»:
 Проректор з наукової роботи
 Київського національного
 економічного університету
 імені Вадима Гетьмана,
 д.с.н. проф. Лариса Антонюк
 «24» квітня 2023 р.

«ЗАТВЕРДЖУЮ»:
 Проректор з науково-педагогічної
 роботи Київського національного
 економічного університету
 імені Вадима Гетьмана
 д.с.н. проф. Анатолій Колот
 «24» квітня 2023 р.

ДОВІДКА

про впровадження у навчальний процес
 результатів дисертаційної роботи

Корчинського Владислава Вікторовича на тему:

«Кластерне моделювання фінансового стану комерційних банків України»

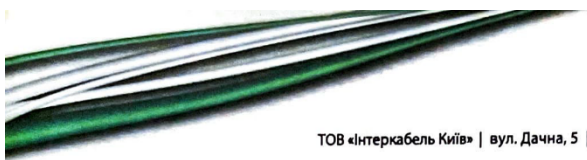
Результати дисертаційної роботи «Кластерне моделювання фінансового стану комерційних банків України» здобувача наукового ступеня доктора філософії Корчинського Владислава Вікторовича застосовуються у навчальному процесі кафедри математичного моделювання та статистики Київського національного економічного університету імені Вадима Гетьмана.

Зокрема, запропонований дисертантом методологічний підхід до аналізу фінансового стану комерційних банків України на підґрунті алгоритмів зниження розмірності t-SNE та UMAP, та ієрархічної агломеративної кластеризації типу Уорда використовується в рамках дисципліни «Machine Learning» для студентів, які навчаються за освітньо-професійною програмою підготовки здобувачів вищої освіти на другому (магістерському) рівні «Економічна кібернетика і Дата Сайнс».

Також, розроблені Корчинським В. В. методологічні засади та підходи до побудови системи моделей та алгоритмів машинного навчання, відбору ознак, попередньої обробки та дослідження даних для подальшого кластерного моделювання фінансового стану вітчизняних комерційних банків використовуються в рамках дисциплін «Кількісні методи в економіці (в R)» та «Інтелектуальні технології моделювання у прийнятті рішень» другого (магістерського) рівня освітньо-професійної програми «Економічна кібернетика і Дата Сайнс».

Завідувач кафедри
 математичного моделювання
 та статистики
 КНЕУ імені Вадима Гетьмана,
 к.ф.-м.н., професор

Галина Великоіваненко



УЧАСНИК SKB-GROUP

ТОВ «Інтеркабель Київ» | вул. Дачна, 5 | 08112, с. Капитанівка, Київська область | Україна
Поштова адреса: а/с 71 | 03115 Київ | Україна
Тел +38 (044) 221-54-05 | Факс +38 (044) 221 54-13
sales@interkabel.ua | interkabel.ua



№ 161/1.1-07 від « 10 » березня 2023 року

ДОВІДКА

Про впровадження результатів дисертаційної роботи

Корчинського Владислава Вікторовича

«Кластерне моделювання фінансового стану комерційних банків України»

Методологічний підхід до оцінювання фінансового стану комерційних банків України, їх діагностування, розроблений та запропонований в дисертаційній роботі Корчинського Владислава Вікторовича, що базується на інструментарії машинного навчання, алгоритмах зниження розмірності та кластеризації, безпосередньо представляє практичний інтерес.

Розроблена система моделей та алгоритмів застосовується при аналізі банківського сектору та діагностуванні окремих банківських установ з точки зору їх фінансового стану, надійності, стійкості, ефективності з метою визначення потенційних банків-партнерів.

Генеральний директор
ТОВ «Інтеркабель Київ»



О.С.Савченко

Документ №11/08-23

від 8 березня 2023 року

ДОВІДКА

Про впровадження результатів дисертаційної роботи

Корчинського Владислава Вікторовича

«Кластерне моделювання фінансового стану комерційних банків України»

Методологічний підхід до оцінювання фінансового стану комерційних банків України, їх діагностування, розроблений та запропонований в дисертаційній роботі Корчинського Владислава Вікторовича, що базується на інструментарії машинного навчання, алгоритмах зниження розмірності та кластеризації, безпосередньо представляє практичний інтерес та застосовується для підтримки операційної та фінансової діяльності Wild Web Art OÜ.

Основною діяльністю компанії є IT послуги та консультації клієнтам по всьому світу, однак вся внутрішня операційна та фінансова діяльність зосереджена на території України. Для диверсифікації активів, компанія інвестує переважно у вітчизняні фінансові інструменти (з метою підтримки національної економіки). Одним з напрямків диверсифікації є довгострокові вкладення коштів на депозитні рахунки. Розроблена система моделей та алгоритмів застосовується при аналізі банківського сектору та діагностуванні окремих банківських установ з точки зору їх фінансового стану, надійності, стійкості, ефективності з метою вибору банків-партнерів.

Виконавчий директор,
управляючий партнер



Лаврінчук С. В.