

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА**

**Навчально-науковий інститут
«Інститут інформаційних технологій в економіці»**

Кафедра математичного моделювання та статистики

Освітньо-професійна програма	Економічна кібернетика
Галузь знань	05 Соціальні та поведінкові науки
Спеціальність	051 Економіка

Форма навчання: очна (денна)

КВАЛІФІКАЦІЙНА БАКАЛАВРСЬКА РОБОТА

на тему «Прогнозування розвитку електронної торгівлі в Україні»

(назва теми)

здобувача Миська Данііла Сергійовича

(ПІБ, підпис)

Науковий керівник: кандидат економічних наук,
доцент Катуніна О. С.

(науковий ступінь, учене звання, ПІБ)

(підпис)

**Робота допущена до захисту перед екзаменаційною
комісією з атестації здобувачів вищої освіти (ЕК)**

Завідувач кафедри кандидат фізико-математичних наук,
професор Великоіваненко Г.І.

Київ 2023

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДВНЗ «КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА»

Навчально-науковий інститут / факультет
«Інститут інформаційних технологій в економіці»
Кафедра математичного моделювання та статистики

ОСВІТНЬО-ПРОФЕСІЙНА ПРОГРАМА

Економічна кібернетика

СПЕЦІАЛЬНІСТЬ

05 Соціальні та поведінкові науки

СПЕЦІАЛІЗАЦІЯ (за наявності)

051 Економіка

ПОГОДЖЕНО

Керівник проектної групи (гарант)
освітньо-професійної програми
О.М.Притоманова

(підпис) (ініціали, прізвище)

20__ р.

ЗАТВЕРДЖУЮ

Завідувач кафедри

Г.І.Великоіваненко

(підпис) (ініціали, прізвище)

20__ р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

здобувачу вищої освіти **Миську Данілу Сергійовичу**
очної (денної) форми навчання

на підготовку кваліфікаційної бакалаврської роботи
на тему «Прогнозування розвитку електронної торгівлі в Україні.»

Тему затверджено наказом ректора Університету від " " 2023 р. №

Кваліфікаційна бакалаврська робота виконується на матеріалах

ТОВ «АГРОС ЮА»

План кваліфікаційної бакалаврської роботи

Розділ 1	Теоретичні засади дослідження та моделювання розвитку електронної торгівлі <i>(назва розділу)</i>
Розділ 2	Математичні моделі аналізу та прогнозування розвитку електронної торгівлі <i>(назва розділу)</i>
Розділ 3	Моделювання електронної торгівлі методами машинного навчання <i>(назва розділу)</i>
Об'єкт дослідження:	Об'єктом дослідження є електронна торгівля як складова інформаційно-мережевої економіки.

Предмет дослідження:	Предметом дослідження є математичні моделі та методи аналізу та прогнозування обсягів електронної торгівлі в економіці країни під впливом низки факторів економічного середовища.
Мета кваліфікаційної бакалаврської роботи:	Метою виконання роботи є побудова та експериментальне використання моделей прогнозування розвитку електронної торгівлі в Україні за допомогою методів інтелектуального аналізу даних та математичного моделювання.

Конкретні завдання, які здобувач повинен виконати для досягнення поставленої мети:

У розділі 1	Проаналізувати структуру та тенденції ринку електронної торгівлі в Україні.
У розділі 2	Вивчити та узагальнити досвід застосування алгоритмів інтелектуального аналізу даних для аналізу та прогнозування розвитку для аналізу процесів формування ринку електронної торгівлі.
У розділі 3	Побудувати систему моделей прогнозування обсягів електронної торгівлі на базі інструментів аналізу часових рядів.

**Завдання підготував
науковий керівник**

(підпис)

О.С. Катуніна

(ініціали, прізвище)

«_____» _____ 20__ р.

**Завдання одержав
здобувач**

(підпис)

Д.С. Мисько

(ініціали, прізвище)

«_____» _____ 20__ р.

РЕФЕРАТ

Кваліфікаційна робота містить 23 ілюстрації, 3 таблиці та 37 джерела, які були використані під час дослідження.

«Прогнозування розвитку електронної торгівлі в Україні»

Об'єкт дослідження: Об'єктом дослідження є електронна торгівля як складова інформаційно-мережевої економіки.

Предмет дослідження: Предметом дослідження є математичні моделі та методи аналізу та прогнозування обсягів електронної торгівлі в економіці країни під впливом низки факторів економічного середовища.

Мета роботи: Метою виконання роботи є побудова та експериментальне використання моделей прогнозування розвитку електронної торгівлі в Україні за допомогою методів інтелектуального аналізу даних та математичного моделювання.

Дослідження: для вивчення обсягів електронної торгівлі застосовано динамічне та статичне моделювання, використані графічні і статистичні методи обробки інформації, методи перевірки стаціонарності часових рядів, ARIMA моделювання одномірних часових рядів, побудовано моделі множинної лінійної регресії для аналізу.

Отримані результати дослідження мають важливе значення для розвитку економіки та бізнесу у сфері електронної комерції. Вони надають практичні інсайти та інформацію для прийняття обґрунтованих рішень щодо стратегій розвитку та управління. Отримані результати можуть бути використані для прогнозування економічних процесів, а також для визначення оптимальних шляхів збільшення обороту в електронній комерції.

Рік виконання кваліфікаційної роботи – 2023.

Рік захисту роботи – 2023.

Ключові слова: *аналіз даних, моделювання, прогнозування, економічні показники, електронна торгівля, електронна комерція, ARIMA.*

В і д г у к

про кваліфікаційну бакалаврську роботу
здобувача навчально-наукового інституту "Інститут інформаційних технологій в
економіці Київського національного економічного університету імені Вадима
Гетьмана
освітньо-професійної програми "Економічна кібернетика"
Миська Данііла Сергійовича
на тему «Прогнозування розвитку електронної торгівлі в Україні»

1. Актуальність теми. В умовах цифрових трансформацій суспільства зростає актуальність стрімкого впровадження сучасних інтелектуальних технологій прогнозування розвитку електронної торгівлі в Україні з урахуванням системи передумов та низки факторів економічного середовища. Електронна комерція стає однією із основних форм економічної діяльності, стрімко збільшується її частка на ринку, тому підприємства та компанії, що все більшою мірою орієнтують свою ринкову діяльність на задоволення потреб кінцевого споживача, залучають новітні інформаційні та модельні технології для прогнозування розвитку сфери електронної торгівлі. Отже, актуальними задачами є аналіз та моделювання обсягів електронної торгівлі в економіці країни та дослідження факторів, що зумовлюють динамічні тенденції цих процесів.

2. Позитивні риси кваліфікаційної бакалаврської роботи. В роботі систематизовано сучасні підходи до моделювання динаміки розвитку електронної торгівлі; проаналізовано структуру та особливості електронної торгівлі в Україні; здійснено огляд ефективних інструментів дослідження ринку електронної торгівлі; розроблено та реалізовано систему моделей прогнозування обсягових показників електронної торгівлі в економіці країни, дослідження та вибору факторів, що впливають на формування ринку електронної комерції.

3. Наявність самостійних розробок автора. У кваліфікаційній бакалаврській роботі автором розв'язано задачу моделювання та прогнозування обсягів електронної торгівлі в Україні, здійснено комп'ютерну реалізацію комплексу моделей. Запропоновані моделі реалізовано на реальних даних, що дає можливість отримати практичні рекомендації для систем управління макро- та мезо-економічного рівня, а також учасників відповідних ринків.

4. Цінність теоретичних висновків та практичних рекомендацій. Запропонована система моделей є цілісним комплексом інструментів дослідження та прогнозування обсягів електронної торгівлі, а тому є затребуваною учасниками відповідних ринків та може використовуватись підприємствами та компаніями, що проводять дослідження ринків. Розглянуті моделі можуть бути ефективно впроваджені в практику дослідницької, управлінської та консалтингової діяльності.

5. Наявність недоліків. В системі розглянутих в роботі моделей прогнозування обсягів електронної торгівлі було б доцільним більшу увагу приділити використанню методів аналізу одномірних часових рядів, зокрема, методів декомпозиції часових рядів та експоненційного згладжування, проте їхня відсутність не є суттєвим недоліком і не впливає на загальну якість розв'язання поставленої задачі.

6. Загальна оцінка кваліфікаційної магістерської роботи та її допущення до захисту перед ЕК: Зважаючи на те, що кваліфікаційна бакалаврська робота студента Миська Данііла Сергійовича на тему «Прогнозування розвитку електронної торгівлі в Україні» є актуальною, дослідження є самостійним і ґрунтовним, вважаю, що дана кваліфікаційна робота відповідає вимогам та може бути допущена до захисту перед ЕК і заслуговує позитивної оцінки.

Науковий керівник

Катуніна О.С., к.е.н, доцент, доцент кафедри математичного моделювання та статистики Київського національного економічного університету імені Вадима Гетьмана

« ___ » _____ 2023 р.

Катуніна О.С.

ЗМІСТ

ВСТУП	3
РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ ТА МОДЕЛЮВАННЯ РОЗВИТКУ ЕЛЕКТРОННОЇ ТОРГІВЛІ	6
1.1 Сутність, передумови виникнення та проблеми дослідження ринку електронної торгівлі.....	6
1.2 Аналіз тенденцій розвитку електронної торгівлі в економіці України.....	8
1.3 Структура та особливості систем електронної комерції	13
РОЗДІЛ 2. МАТЕМАТИЧНІ МОДЕЛІ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ РОЗВИТКУ ЕЛЕКТРОННОЇ ТОРГІВЛІ	26
2.1 Огляд напрямів використання математичних моделей у дослідженнях ринку електронної торгівлі	26
2.2 Інструментарій розвідувального аналізу та згладжування одномірних часових рядів факторів ринку електронної торгівлі.....	31
2.3 Напрями використання методів інтелектуального аналізу даних та машинного навчання для аналізу та прогнозування тенденцій розвитку електронної торгівлі.....	33
РОЗДІЛ 3. МОДЕЛЮВАННЯ ЕЛЕКТРОННОЇ ТОРГІВЛІ МЕТОДАМИ МАШИННОГО НАВЧАННЯ	39
3.1 Аналіз результатів візуалізації та препроцесингу даних про стан ринку електронної комерції в Україні.....	39
3.2 Аналіз результатів моделювання тенденцій та системи факторів розвитку електронної комерції на базі методів інтелектуального аналізу даних	43
3.3 Аналіз результатів моделювання динаміки розвитку електронної комерції..	50
ВИСНОВКИ	59
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	61
ДОДАТКИ	65

ВСТУП

Актуальність теми дослідження. В епоху розвитку цифрових технологій впровадження систем електронної торгівлі має тенденцію до стрімкого зростання та збільшення впливу на сфери суспільного життя. Глобалізація процесів електронної торгівлі та визначення її перспективного розвитку у майбутньому створює відповідні умови, що обумовлюють необхідність здійснення аналізу сучасних форм ведення електронної комерції.

Аналіз стану електронної комерції в економіці України довів, що зростання впливу цифрових технологій на ведення бізнесу залежить від стрімкого розвитку технологічних можливостей через вплив всесвітньої мережі Інтернет. Електронна комерція стрімко набирає популярність, а завдяки розгалуженій мережі Інтернету з'являються зручні форми доступу покупців до онлайн-ринків. Нові форми торгівлі також стають вигідними для постачальників, дозволяючи їм легше входити на ринок і розширювати номенклатуру своєї продукції. В результаті електронна форма ведення комерції сприяє зростанню обсягів товарів та послуг, які замовляються і доставляються дистанційно, що в результаті сприяє максимальному задоволенню потреб споживачів.

З розвитком нових онлайн-ринків, продуктів і бізнес-моделей, пов'язаних з електронною комерцією, відбуваються значні зміни в способах ведення бізнесу та економічної діяльності. Впровадження сучасних інформаційних технологій мережевої економіки істотно змінює традиційні підходи до торгівлі та сприяє появі нових мережевих платформ, а також розгалуженню форматів комунікації учасників ринку. Набувають пріоритетного значення розвиток методичних підходів та інструментарію функціонування систем електронної торгівлі, в тому числі математичні моделі та методи аналізу, оцінювання, прогнозування попиту та обсягів реалізації продукції.

Аналіз останніх досліджень та публікацій. У науковій літературі приділяється значна увага теоретичним та практичним аспектам функціонування

електронного бізнесу та електронної комерції, зокрема, питанням моделювання попиту, обсягів реалізації продукції та дослідженню механізмів впливу попитотвірних факторів. Серед робіт зарубіжних учених та дослідників із зазначеної тематики найбільш відомими є публікації Д. Белла, Л. Герстнера, Гр. Дункана, М. Кастельса, Л. Клімченю, О. Кобелева, Д. Козьє, А. Саммера, Л. Стюарта, Яна Тіме, В. Тріза, Р. Фроста та інших. Водночас питання розроблення та впровадження математичного та модельного інструментарію розбудови систем електронної торгівлі потребують подальшого дослідження і саме вони зумовили вибір теми даної кваліфікаційної роботи.

Мета і завдання дослідження. Метою виконання даної роботи є побудова та експериментальне використання моделей прогнозування розвитку електронної торгівлі в Україні за допомогою методів інтелектуального аналізу даних та математичного моделювання.

Досягнення поставленої мети обумовлює необхідність вирішення таких завдань:

1. Проаналізувати структуру та тенденції ринку електронної торгівлі в Україні.
2. Вивчити та узагальнити досвід застосування інструментарію математичного моделювання для аналізу процесів формування ринку електронної торгівлі.
3. Обґрунтувати вибір алгоритмів інтелектуального аналізу даних для аналізу та прогнозування розвитку електронної комерції.
4. Побудувати систему моделей прогнозування обсягів електронної торгівлі на базі інструментів аналізу часових рядів.

Об'єкт дослідження. Об'єктом дослідження є електронна торгівля як складова інформаційно-мережевої економіки.

Предмет дослідження. Предметом дослідження є математичні моделі та методи аналізу та прогнозування обсягів електронної торгівлі в економіці країни під впливом низки факторів економічного середовища.

Методи дослідження. Для вирішення поставлених у роботі завдань використано низку загальнонаукових методів, а саме: описово-аналітичний метод; метод системного й просторового аналізу; метод логічного узагальнення; методи діалектики: індукція й дедукція; метод аналізу й синтезу. Для дослідження обсягів електронної торгівлі застосовано динамічне та статичне моделювання, використані графічні і статистичні методи обробки інформації, методи перевірки стаціонарності часових рядів, ARIMA моделювання одномірних часових рядів, побудовано моделі множинної лінійної регресії для аналізу та прогнозування залежності частки електронної торгівлі у ВВП від низки факторів економічного середовища.

Теоретична, методична та практична значущість отриманих результатів. Кваліфікаційна робота має теоретичну значущість, оскільки вона сприяє розширенню знань про використання інструментів математичного моделювання для аналізу розвитку електронної торгівлі. Практична значущість роботи полягає у розробці інструментів прогнозування обсягових показників електронної торгівлі країни з урахуванням впливу низки факторів економічного середовища, що допоможе різним зацікавленим сторонам при прийнятті обґрунтованих рішень.

Інформаційна база дослідження. Інформаційною базою дослідження послужили законодавство України, праці вітчизняних та зарубіжних науковців та практиків, наукові статті, статистичні дані офіційних сайтів та спеціалізованих інтернет-джерел, науково-методична література.

Структура роботи. Робота складається зі вступу, трьох розділів, списку використаних джерел та додатків. Загальний обсяг роботи становить 68 сторін.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ ТА МОДЕЛЮВАННЯ РОЗВИТКУ ЕЛЕКТРОННОЇ ТОРГІВЛІ

1.1 Сутність, передумови виникнення та проблеми дослідження ринку електронної торгівлі

Електронна комерція або електронна торгівля — це купівля та продаж товарів і послуг через електронну мережу, насамперед через Інтернет. Активний розвиток Інтернету сприяє формуванню мережі та побудови різноманітних платформ, які істотно впливають на сферу торгівлі. Тому Інтернет став не лише місцем обміну інформацією, а й майданчиком для купівлі та продажу матеріальних благ і послуг. Світи електронної комерції та традиційної роздрібною торгівлі, які поширені в суспільстві, здійснюють зливання, і стають багатоканальними [13].

Всесвітня мережа Інтернет як канал збуту розширює ринок, пропонує споживачеві більше інформації щодо пропонуємої продукції та можливість порівняти різні властивості та відповідну ціну товару. З огляду на стрімкий розвиток веб- та інтернет-технологій, електронна комерція завойовує всі куточки сфер діяльності суспільства і перетворюється на самостійну галузь економіки. Середовище електронної комерції стимулює модернізацію бізнес-процесів (виробництво, збут, маркетинг тощо), а також торгово-економічні відносини [15].

Сучасна форма пропозицій товарів, послуг та безпосередньо сама електронна комерція вже стала невід'ємною частиною життя багатьох людей. Україна не є винятком у цій глибокій трансформації, її комерційні відносини активно переходять в онлайн-світ. За останні шість років ринок електронної комерції в Україні спостерігав щорічні темпи зростання близько 24,7%, навіть перевищивши зростання світового ринку в 21,5%. У 2019 році загальний обсяг ринку в Україні досяг 76 млрд грн, або близько 7% від загального обсягу роздрібних продажів. В Україні поява і розвиток електронної комерції відбулися

дещо пізніше, ніж в розвинутих країнах Європи [19].

Згадані тенденції є можливістю розширення новим каналом для вже віддалених офлайн-роздрібних компаній, перспективу раннього виходу на ринок для нових гравців, а також сприятливе бізнес-середовище для тих, хто вже захопив значну частку ринку в українському електронному ринку. Можливості супроводжуються більш інтенсивним конкурентним середовищем.

Глибоке розуміння ключових факторів, які спонукають споживачів робити покупки в Інтернеті, стає все більш важливим для гравців у середовищі електронної комерції. Успіх компанії багато в чому залежить від її здатності підтримувати міцні стосунки зі своїми клієнтами. Успіх ритейлерів в епоху цифрових технологій вимагає розвитку глибокого розуміння та усвідомлення портрета клієнта. Тому комплексне обґрунтування рішень про купівлю та переваг споживачів стає все більш важливим у процесі розробки та вдосконалення корпоративних маркетингових стратегій [13].

Переконати споживача робити покупки онлайн – це складне завдання. На це рішення впливає кілька факторів, зокрема психологічні, соціальні та демографічні. Швидке щорічне зростання роздрібних продажів електронної комерції сприяло появі нового покоління споживачів зі специфічними потребами, уподобаннями та цінностями. У електронної комерції існує п'ять різних моделей доходу, які використовуються компаніями для заробітку.

Ось ці моделі:

- Модель продажу: Найпоширеніша модель, в якій компанії продають товари або послуги, отримуючи прибуток.
- Модель підписки: Компанії пропонують продукти або послуги на основі підписки, отримуючи стабільний дохід від постійних платежів.
- Рекламна модель: Онлайн-творці та впливові особистості розвивають свої бренди, укладаючи рекламні угоди з іншими компаніями.
- Партнерська модель: Творці з великою аудиторією можуть співпрацювати з іншими компаніями, продавати їх продукти через свої платформи та отримувати комісійні від кожної угоди.

- Модель комісії за транзакцію: Компанії, які обробляють фінансові операції, отримують дохід у вигляді комісійних з кожної здійсненої транзакції [13].

1.2 Аналіз тенденцій розвитку електронної торгівлі в економіці України

Роздрібний сегмент електронної комерції стрімко розвивається у всьому світі. За даними Statista, його обсяг у 2019 році склав \$3,5 трлн, або 14% від загального обсягу роздрібних продажів. Експерти прогнозують, що у 2023 році частка електронної комерції досягне 21,8%, що становить понад 6,5 трильйонів доларів. Загалом, за останні шість років світовий ринок електронної комерції зріс на CAGR на 21,5% [4].

На основі сторін економічних відносин та їх взаємодії між ними електронна комерція в даний час охоплює всі чотири основні сегменти ринку: бізнес-бізнес (B2B), бізнес-споживач (B2C), споживач-споживач (C2C) і від споживача до бізнесу (C2B).

Враховуючи участь України в процесах глобалізації, електронна комерція стала однією з найперспективніших ніш бізнесу в Україні. За оцінками однієї з найбільших українських продуктових ІТ-компаній EVO, обіг фізичних товарів і послуг, придбаних онлайн в Україні, у 2019 році склав 76 млрд грн, або 6,9% від загального обсягу роздрібних продажів (рис 1.1). Однак, якщо порівняти ці обсяги зі світовими, стає зрозуміло, що український ринок електронної комерції все ще знаходиться в зародковому стані. Оскільки ринок ще не зрілий, він розвивається надзвичайно швидко. Примітно, що протягом останніх шести років компанія досягла надзвичайного зростання, що підтверджується сукупним річним темпом зростання (CAGR) у 24,7%. При цьому сформований сегмент роздрібної торгівлі в Україні зберіг темп зростання на рівні 3,9% [20].

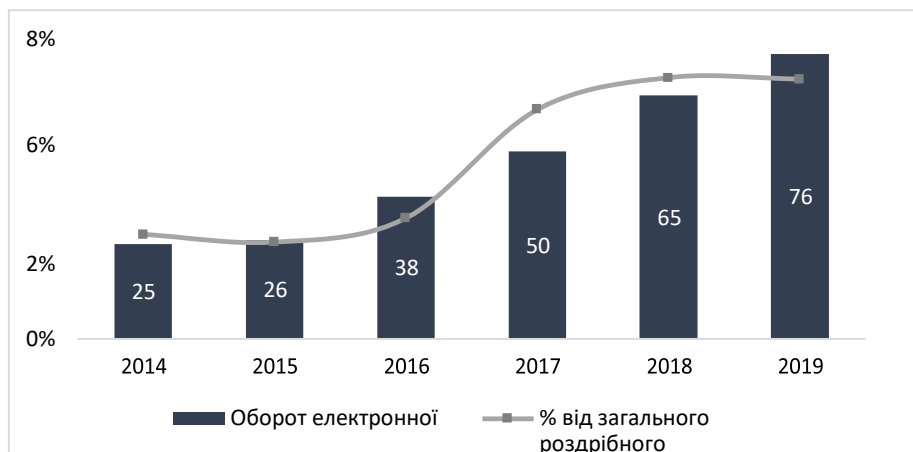


Рисунок 1.1 – Український роздрібний e-commerce (2014-2019), млрд грн[20]

Слід зазначити, що в більшості країн європейського простору, особливо в економічно розвинутих країнах, частка розрахунків за придбані в електронному вигляді товари та послуги, ступінь використання банківських карток і різноманітних платіжних систем у загальному обсязі платежів є досить високою. В основному це пов'язано з високим рівнем правової захищеності таких операцій, інформаційною культурою компанії та зручністю використання таких інструментів розрахунків. Усе це також сприяє підвищенню привабливості електронної комерції не лише для нових торгових компаній, а й для компаній у сфері послуг, особливо у фінансовому секторі [21].

Незважаючи на величезне зростання в річному обчисленні, яке навіть перевищує загальне зростання світового ринку, українське онлайн-середовище все ще має багато обмежень, скорочення яких може відкрити ще більший потенціал для розширення. Рівень проникнення інтернету в країні сягнув 67% у 2019 році, тоді як середній показник для Східної Європи на той час становив 71%.

Ісландія, наприклад, майже досягла повного охоплення здійснення електронних послуг населення (99% у 2018 році). У той же час кількість користувачів Інтернету в Україні стрімко зростає в порівнянні з минулим роком, про що свідчить річний показник зростання на 5,7% між 2019 і 2020 роками. Якщо

дивитися глибше, то частка споживачів, які користуються Інтернетом і робили покупки онлайн є порівняно невеликою. Лише 21% населення України робить покупки онлайн, тоді як серед інтернет-користувачів 34% (52% у Європі) роблять покупки онлайн. Щоб підкреслити контраст, 88% користувачів Інтернету в Швейцарії роблять покупки онлайн.

Пояснення цього може полягати в тому, що сегментація ринку України не настільки розвинена та багатостороння. Середні витрати на покупки в Інтернеті в Європі становлять 276 євро, тоді як в Україні люди витрачають в середньому 200-1000 грн [19].

За статистикою EVO Group, яка включає Prom, Bigl та інші проекти, обсяг ринку електронної комерції в Україні у 2019 році зрос на 17% і склав 76 млрд грн. Незважаючи на це, відмічається зменшення середнього чеку на 7-10%, а вартість доставки зросла на 25-30% і становить від 5 до 15% від вартості замовлення.

За прогнозами експертів EVO Group, у 2020 році очікується менше зростання - приблизно 15% (всього 87,2 млрд грн). Онлайн-торгівля продовжує розвиватися швидше, ніж традиційний роздрібний ринок в Україні. Це призводить до того, що більшість продавців інвестують у швидкий і глобальний розвиток онлайн-каналів, не тільки в сфері побутової техніки, одягу та взуття, але й в продукти харчування. Це стимулює розвиток ринку доставки готових товарів, продуктів і страв [19].

Важливим аспектом розвитку електронної комерції в Україні є високий потенціал для зростання. Навіть при низькому рівні доходів населення та девальвації національної валюти, ринок продовжує стабільно зростати. Серед найбільших операторів електронної комерції можна виділити Prom.ua, Citrus, Rozetka, Tickets.ua, modnaKasta, LeBoutique, MoYo, Allo та інших. Найкращі сайти електронної комерції в Україні включають [20]:

- OLX, який займає до 40% всіх онлайн-замовлень і є рекламним майданчиком;
- Інтернет-супермаркет Rozetka (32,5%);
- Prom - Marketplace, що об'єднує багато інтернет-магазинів (26,5%).

36% українських споживачів вважають оплату покупок банківською карткою на сайті найкращим варіантом, тоді як 23% віддають перевагу готівці. 14% респондентів використовують банківські перекази або оплату при отриманні на пошті. Проте існують кілька перешкод для оплати покупок карткою на сайті електронної комерції в Україні, зокрема:

- 44% українських респондентів не бажають робити передоплату.
- 26% відзначають, що на деяких сайтах відсутній спосіб оплати карткою.
- 10% мають уявлення про небезпеку онлайн-платежів.
- 9% стикаються з високою вартістю покупок в інтернет-магазинах.

За дослідженням сектору електроніки в Україні, найпопулярнішими онлайн-покупками серед українських споживачів є парфумерія, косметика, взуття, аксесуари, електроніка, товари повсякденного попиту, побутова хімія, товари для дому та саду. Зокрема, 64% інтернет-користувачів купують взуття та одяг, а 6% виявляють бажання, але не можуть зробити такі покупки. Близько 11% українських респондентів здійснюють покупки онлайн з доставкою додому, а також бажають зробити такі покупки, але стикаються з обмеженою інформацією або неможливістю скористатися такими послугами [19].

Близько 29% українців роблять покупки в Інтернеті або оплачують рахунки таким способом. Електронна комерція серед українців розвивається, але залишається невеликою відсотковою часткою. У 2017 році близько 13,6 мільйонів людей відвідали сайт електронної комерції принаймні один раз, тоді як середній споживач витратив 334 долари США на покупки в Інтернеті.

Найбільшим сегментом української електронної комерції є електроніка та медіа. Він генерує 28% доходу від електронної комерції в Україні. Далі йде мода з 26%. За ним йдуть меблі та техніка з 20% доходу від електронної комерції. Останні два: іграшки, хобі та DIY з 15% і їжа та засоби особистої гігієни з 10% [21].

Зауважується, що в Україні спостерігається слабкий розвиток інфраструктури електронної комерції, включаючи низьке проникнення Інтернету та обмеженість інструментів електронних платежів. Крім того, Україна має

найнижчий ВВП на душу населення серед країн Східної Європи, що призводить до низької купівельної спроможності громадян і обмеженого рівня споживчого попиту, включаючи сегмент електронної комерції. У таблиці 1.1 показано форми електронного бізнесу.

Таблиця 1.1 – Форми електронного бізнесу

	Форма	Характеристика
Традиційні форми	Бізнес – споживач (B2C/C2B – Business-to- Consumer/Consumer- to-business)	Форма взаємодії між компаніями та споживачами, що дозволяє виробникам ефективно рекламувати свої товари та послуги безпосередньо споживачам, надаючи останнім більше можливостей для вибору. Цей метод є однією з ранніх стратегій електронного бізнесу, яка передбачає прямий шлях від виробників до кінцевих споживачів, обходячи посередників.
	Бізнес – бізнес (B2B –Business-to- Business)	Форма оперативної співпраці між підприємствами на різних рівнях, що базується на вже встановлених бізнес-процесах та затверджених схемах постачання та збуту, забезпечує сталу внутрішню організацію. Ця форма дозволяє об'єднати внутрішні мережі партнерів для спільного електронного обміну документами, створює систему прямого розміщення замовлень з можливістю деталізації їх параметрів та відстеження їх виконання в режимі реального часу, що сприяє щільнішій комунікації між підприємствами. Цей підхід є надзвичайно популярним у сучасному бізнесі.
	Споживач – споживач (C2C –Consumer-to- Consumer)	Форма електронного спілкування, де приватні особи здійснюють купівлю, продаж, обмін товарів, послуг та інформації в онлайн-середовищі. Наприклад, дошки оголошень та Інтернет-аукціони є прикладами такої форми взаємодії.
Альтернативні форми	Бізнес – бізнес (B2B –Business-to- Business)	Взаємодія між компаніями за допомогою електронних засобів на всіх рівнях дозволяє ефективно встановлювати двосторонні та багатосторонні економічні зв'язки. Це означає, що компанії можуть самостійно реалізовувати свою продукцію, розміщувати замовлення та моніторити їх виконання шляхом використання баз даних постачальників та інших інструментів.
	Бізнес – споживач (B2C –Business-to- Consumer)	Форма електронної комерції, яка спрямована на кінцевого споживача. Одним з найтипівідоміших прикладів такого підходу є електронна роздрібна торгівля в Інтернеті, яка прямо взаємодіє зі споживачами.

Бізнес – адміністрація (B2A –Business-to-Administration)	Взаємодія між компаніями та адміністративними органами, включає в себе комерційні зв'язки між комерційними структурами та державними та міжнародними організаціями. Цей сегмент електронної комерції є ще недостатньо розвинутим, але має потенціал для подальшого росту, за умови, що уряди країн активніше використовуватимуть свої можливості для підтримки компаній.
Споживач – адміністрація (C2A – Consumer-to-Administration)	Взаємодія між споживачів та адміністрацією.

Джерело: . розроблено автором на основі [20]

Крім того, низький рівень цифрової культури серед населення та недовіра до онлайн-платежів ускладнюють розвиток електронної комерції в Україні. Проте, варто відзначити значний потенціал розвитку ринку, адже країна має високі темпи зростання.

Однією з основних перешкод для розвитку електронної комерції в Україні є недосконалість інфраструктури та правового регулювання ринку. Проте, розвиток електронної комерції може стати потужним двигуном економічного зростання в країні, завдяки впровадженню новітніх технологій, оптимізації бізнес-процесів та інтеграції торгових відносин України в глобальну інформаційну економіку. Крім того, електронна комерція відкриває нові можливості для транскордонних транзакцій через цифрові торгові платформи, що дозволяє залучати користувачів з усього світу [23].

1.3 Структура та особливості систем електронної комерції

У сучасному бізнесі відбуваються значні зміни, пов'язані зі зростанням можливостей компаній та підвищенням конкуренції. Це призводить до поліпшення якості товарів та послуг і вимагає нових підходів до організації та управління бізнесом. Багато компаній впроваджують різноманітні комп'ютерні системи та мережі для автоматизації своїх процесів і досягнення успіху. [15].

У міру проникнення сучасних інформаційних і комунікаційних технологій у діяльність різноманітних суспільних інститутів і виробництв діапазон комерційних відносин на основі активного використання Інтернет істотно розширюється. Вже сьогодні електронна комерція стає частиною повсякденного життя, а в самому найближчому майбутньому така форма організації ділових відносин стане звичайною.

У зв'язку з цим у сучасній економічній науці відбувається поступова трансформація понять, що спричиняє появу різних визначень термінів «електронна комерція», «електронний бізнес».

У більшості випадків поняття електронного бізнесу трактується як загальне поняття бізнес-активності в Інтернеті. У широкому розумінні, електронний бізнес виходить за межі електронної комерції. Він охоплює всі бізнес-активності, які використовують можливості глобальних інформаційних мереж для трансформації внутрішніх та зовнішніх зв'язків компанії з метою отримання прибутку. [4].

Відповідно до існуючих стандартів ООН, бізнес визнається електронним, якщо хоча б дві з чотирьох його складових (виробництво товару чи послуги, організація попиту, доставка та фінансові розрахунки) здійснюються за допомогою мережі Інтернет.

Електронна комерція охоплює різноманітні моделі бізнесу, включаючи бізнес-бізнес (B2B), бізнес-споживач (B2C), споживач-споживач (C2C) та споживач для бізнесу (C2B). Крім того, існують різні інтернет-послуги, які надаються в мережі, такі як фінансові послуги (трейдинг, банкінг, страхування), інформаційні послуги (доступ до пошукових систем, електронні каталоги, електронна пошта, інтернет-телефонія) та послуги в різних галузях, таких як медицина, транспорт, туризм, освіта та консультування. Крім того, електронна сфера включає Інтернет-сервіси, такі як мережеві ЗМІ, електронні версії газет, журналів, радіо- та телестанцій, а також діяльність електронних фірм, які займаються розробкою програмного забезпечення, системною інтеграцією, просуванням та супроводженням веб-сайтів, а також хостингом та реєстрацією доменних імен. [23].

Традиційна комерція, або ведення бізнесу, являє собою обмін товарами або послугами між, принаймні, двома сторонами і включає всі види діяльності, які кожна зі сторін використовує для здійснення угоди. Тобто комерція – це торгівля.

Тоді, придбання товару відноситься до електронної комерції, якщо, як мінімум, маркетинг (організація попиту) та фінансові розрахунки між виробником продукції та її споживачем здійснюються за допомогою мережі Інтернет.

Електронна комерція дозволяє постачальникам і замовникам рівною мірою використовувати нові можливості (табл.1.2.).

Таблиця 1.2 – Можливості та переваги електронної комерції

Можливості постачальників	Можливості споживачів	Короткий опис переваг
Глобальна присутність	Глобальний вибір	Завдяки глобальній природі Інтернету, електронна комерція дозволяє навіть найдрібнішим постачальникам досягати глобальної присутності і займатися бізнесом у всесвітньому масштабі. Споживачі також отримують можливість глобального вибору з усіх потенційних постачальників, що пропонують необхідні товари або послуги, незалежно від їх географічного розташування.
Підвищення конкурентоспроможності	Отримання високоякісних послуг	Постачальники за допомогою електронної комерції можуть підвищити свою конкурентоспроможність, наближаючись "ближче до замовника". Багато компаній використовують технології електронної комерції, щоб надати розширену підтримку перед та після продажу. В результаті споживач отримує покращену якість обслуговування.
Повна інформованість про потреби споживачів	Персоналізація товарів і послуг	В наслідок електронній взаємодії, компанії можуть отримувати детальну інформацію про запити кожного окремого споживача та автоматично надавати продукти і послуги відповідно до їх вимог. Простим прикладом може бути електронний журнал, який налаштовується під конкретного читача, надаючи йому статті, які цікаві саме йому при кожному доступі, а також видаляючи вже прочитані матеріали.
Скорочення шляху товару до споживача	Швидка реакція на попит	Електронна комерція сприяє значній оптимізації товарних потоків, оскільки товари доставляються безпосередньо від виробника до споживача, уникаючи традиційних проміжних точок, таких як оптові та роздрібні торгові підприємства.

Зниження витрат	Зниження цін	В електронному оформленні замовлень витрати на обслуговування операцій значно знижуються, що в свою чергу призводить до зниження цін для споживачів.
Нові бізнес-моделі	Нові продукти та послуги	Нові моделі бізнесу, такі як віртуальні підприємства, віртуальні агенти, технології, аутсорсинг і телеробота, дуже ефективно покращують комерційну діяльність. Крім перетворення наявного ринку товарів і послуг, електронна комерція створює можливість для з'явлення абсолютно нових продуктів і послуг. Наприклад, вона відкриває двері для віддалених банківських, страхових, брокерських послуг, а також служб електронного постачання та підтримки.

Джерело: розроблено автором на основі [20]

На сьогоднішній день під електронною комерцією (e-commerce) розуміють здійснення будь-яких форм ділових угод за допомогою інформаційних мереж, до яких належать:

- надання інформації про товари та послуги в режимі онлайн;
- оформлення замовлень та представлення до оплати рахунків в режимі онлайн;
- оплата та обробка транзакцій в режимі онлайн.

Таким чином, електронна комерція займається лише питаннями здійснення продажів, а електронний бізнес – організацією бізнес-процесів в мережі Інтернет в цілому. Електронна комерція є одним з перспективних напрямків сучасного бізнесу, що складається з п'яти взаємозалежних сфер комерційної діяльності компаній:

1) реклама – зосередження динамічними та інтерактивними Інтернет - засобами уваги користувачів на комерційному сайті фірми. Вимоги до якості Інтернет-реклами, стилю її подання стимулювали розробку нових інформаційних технологій та засобів програмування, таких як Java-анімація та динамічний HTML (DHTML) [23];

2) представлення товару – передбачає мультимедійний показ продукції засобами мережі Інтернет. Для показу товарів використовуються відеоролики, різноманітні звукові ефекти, тривимірні зображення та анімаційний супровід;

3) виконання операцій – складається з вибору бажаних товарів за допомогою вітрини, заповнення електронної кошика, роботи з електронною платіжною системою, а також із супроводу виконання замовлень;

4) післяпродажна підтримка – передбачає гарантійне обслуговування продукції;

5) побудова відносин – включає проведення комплексу системних заходів, які дозволяють підтримувати комерційні відносини між споживачами та фірмою. Прикладом такої діяльності може бути конвеєрна розсилка інформації про нові товари, акції тощо. В якості засобів зворотного зв'язку з клієнтами можуть використовуватись форуми, або телеконференції [5].

Розглянемо бізнес-процеси, що придатні для:

– електронної комерції: продаж книг і CD, Online-доставка програмного забезпечення, реклама транспортних послуг, інформація про товар у дорозі [22];

– традиційної комерції: продаж модного одягу, продаж продуктів, що швидко псуються, продаж дрібниць, продаж ювелірних виробів та антикваріату;

– комбінування електронної та традиційної комерції: продаж автомобілів, банківські послуги, капіталовкладення, страхові послуги.

Щоб здійснювати, або отримувати оплату через мережу Інтернет, електронний магазин сам повинен бути підключений до однієї, або декількох платіжних Інтернет-системам.

Як зазначалось раніше, на Україні при здійсненні електронних платежів використовуються системи на основі банківських платіжних карт, та системи Інтернет-банкінгу.

Системи на основі банківських платіжних карт. При підключенні до певної платіжної системи, яка функціонує на основі банківських платіжних карт, електронний магазин зможе отримувати платежі за допомогою мережі Інтернет від тих своїх клієнтів, які мають карти, що підтримуються даною платіжною системою [24].

Наприклад, платіжна система «Портмоне» (адреса сайту – www.portmone.com.ua) – підтримує платежі кредитними картами Visa та

MasterCard таких банків, як: «ПриватБанк», «Monobank» та інших. Дана платіжна система найбільше орієнтована на оплату комунальних послуг за допомогою мережі Інтернет.

За типом обслуговування особового рахунку банківські платіжні картки поділяються на дебетові та кредитні.

Дебетові платіжні картки – це картки, які використовуються для оплати товарів та послуг, отримання готівки в банкоматах. Такі картки дозволяють розпоряджатись грошовими коштами лише в межах доступного залишку на рахунку. Функцією дебетових карт є заміна паперових грошей в обігу та здійсненні безготівкових платежів власними коштами клієнтів [30].

Кредитні платіжні картки – це картки, які на відміну від дебетових, дозволяють їх власникам здійснювати розрахунок за рахунок коштів, наданих банком в межах встановленого ліміту.

Картки міжнародних платіжних систем. Деякі види банківських карток використовуються в міжнародних системах платежів. До найбільш популярних міжнародних платіжних систем відносять: Visa (Electron, Classic, Gold, Platinum) та Mastercard (Cirrus, Maestro, Standard, Gold, Platinum) [24].

Найбільш розповсюдженими в світі є картки Visa Electron, та Mastercard Cirrus, або Maestro. У більшості випадків вони є дебетовими та, як правило, не дозволяють здійснювати електронні платежі за допомогою мережі Інтернет. Також вони мають мінімальний платіжний ліміт на операції, не містять рельєфних елементів на поверхні, що обмежує можливість використання цього виду карток в терміналах, заснованих на знятті відбитків з них. Це підвищує безпеку їх використання, такі картки є найбільш дешевими за вартістю їхнього випуску та обслуговування.

Також популярними та зручними у користуванні є картки Visa Classic та Mastercard Standard. Ці картки мають стандартний набір функцій, до яких входять: підтримка платежів в більшості торгівельних точок, бронювання товарів та послуг тощо. Вони бувають як дебетові, так і кредитні, а також дозволяють своїм власникам розраховуватись за допомогою мережі Інтернет [25].

Gold або Platinum картки є ексклюзивними картками, які надаються з метою підкреслити високий статус власників та слугувати символом престижу. Вони забезпечують додаткові гарантії щодо платоспроможності власника і мають більш високі платіжні та кредитні ліміти. Ці картки також використовуються для підвищення іміджу свого власника.

Технічно, картки Visa Classic, Visa Gold, MasterCard Standard та MasterCard Gold нічим, окрім дизайну, вартості випуску та обслуговування не відрізняються. Власникам карток Gold та Platinum деякі магазини пропонують знижки та інші переваги, наприклад, можливість користування в аеропорту залом бізнес-класу незалежно від категорії квитка на літак.

На поширене запитання клієнтів, «яку ж картку краще обрати MasterCard, або Visa, які в них відмінності та особливості?» співробітник банку небезпідставно може відповісти, що ці платіжні системи практично ідентичні і не мають суттєвої різниці. Але, на практиці іноді зустрічаються банкомати, що приймають картки тільки однієї з цих платіжних систем [25].

Також слід мати на увазі, що на сьогоднішній день для оплати товарів та послуг на території Європи більш актуальним є використання карток платіжної системи Mastercard, оскільки при використанні Visa відбувається подвійний перерахунок курсу валюти: гривні в долари США, а потім долари США в євро. Це збільшує накладні витрати на здійснення платіжних операцій, в порівнянні з Mastercard, де сума платежу відразу конвертується з валюти рахунку картки в євро.

Системи Інтернет-банкінгу призначені для забезпечення клієнтів послугами з керування банківськими рахунками через Інтернет. Ці системи дозволяють отримувати виписки про здійснені операції. Важливо зауважити, що для використання систем Інтернет-банкінгу магазину не обов'язково бути підключеним до цих систем. Для здійснення безготівкових платежів достатньо, щоб магазин приймав такі види розрахунків. [26].

Для того, щоб мати можливість використовувати послуги Інтернет-банкінгу, клієнт повинен укласти відповідний договір з банком. Після цього він отримує доступ до свого рахунку та може здійснювати платежі за допомогою

мережі Інтернет, використовуючи звичайний браузер та передбачені системою Інтернет-банкінгу стандартні процедури. [26]

Як правило, послуги Інтернет-банкінгу включають:

- Виписки банківських рахунків;
- Надання інформації про банківські продукти, наприклад, депозити і кредити;
- Прийом заявок клієнтів на відкриття депозитів, отримання кредитів і банківських платіжних карток;
- Внутрішні перекази на рахунки в банку за бажанням клієнтів;
- Перекази на рахунки в інших банках;
- Конвертація коштів;
- Оплата послуг та інше.

В якості прикладів українських платіжних систем, які підтримують послуги Інтернет-банкінгу, можна навести наступні:

- «Приват-24», ПриватБанк
- «Monobank», Monobank

Більш детально користування послугами інтернет-банкінгу на прикладі системи «Приват-24» ми будемо розглядати пізніше.

На Україні при здійсненні платежів за допомогою мережі Інтернет використовуються наступні платіжні системи, що засновані на принципах «електронної готівки»: системи на основі смарт-карт та системи на основі електронних грошей. Розглянемо їх більш детально [26].

Системи на основі смарт-карт. Для роботи з системами на основі смарт-карт, клієнтам необхідно мати саму смарт-карту, емітовану банком, яка підключена до його платіжної системи. Для здійснення платежів необхідно використовувати спеціальний пристрій – смарт-картрідер та відповідне програмне забезпечення.

Прикладом української платіжної системи на основі смарт-карт є система «Інтерплат».

Банк-емітент, також відомий як банк покупця, це банк, який видав (емітував) картку клієнта. Він відкриває рахунок клієнта, пов'язаний з картою,

здійснює авторизацію платежів, щоб підтвердити наявність коштів на рахунку, і здійснює переказ коштів з карткового рахунку клієнта на рахунок банку-еквайєра.

Банк-еквайєр, також відомий як банк продавця або розрахунковий банк, це банк, пов'язаний з електронним магазином. Він забезпечує розрахунки з магазином за операціями, що використовують кредитні або смарт-карти. Банк-еквайєр відкриває рахунок продавця (Merchant Account), який підтверджує згоду банку на приймання платежів за товари з використанням пластикових карток. [27].

Системи на основі електронних грошей. Для того, щоб використовувати ці системи для платежів за допомогою мережі Інтернет, необхідно підключитись до будь-якої з них.

Для цього, користувач повинен встановити на своєму комп'ютері клієнтське програмне забезпечення, що необхідне для роботи з такою системою; відкрити за допомогою нього рахунок та поповнити його, наприклад, з передплатеної картки.

На сайтах вказаних платіжних систем є перелік магазинів та компаній, що підключені до них. Ці магазини можуть отримувати оплату за товари та послуги електронними грошима.

Електронні гроші мають свої переваги та недоліки. Вони особливо корисні та зручні при здійсненні масових платежів невеликими сумами. Наприклад, використовуючи електронні гроші, легко розраховуватись в метро, кінотеатрах, клубах, оплачувати комунальні послуги, штрафи, здійснювати покупки в інтернет-магазинах та багато іншого. Процес платежу електронними грошима є швидким, не виникає черг, не потрібно видавати здачу, гроші майже миттєво переходять від платника до одержувача [28].

Переваги електронних грошей включають:

- Відсутність потреби в отриманні здачі при здійсненні платежів.
- Немає обмежень на розмір суми платежу, не потрібно мати готівку відповідної величини, як у разі з готівкою.
- Емісія електронних грошей має дуже низьку вартість, оскільки не потрібно виготовляти монети або банкноти, використовувати метал, папір, фарби тощо.

- Фізичне перерахування грошей не потрібне, ця функція здійснюється платіжним інструментом.
- Забезпечення фізичної безпеки електронних грошей простіше, ніж у разі з готівкою.
- Момент платежу фіксується електронними системами, що знижує вплив людського фактора до мінімуму.
- Електронні гроші не потрібно рахувати, пакувати або перевозити, не потрібно організовувати спеціальні сховища.
- Електронні гроші не втрачають своїх фізичних властивостей з плином часу, вони не піддаються старінню.

Однак, використання електронних грошей також супроводжується значними недоліками:

- Відсутність стабільного правового регулювання, оскільки багато країн ще не визначили свою позицію щодо електронних грошей на законодавчому рівні.
- У разі фізичного пошкодження носіїв електронних грошей втрачається їхня грошова вартість, і відновлення стає неможливим.
- Криптографічні методи захисту, які застосовуються в платіжних системах електронних грошей, ще не мають довгострокової успішної історії експлуатації.
- Зацікавлені особи можуть намагатись відстежувати персональні дані платників та маніпулювати обігом електронних грошей поза банківською системою.
- Теоретично можливе розкрадання електронних грошей за допомогою новаторських методів, використовуючи недостатню зрілість технологій їхнього захисту.

Компанії-розробники протоколу Secure Electronic Transaction (SET) визначили відкриту специфікацію, яка стала неофіційним стандартом для електронної комерції. Ця специфікація включає захист інформації шляхом шифрування, аутентифікацію учасників транзакцій за допомогою цифрових

підписів та сертифікатів, а також взаємодію між різними виробниками за допомогою відкритих протоколів.

Протокол SET встановлює наступні вимоги до безпеки операцій електронної комерції:

- Забезпечення конфіденційності даних оплати та конфіденційності інформації про замовлення, що передається разом з даними оплати.
- Збереження цілісності даних платежів за допомогою цифрових підписів.
- Використання криптографії з відкритим ключем для проведення аутентифікації.
- Аутентифікація утримувача картки за його кредитною карткою за допомогою цифрових підписів та сертифікатів утримувача.
- Аутентифікація продавця та його здатність приймати платежі за допомогою цифрових підписів та сертифікатів продавця.
- Підтвердження того, що банк продавця є діючою організацією, яка може приймати платежі по картках, за допомогою цифрових підписів та сертифікатів банку продавця.
- Готовність до оплати транзакцій після успішної аутентифікації сертифікату з відкритим ключем для всіх сторін.
- Забезпечення безпеки передачі даних шляхом широкого використання криптографії та інших заходів.

Однією з основних переваг протоколу SET є використання цифрових сертифікатів стандарту X.509, версії 3. Ці сертифікати пов'язують утримувача картки, продавця та банк продавця з банківськими установами, які підтримують платіжні системи VISA та MasterCard. Це забезпечує додатковий рівень ідентифікації та довіри в електронному бізнесу[29].

Після вивчення різних підходів до визначення електронної комерції, ми встановили, що це широкий глобальний простір, де товари та послуги продаються та купуються за допомогою Інтернету та інших інформаційно-комунікаційних технологій. Оскільки немає однозначного міжнародного розуміння, ми вважаємо терміни "електронна комерція" та "електронна торгівля" синонімами.

В Україні електронна комерція поступово розвивається і набуває значення в економіці країни. Зростання кількості користувачів Інтернету та поширення мобільних пристроїв сприяють розвитку електронної комерції. Однак існують проблеми, такі як недостатній рівень довіри, недостатня розбудова інфраструктури та неефективне правове регулювання, які потребують уваги та вдосконалення.

На сьогоднішній день існують кілька передумов для розвитку та поширення електронної торгівлі. Серед них варіативність та швидкий розвиток світових та національних ринків, прогрес наукових досліджень та впровадження інноваційних технологій, розвиток інформаційного суспільства, зростання кількості користувачів Інтернету та поширення мобільних пристроїв, що впливають на споживчу поведінку та вподобання (табл. 1.3).

Таблиця 1.3 – SWOT-аналіз електронної комерції

	Сильні сторони <i>Можливості</i>	Слабкі сторони <i>Загрози</i>
Зовнішнє середовище	<ol style="list-style-type: none"> 1. Зростання кількості користувачів Інтернету підвищується. 2. Відбувається зміна свідомості споживачів. 3. Компанії електронної комерції завжди мають потенціал для розширення шляхом проникнення в нові сегменти ринку та дошарки населення. 4. Управління не обмежується часом і простором. 5. Бізнес у сфері електронної комерції демонструє високі темпи зростання. 6. Реклама в електронній комерції є більш економічно ефективною порівняно з традиційними системами. 	<ol style="list-style-type: none"> 1. Інтенсивна конкуренція на ринку. 2. Проблема забезпечення конфіденційності інформації, яка може бути розкрита на глобальному рівні. 3. Вплив макрофакторів, таких як зміни в політичному, економічному та культурному оточенні. 4. Проблема шахрайства.
	<i>Переваги</i>	<i>Недоліки</i>

<ol style="list-style-type: none"> 1. Розвиток електронної комерції не обмежений територіальними рамками та не потребує прив'язки до певного місця. 2. Онлайн купівля-продаж значно зменшує час та фізичні зусилля суб'єктів ринку. 3. Немає обмежень у часі для здійснення трансакцій. 4. Швидкий обмін інформацією між постачальниками та споживачами. 5. Гнучка сегментація ринку. 6. Висока доступність товарів та послуг. 	<ol style="list-style-type: none"> 1. Відсутність адекватного нормативно-правового забезпечення, яке б забезпечувало безпеку в електронній комерції. 2. Відсутність прямого фізичного контакту між споживачем і продавцем може обмежувати можливості отримання знижок. 3. Довгі строки доставки товарів. 4. Обмежена можливість реклами через відсутність доступу до масових пропагандистських інструментів в глобальній мережі. 5. Відсутність можливості впливу на вибір споживача, пов'язана з відсутністю прямого фізичного контакту.
--	--

Джерело: розроблено автором на основі [15]

Аналіз регулювання електронної торгівлі показав, що протягом останніх десятиліть було сформовано відповідну нормативну базу для правового регулювання на рівні окремих країн, а також в межах інтеграційних об'єднань та міжнародних організацій. Однак розвиток технологій та інструментів електронної торгівлі, а також впровадження нових бізнес-моделей ставлять перед національними та міжнародними регуляторами нові завдання, оскільки складніші трансакції вимагають подальшого розвитку та удосконалення заходів для регулювання електронної торгівлі.

РОЗДІЛ 2

МАТЕМАТИЧНІ МОДЕЛІ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ РОЗВИТКУ ЕЛЕКТРОННОЇ ТОРГІВЛІ

2.1 Огляд напрямів використання математичних моделей у дослідженнях ринку електронної торгівлі

Аналіз існуючих математичних моделей, методів і підходів для досліджень ринку електронної комерції є важливою складовою вивчення цієї сфери бізнесу. Дослідження ринку електронної комерції допомагають розуміти поведінку споживачів, тенденції ринку та прогнозувати його розвиток. У науковій літературі і практичних дослідженнях розроблено різноманітні математичні моделі та методи, які допомагають аналізувати ринок електронної комерції.

Одним з основних підходів є статистичний аналіз даних, що дозволяє виявляти залежності, тренди та кореляції між різними факторами. Наприклад, за допомогою методу регресійного аналізу можна встановити зв'язок між витратами на рекламу і обсягом продажів, показниками відвідуваності сайту та конверсією покупок. [9].

Окрім того, моделі попиту та прогнозування ринкових тенденцій є важливим інструментом для дослідження ринку електронної комерції. Ці моделі можуть використовувати такі методи, як аналіз часових рядів, методи експертних оцінок, системи штучного інтелекту тощо. Вони дозволяють прогнозувати попит на товари та послуги, передбачати зміни у споживчих пререференціях та поведінці покупців.

Крім того, аналітичні моделі та алгоритми машинного навчання використовуються для виявлення патернів і залежностей у великих обсягах даних, наприклад, для сегментації ринку та прогнозування споживчих тенденцій.

Усі ці математичні моделі та методи допомагають дослідникам електронної комерції здійснювати аналіз ринкових даних, розуміти споживачів та прогнозувати

розвиток ринку. Використання цих моделей є важливим інструментом для прийняття обґрунтованих рішень у сфері електронної комерції.

Динамічний підхід до моделювання ринку електронної комерції зосереджується на охопленні мінливого характеру ринку з часом. Він розглядає такі фактори, як поведінка споживачів, ринкові тенденції, технологічний прогрес і динаміка конкуренції. Цей підхід зазвичай передбачає аналіз часових рядів, методи прогнозування та імітаційні моделі. Ось деякі ключові особливості динамічного підходу:

Аналіз часових рядів: включає багато категорій або варіацій даних, аналітикам іноді доводиться створювати складні моделі. Однак аналітики не можуть врахувати всі відхилення, і вони не можуть узагальнити конкретну модель для кожного зразка. Занадто складні моделі або моделі, які намагаються зробити занадто багато речей, можуть призвести до відсутності відповідності. Відсутність відповідності або переобладнання моделей призводить до того, що ці моделі не розрізняють випадкову помилку та справжні зв'язки, залишаючи аналіз спотвореним і прогнози невірними.

Моделі аналізу часових рядів включають:

- Classification: визначає та призначає категорії даних.
- Curve fitting: малює дані вздовж кривої, щоб вивчити зв'язки змінних у даних.
- Descriptive analysis: визначає моделі в даних часових рядів, наприклад тренди, цикли або сезонні коливання.
- Explanative analysis: намагається зрозуміти дані та зв'язки в них, а також причинно-наслідкові зв'язки.
- Exploratory analysis: висвітлює основні характеристики даних часових рядів, зазвичай у візуальному форматі.
- Forecasting: прогнозує майбутні дані. Цей тип базується на історичних тенденціях. Він використовує історичні дані як модель для майбутніх даних, прогножуючи сценарії, які можуть статися в майбутніх точках графіка.
- Intervention analysis: вивчає, як подія може змінити дані.

- **Segmentation:** розбиває дані на сегменти, щоб показати базові властивості вихідної інформації [31].

Методи прогнозування: використовуючи статистичні методи та економетричні моделі, дослідники можуть прогнозувати майбутні ринкові умови, моделі попиту та показники ефективності. Ці прогнози допомагають у прийнятті рішень і стратегічному плануванні.

Три загальні типи:

Після постановки проблеми менеджером і прогнозистом, прогнозист має можливість вибрати підхід. В цьому контексті існують три основні типи методів: якісні методи, методи аналізу часових рядів та прогнозування, а також моделі причинно-наслідкових зв'язків.

Перший тип методів використовує якісні дані, такі як експертні оцінки та інформацію про особливі події, які можуть бути або не бути пов'язані з минулим.

Другий тип методів цілком зосереджений на аналізі шаблонів і змінах у шаблонах, полагаючись на історичні дані.

Третій тип методів використовує детальну і специфічну інформацію про взаємозв'язки між елементами системи і може формально враховувати спеціальні події. Він є потужним інструментом і, як і в методах аналізу часових рядів та прогнозування, минуле має важливе значення для причинно-наслідкових моделей.

Ці відмінності означають (цілком правильно), що один і той самий тип методики прогнозування не підходить для прогнозування продажів, скажімо, на всіх етапах життєвого циклу продукту — наприклад, методика, яка спирається на історичні дані, не буде корисною для прогнозування майбутнього абсолютно нового продукту, який не має історії [32].

Ось три найпоширеніші методи вивчення даних:

Моделі Box-Jenkins ARIMA: ці однофакторні моделі використовуються для кращого розуміння окремої змінної, що залежить від часу, наприклад температури в часі, і для прогнозування майбутніх точок даних змінних. Ці моделі працюють на основі припущення, що дані стаціонарні. Аналітики повинні врахувати та усунути якомога більше відмінностей і сезонності в минулих точках даних. На щастя,

модель ARIMA включає терміни для врахування ковзних середніх, оператори сезонної різниці та умови авторегресії в рамках моделі.

Модель ARIMA:

$$Y_t = a + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \phi_q \epsilon_{t-q}, \quad (2.1)$$

де $Y(t)$ - поточне значення часового ряду;

$\epsilon(t)$ - білий шум;

$\beta_1, \beta_2, \dots, \beta_p$ - коефіцієнти регресії;

$\phi_1, \phi_2, \dots, \phi_q$ - коефіцієнти авторегресії помилок.

Метод Холта-Вінтерса – це техніка експоненціального згладжування. Він призначений для прогнозування результатів за умови, що точки даних включають сезонність.

Формула метод Холта-Вінтерса:

Рівняння рівня:

$$L_t = \alpha * (Y_t - S_{t-m}) + (1 - \alpha) * (L_{t-1} + T_{t-1}), \quad (2.2)$$

Рівняння тенденції:

$$T_t = \beta * (L_t - L_{t-1}) + (1 - \beta) * T_{t-1}, \quad (2.3)$$

Сезонне рівняння:

$$S_t = \gamma * (Y_t - L_t) + (1 - \gamma) * S_{t-m}, \quad (2.4)$$

Рівняння прогнозу:

$$F(t + k) = L_t + k * T_t + S(t + k - m), \quad (2.5)$$

де $Y(t)$ представляє фактичне значення часового ряду в момент часу t ;

$L(t)$ представляє компонент рівня в момент часу t ;

$T(t)$ представляє складову тренду в момент часу t ;

$S(t)$ представляє сезонну складову в момент часу t ;

m означає тривалість сезонного циклу;

α, β та γ є параметрами згладжування, які контролюють ваги, надані поточному спостереженню, рівню та тенденції оновлення компонентів;

k представляє кількість періодів вперед для прогнозування [32].

Статичний підхід до моделювання ринку електронної комерції зосереджується на моментальному знімку ринку в певний момент часу. Він

забезпечує статичне бачення ринкових умов і не враховує тимчасові зміни. Ось деякі характеристики статичного підходу:

Перехресний аналіз: це метод, який досліджує зв'язок між двома чи більше змінними для виявлення закономірностей, асоціацій або залежностей. Це допомагає зрозуміти, як зміни однієї змінної впливають на іншу. Формула для розрахунку перехресного аналізу залежить від типу аналізу, який виконується, наприклад кореляційний аналіз, тест хі-квадрат або регресійний аналіз.

Формула для розрахунку коефіцієнта кореляції (r):

$$r = (\Sigma((X - \bar{X})(Y - \bar{Y}))) / (n * Sx * Sy), \quad (2.6)$$

де X і Y – окремі точки даних;

\bar{X} і \bar{Y} – середні значення;

X і Y , n – кількість точок даних;

Sx і Sy – стандартні відхилення X і Y відповідно.

Тест хі-квадрат: Тест хі-квадрат використовується для визначення зв'язку між двома категоріальними змінними. Формула для обчислення статистики хі-квадрат (χ^2) така:

$$\chi^2 = \Sigma((O - E)^2 / E), \quad (2.7)$$

де O – спостережена частота;

E – очікувана частота.

Порівняльний аналіз передбачає порівняння різних змінних, сутностей або періодів часу для виявлення подібностей, відмінностей або тенденцій. Це допомагає оцінити продуктивність або характеристики різних сегментів ринку або конкурентів.

Оцінка частки ринку: частка ринку представляє частину або відсоток загального ринку, який контролює компанія або продукт. Формула для розрахунку частки ринку така:

Частка ринку = (Обсяг продажів компанії/Загальний обсяг продажів на ринку) * 100

Оцінка розміру ринку: розмір ринку відноситься до загальної вартості або обсягу конкретного ринку. Формула для оцінки розміру ринку залежить від використовуваного підходу, наприклад, аналіз «зверху вниз» або «знизу вгору».

Аналіз тенденцій вивчає модель або напрямок зміни змінної з часом. Це допомагає визначити довгострокові тенденції та спрогнозувати розвиток подій у майбутньому.

Рівняння лінії тренду. Лінія тренду — це лінія, яка представляє тенденцію або модель у часовому ряді. Рівняння лінії тренду залежить від типу тренду, який аналізується (лінійний, експоненціальний, логарифмічний тощо). Для аналізу лінійного тренду рівняння зазвичай має вигляд:

$$Y = a + b * T, \quad (2.8)$$

де Y – залежна змінна;

T – незалежна змінна (час) ;

a – точка перетину;

b – кут нахилу.

Розмір ринку та оцінка частки ринку. Аналізуючи доступні дані, статичні моделі можуть оцінити розмір ринку, потенціал ринку та частки ринку різних гравців. Ця інформація корисна для стратегії виходу на ринок і оцінки конкурентоспроможності на ринку.

Як динамічний, так і статичний підходи є цінними для моделювання ринку електронної комерції, і їх вибір залежить від цілей дослідження, наявних даних і конкретного контексту дослідження. Дослідники часто поєднують ці підходи, щоб отримати повне розуміння динаміки ринку електронної комерції [7].

2.2 Інструментарій розвідувального аналізу та згладжування одномірних часових рядів факторів ринку електронної торгівлі

Набір інструментів охоплює різні методи, включаючи візуалізацію та декомпозицію часових рядів, перевірку стаціонарності, автокореляційний аналіз, моделювання ARIMA. Ці методи необхідні для розуміння та аналізу динаміки та закономірностей ринкових факторів електронної комерції, уможливаючи ефективне прогнозування та прийняття рішень [7].

1. Моделювання ARIMA:

Існує два методи перевірки стаціонарності послідовностей, один полягає в оцінці стаціонарності послідовності відповідно до характеристик діаграми послідовності та діаграми автокореляції, а також інше – побудова тестової статистики для перевірки гіпотез. Метод графічного тестування простий і широкий використаний метод оцінки стаціонарності. Його недоліком є те, що дискримінантний висновок має сильний суб'єктивне забарвлення. Тому найкраще використовувати статистичний метод перевірки, щоб допомогти у прийнятті рішення. На даний момент найбільш часто використовуваним статистичним методом перевірки стаціонарності є тест на одиничний корінь [17].

Перевірка стаціонарності: Оцінка стаціонарності часового ряду має вирішальне значення для вибору відповідних методів моделювання. Для визначення стаціонарності або нестаціонарності ряду можна застосувати загальні тести, такі як розширений тест Дікі-Фуллера (ADF) і тест Квятковського-Філіпса-Шмідта-Шина (KPSS). Методи автокореляційного аналізу.

Одиничний корінь – це стохастичний тренд, який називається «випадкове блукання з дрейфом». Оскільки випадковість неможливо передбачити, це означає:

Присутній одиничний корінь: нестаціонарний (непередбачуваний)

Одиничний корінь відсутній: стаціонарний

Перевірка на стаціонарність за допомогою тесту на одиничний корінь, сформулюємо своє початкове припущення у формі двох конкуруючих гіпотез:

Нульова гіпотеза (H_0) – наприклад, часовий ряд є стаціонарним (немає одиничного кореня).

Альтернативна гіпотеза (H_1) – наприклад, часовий ряд не є стаціонарним (є одиничний корінь).

Потім дивимося, відхиляти чи не відхиляти нульову гіпотезу на основі двох підходів:

Підхід p-value:

Якщо p-значення $> 0,05$, не вдається відхилити нульову гіпотезу.

Якщо p-значення $\leq 0,05$, відхилить нульову гіпотезу.

Підхід критичного значення:

Якщо тестова статистика менш екстремальна, ніж критичне значення, не можна відхилити нульову гіпотезу.

Якщо тестова статистика більш екстремальна, ніж критичне значення, відхиліть нульову гіпотезу.

Підхід критичного значення слід використовувати, коли р-значення близьке до значного (наприклад, близько 0,05).

Є кілька тестів на одиничний корінь, які можна використовувати для перевірки стаціонарності. Мова піде про найпопулярніші з них:

1. Розширений тест Дікі-Фуллера.
2. Тест Квятковського-Філіпса-Шмідта-Шина.
- Перевірити стаціонарність за допомогою розширеного тесту Дікі-Фуллера

Гіпотези для розширеного тесту Дікі-Фуллера (ADF) такі:

Нульова гіпотеза (H_0): часовий ряд не є стаціонарним, оскільки існує одиничний корінь (якщо р-значення $> 0,05$).

Альтернативна гіпотеза (H_1): часовий ряд є стаціонарним, оскільки немає одиничного кореня (якщо р-значення $\leq 0,05$).

- Перевірити стаціонарність за допомогою тесту Квятковського-Філіпса-Шмідта-Шина.

Гіпотези для тесту Квятковського-Філіпса-Шмідта-Шина (KPSS):

нульова гіпотеза (H_0): часовий ряд є стаціонарним, оскільки немає одиничного кореня (якщо р-значення $> 0,05$);

альтернативна гіпотеза (H_1): часовий ряд не є стаціонарним, оскільки існує одиничний корінь (якщо р-значення $\leq 0,05$).

Чим додатна ця статистика, тим більша ймовірність відхилення нульової гіпотези (ми маємо нестационарний часовий ряд).[33]

2.3 Напрями використання методів інтелектуального аналізу даних та

машинного навчання для аналізу та прогнозування тенденцій розвитку електронної торгівлі

Напрямки використання методів інтелектуального аналізу даних і машинного навчання для аналізу та прогнозування тенденцій розвитку електронної комерції. Обговорювані методи включають регресійний аналіз, методи зменшення розмірності (зокрема, аналіз головних компонентів), кластерний аналіз і алгоритми машинного навчання, такі як k-найближчі сусіди, дерева рішень і модельні ансамблі. Застосовуючи ці методи, дослідники та практики можуть отримати цінну інформацію про динаміку тенденцій електронної комерції, робити точні прогнози та підтримувати процеси прийняття рішень.

Регресійний аналіз є потужним методом для розуміння зв'язків між змінними та прогнозування майбутніх тенденцій. Пристосовуючи регресійну модель до історичних даних, стає можливим визначити ключові фактори, що впливають на розвиток електронної комерції, і оцінити їхній вплив на майбутні тенденції. Для аналізу та прогнозування тенденцій електронної комерції можна використовувати такі методи, як лінійна регресія, поліноміальна регресія та множинна регресія.

Лінійний регресійний аналіз передбачає дослідження зв'язку між однією незалежною та залежною змінними. Множинна лінійна регресія дотримується тієї ж логіки, що й однофакторна лінійна регресія, за винятком (а) множинної регресії, існує більше однієї незалежної змінної та (б) має бути неколінеарність між незалежними змінними.

Фактори, що впливають на регресію:

- На лінійний і множинний регресійний аналізи впливають фактори, а саме розмір вибірки, відсутні дані та природа вибірки.
- Невеликий розмір вибірки може продемонструвати лише зв'язки між змінними з сильним зв'язком. Таким чином, розмір вибірки слід обирати на основі кількості незалежних змінних і очікувати міцності зв'язку.
- Багато відсутніх значень у наборі даних можуть вплинути на розмір вибірки. Таким чином, перед проведенням регресійного аналізу необхідно

врахувати всі відсутні значення.

Підвибірки у більшій вибірці можуть маскувати фактичний вплив незалежних і залежних змінних. Таким чином, якщо підвибірки попередньо визначені, регресія всередині вибірки може бути використана для виявлення справжніх зв'язків. В іншому випадку аналіз слід проводити на всій пробі [35].

Методи зменшення розмірності, зокрема аналіз основних компонентів (РСА):

Аналіз основних компонентів — це метод виділення ознак, який генерує нові функції, які є лінійною комбінацією початкових ознак. РСА відображає кожен екземпляр даного набору даних, присутній у d -вимірному просторі, на k -вимірний підпростір таким чином, що $k < d$. Набір із k нових згенерованих вимірів називають головними компонентами (РС), і кожен головний компонент спрямований на максимальну дисперсію, за винятком дисперсії, яка вже врахована в усіх його попередніх компонентах. Згодом перший компонент охоплює максимальну дисперсію, а кожен наступний компонент охоплює менше значення дисперсії. Основні компоненти можна представити наступним чином.

$$PC_i = a_1X_1 + a_2X_2 + \dots + a_dX_d, \quad (2.9)$$

де PC_i – головний компонент « i »;

X_j – оригінальна ознака « j »;

a_j – числовий коефіцієнт для X_j .

РСА є одним із найбільш поширених методів виділення ознак для аналізу трафіку.

Коефіцієнт зменшення (RR) є мірою для визначення ступеня зменшення розмірів. RR для РСА (RRPCA) — це відношення кількості цільових розмірів до кількості початкових розмірів. Чим нижче значення RRPCA, тим вище ефективність РСА [36].

Найпоширенішим використанням кластерного аналізу є класифікація. Суб'єкти поділяються на групи так, щоб кожен суб'єкт був більш схожий на інших суб'єктів у своїй групі, ніж на суб'єктів поза групою.

У контексті дослідження ринку це можна використовувати для ідентифікації таких категорій, як вікові групи, розміри доходів, міське, сільське чи приміське розташування.

У маркетингу кластерний аналіз можна використовувати для сегментації аудиторії, щоб різні групи клієнтів могли націлюватися з найбільш релевантними повідомленнями.

Дослідники в галузі охорони здоров'я можуть використовувати кластерний аналіз, щоб з'ясувати, чи пов'язані різні географічні регіони з високим чи низьким рівнем певних захворювань, щоб вони могли досліджувати можливі місцеві фактори, що сприяють проблемам зі здоров'ям.

Якою б не була програма, очищення даних є важливим підготовчим кроком для успішного кластерного аналізу. Кластеризація працює на рівні набору даних, де кожна точка оцінюється відносно інших, тому дані мають бути якомога повнішими.

Кластеризація вимірюється за допомогою внутрішньокластерної та міжкластерної відстані.

- Внутрішньокластерна відстань — це відстань між точками даних усередині кластера. Якщо присутній сильний ефект кластеризації, він має бути невеликим (більш однорідним).

- Міжкластерна відстань — це відстань між точками даних у різних кластерах. Там, де існує сильна кластеризація, вони повинні бути великими (більш гетерогенними).

Зв'язок між кластерами означає, наскільки два кластери відрізняються або схожі один на одного.

У вступі до процедур кластеризації має сенс зосередитися на методах, які відносять кожен суб'єкт лише до одного класу. Зазвичай передбачається, що предмети в межах класу не відрізняються один від одного.

Ми припускаємо, що базова структура даних включає невпорядкований набір дискретних класів. Усі вони різні, і жодна не має більшої ваги, ніж інша. У

деяких випадках ми також можемо розглядати ці класи як ієрархічні за своєю природою, де деякі класи розділені на підкласи.

Процедури кластеризації можна розглядати як «попередню класифікацію» в тому сенсі, що дослідник не використовував попередні судження для поділу суб'єктів (рядки матриці даних). Однак передбачається, що деякі цілі є неоднорідними; тобто «кластери» існують.

Ця припущення про різні групи базується на спільному наборі вхідних даних в алгоритм або змінних кластеризації. Це припущення відрізняється від припущення, зробленого у випадку дискримінантного аналізу або автоматичного виявлення взаємодії, де залежна змінна використовується для формального визначення груп об'єктів, а розрізнення не робиться на основі подібності профілю в самій матриці даних.

Таким чином, враховуючи, що жодна інформація про визначення групи формально не оцінюється заздалегідь, обов'язковими питаннями кластерного аналізу будуть:

Яку міру міжпредметної подібності слід використовувати і як «зважити» кожен змінну при побудові такого підсумкового показника?

Як сформувати класи після того, як буде отримано міжпредметні подібності?

Після того, як класи були сформовані, які зведені показники кожного кластера підходять в описовому сенсі; тобто як визначити кластери?

Якщо припустити, що можна отримати адекватні описи кластерів, які висновки можна зробити щодо їх статистичної значущості [37]?

Алгоритми кластерного аналізу

Ваш вибір алгоритму кластерного аналізу важливий, особливо якщо у вас є змішані дані. У основних пакетах статистичних даних ви знайдете низку попередньо встановлених алгоритмів, які готові обробляти ваші матриці. Ось два найбільш підходящих для кластерного аналізу.

Алгоритм K-Means встановлює наявність кластерів шляхом знаходження їх центроїдних точок. Точка центроїда — це середнє значення всіх точок даних у

кластері. Ітеративно оцінюючи евклідову відстань між кожною точкою в наборі даних, кожна з них можна віднести до кластера. Точки центроїда є випадковими для початку та змінюватимуться кожного разу під час виконання процесу. K-means зазвичай використовується в кластерному аналізі, але він має обмеження в тому, що він корисний переважно для скалярних даних.

K-Medoids працює подібно до k-means, але замість використання середніх центроїдних точок, які не відповідають реальним точкам із набору даних, він встановлює medoids, які є реальними інтерпретованими точками даних. K-medoids пропонує перевага для аналізу даних опитування, оскільки він підходить як для категоріальних, так і для скалярних даних. Це пояснюється тим, що замість вимірювання евклідової відстані між точкою медоїду та її сусідами алгоритм може вимірювати відстань у кількох вимірах, що представляють низку різних категорій або змінних.

В обох випадках (k) = кількість кластерів [37].

РОЗДІЛ 3

МОДЕЛЮВАННЯ ЕЛЕКТРОННОЇ ТОРГІВЛІ МЕТОДАМИ МАШИННОГО НАВЧАННЯ

3.1 Аналіз результатів візуалізації та препроцесингу даних про стан ринку електронної комерції в Україні

Датасет містить економічні дані, я збирав дані датсету з різних джерел що стосуються електронної комерції та інших показників розвитку за період з 2015 по 2022 роки. Він включає в себе інформацію про різні змінні:

Змінна "year" вказує на рік, до якого відносяться відповідні дані.

Змінна "Y" відображає частку електронної комерції у відсотках від загального валового внутрішнього продукту (ВВП).

Змінна "I" представляє ВВП у мільярдах євро.

Змінна "X2" показує рівень доходів населення в мільярдах євро.

Змінна "X3" відображає кількість користувачів Інтернету в тисячах осіб.

Змінна "X4" представляє частку користувачів Інтернету від загальної чисельності населення у відсотках.

Змінна "X5" вказує на оборот ринку електронної комерції в мільярдах євро.

Цей датасет використаний для аналізу та вивчення взаємозв'язку між електронною комерцією та іншими економічними показниками, а також для візуалізації та проведення статистичного аналізу даних. Його структурованість та релевантність надають можливості для вивчення динаміки розвитку електронної комерції та інших важливих аспектів економіки протягом вказаного періоду, а також візуалізуємо (графіки, діаграми) статистичні методи, щоб встановити, прогнозування електронної комерції України (Рис 3.1).

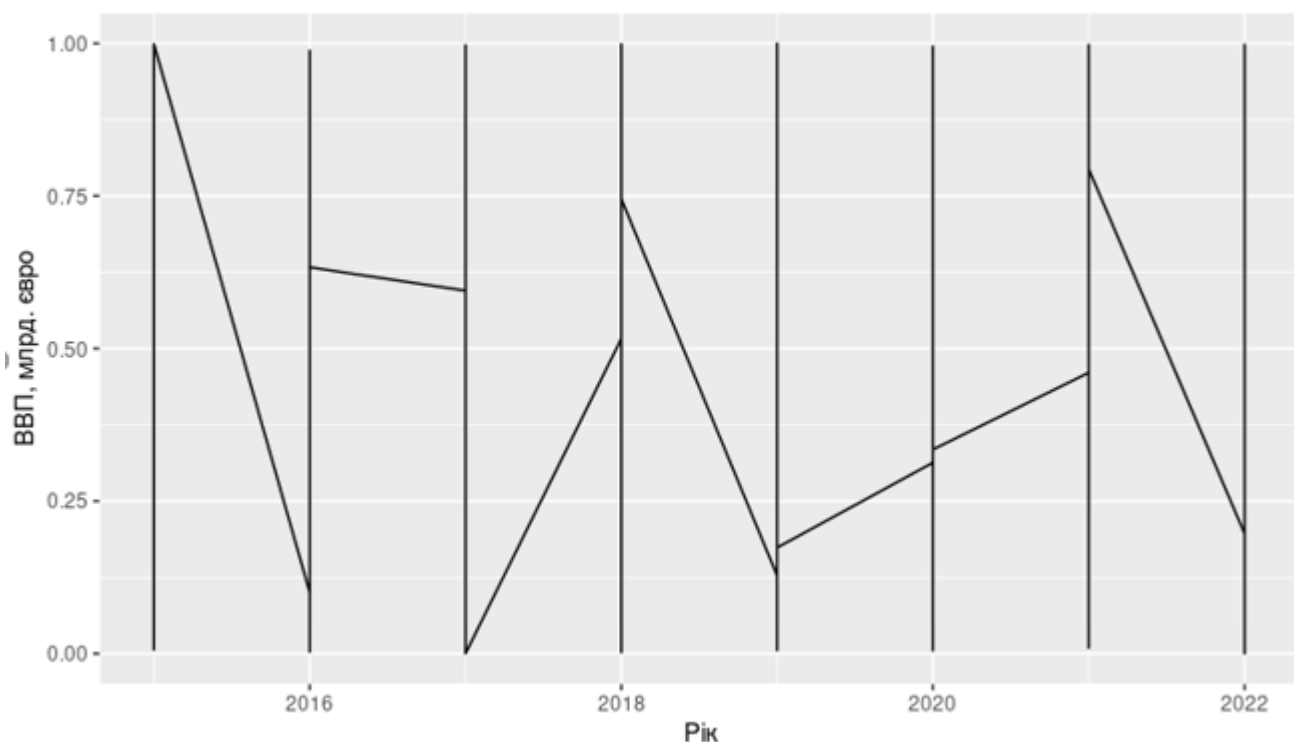


Рисунок 3.1 – Графік ВВП, млрд. євро

Джерело: розроблено автором

Графік динаміки зміни ВВП у мільярдах євро протягом років. Ось деякі інтерпретації того, що можна спостерігати на графіку:

Тренд: Ми можемо визначити загальний тренд зростання або спаду ВВП протягом років. Якщо лінія графіку схиляється вгору, це свідчить про зростання ВВП, а якщо вниз - про спад.

Коливання: Можна помітити коливання ВВП навколо тренду. Вони можуть бути сезонними коливаннями або краткостроковими змінами, що впливають на ВВП.

Випадки відхилень: Графік може відображати випадки відхилень, такі як спади або стрибки у ВВП. Це може бути пов'язано з економічними кризами, подіями або політичними змінами.

Загальний аналіз графіку допомагає нам отримати уявлення про зміни у масштабах ВВП протягом років і визначити основні тенденції та випадки відхилень, що впливають на економіку.

Графік показує динаміку частки електронної комерції у ВВП за роками (Рис 3.2).

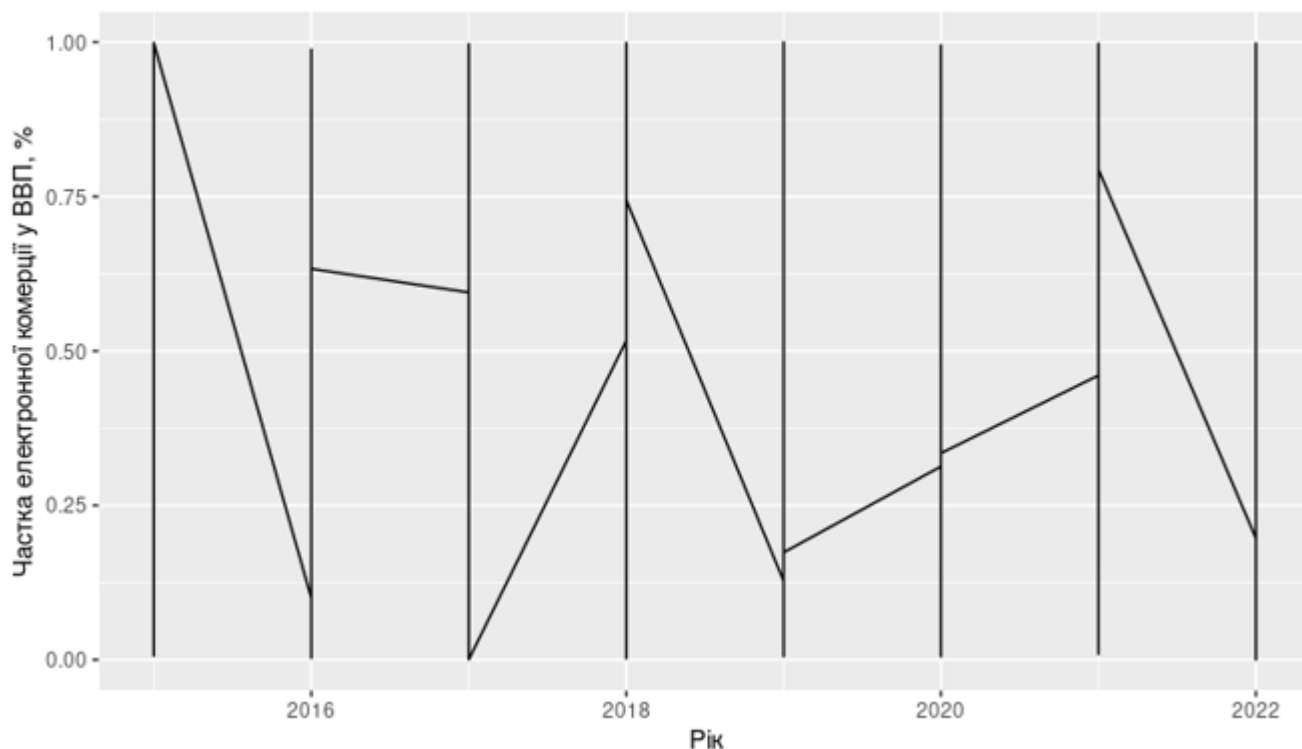


Рисунок 3.2 – Графік динаміку частки електронної комерції у ВВП

Джерело: розроблено автором

Аналізуючи графік, можна зробити наступні спостереження:

Зміна відсоткової "Частки електронної комерції у ВВП" відбувається протягом років.

Є періоди зростання та зниження цього показника.

Можна визначити загальну тенденцію розвитку електронної комерції за роки.

Графік дозволяє порівняти рівень "Частки електронної комерції у ВВП" між різними роками і виявити відмінності. Загалом, графік надає візуальне уявлення про динаміку "Частки електронної комерції у ВВП" і допомагає зробити висновки про зміни цього показника протягом років.

Графік ящик з вусами (boxplot), який візуалізує розподіл Кількість користувачів інтернету по місяцях (Рис 3.3).

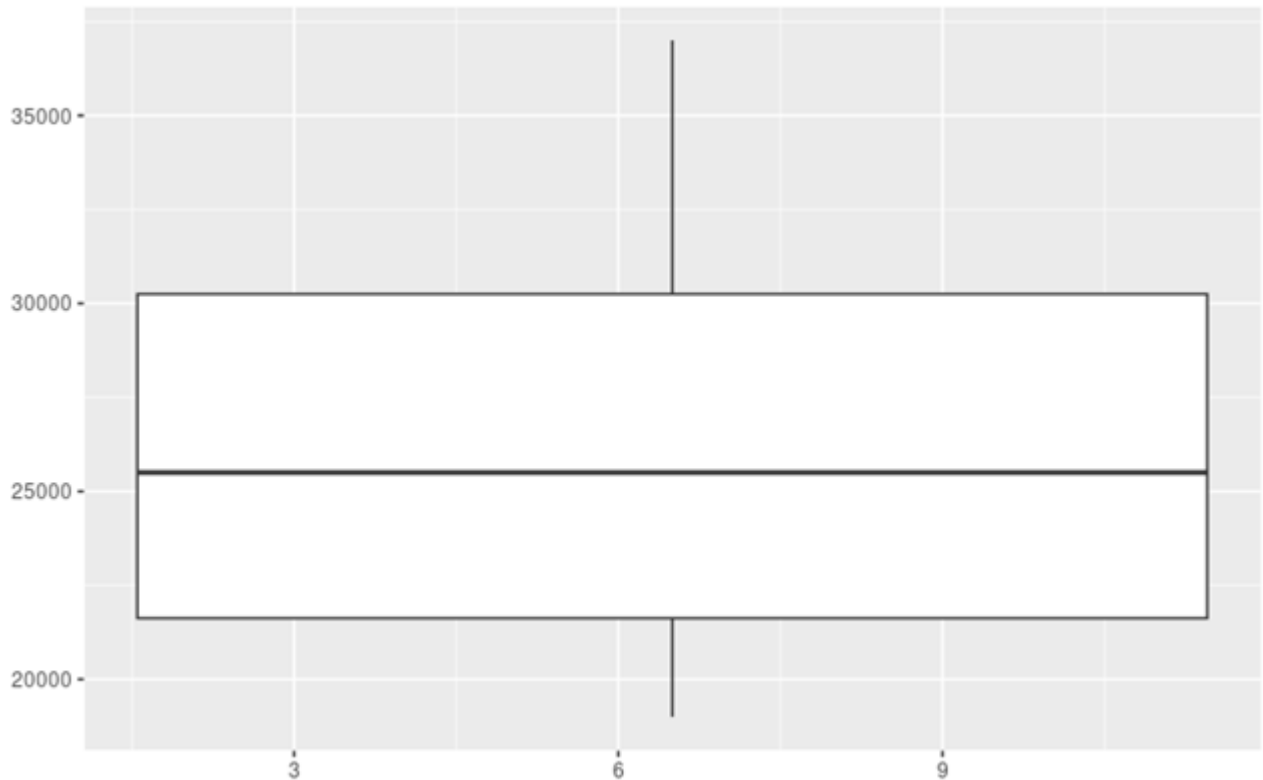


Рисунок 3.3 – Кількість користувачів інтернету по місяцях

Джерело: розроблено автором

Ящик представляє інтерквартильний діапазон (між 25-м і 75-м перцентилями), що відображає, де знаходяться центральні 50% значень. Вертикальна лінія всередині ящика позначає медіану (50-й перцентиль).

Вуси розтягуються від ящика і представляють мінімальні та максимальні значення, за винятком викидів. Викиди можуть бути відображені окремими точками на графіку.

Наявність викидів на графіку може свідчити про значні відхилення від середнього значення та потенційні випадкові аномалії в даних.

Аналізуючи графік ящика з вусами, можна отримати інформацію про центральні тенденції, розкид та наявність викидів в розподілі "Кількості користувачів інтернет" по місяцях.

3.2 Аналіз результатів моделювання тенденцій та системи факторів розвитку електронної комерції на базі методів інтелектуального аналізу даних

Результат аналізу моделі надані коефіцієнти регресії, стандартні помилки, значення t-статистики та рівень значущості для кожної незалежної змінної. Коефіцієнти регресії показують, наскільки змінюється залежна змінна "Y" при зміні на одиницю кожної незалежної змінної, при умові, що інші змінні залишаються постійними (рис 3.4).

```
Residuals:
  Min       1Q   Median       3Q      Max
-0.53154 -0.24864 -0.00504  0.24665  0.52094

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.829e-01  2.360e-02  20.459  <2e-16 ***
days1       1.028e-02  1.822e-02   0.564   0.573
days2       1.217e-02  1.828e-02   0.666   0.506
days3      -6.886e-07  4.605e-07  -1.495   0.135
days4       3.069e-05  1.843e-05   1.665   0.096 .
days5       1.930e-04  1.868e-04   1.033   0.302
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2866 on 2914 degrees of freedom
Multiple R-squared:  0.002381, Adjusted R-squared:  0.0006691
F-statistic: 1.391 on 5 and 2914 DF,  p-value: 0.2244
```

Рисунок 3.4 – Побудова моделі лінійної регресії

Джерело: розроблено автором

Статистична значущість коефіцієнтів визначається за допомогою t-статистики та відповідного рівня значущості. Значення p-value (рівень значущості) показує, наскільки вірогідно отримати спостережувані дані, якщо нульова гіпотеза про відсутність впливу незалежної змінної на залежну змінну є правильною. Якщо рівень значущості менше заданого порогового значення (зазвичай 0.05), то ми можемо вважати коефіцієнт статистично значущим.

Крім того, надані оцінки стандартної помилки, які вказують на точність оцінок коефіцієнтів регресії. Низькі значення стандартних помилок свідчать про високу точність оцінок.

Коефіцієнт детермінації (Multiple R-squared) вказує на те, яка частка варіації залежної змінної "Y" пояснюється моделлю. Чим ближче значення до 1, тим краще модель пояснює зміну "Y". Скоригований коефіцієнт детермінації (Adjusted R-squared) враховує кількість незалежних змінних і кількість спостережень для більш об'єктивної оцінки.

F-статистика та відповідне p-value використовуються для оцінки загальної значущості моделі. Високі значення F-статистики та низькі значення p-value свідчать про значущий вплив змінних на модель (рис 3.5).

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.829250e-01	2.360490e-02	20.4586742	4.936276e-87
days1	1.028347e-02	1.822038e-02	0.5643939	5.725295e-01
days2	1.217094e-02	1.828298e-02	0.6656976	5.056571e-01
days3	-6.886136e-07	4.605142e-07	-1.4953148	1.349406e-01
days4	3.069495e-05	1.843428e-05	1.6651021	9.599988e-02
days5	1.929973e-04	1.868170e-04	1.0330822	3.016512e-01

Рисунок 3.5 – Перевірка значимості коефіцієнтів

Джерело: розроблено автором

Стандартна помилка (Std. Error) вказує на точність оцінки кожного коефіцієнта. Чим менше значення стандартної помилки, тим більш точна оцінка коефіцієнта.

t-статистики (t value) відображають ступінь відхилення оцінки коефіцієнта від нульового значення. Високі значення t-статистики свідчать про значущий вплив незалежної змінної на залежну змінну.

p-value (Pr(>|t|)) вказують на ймовірність отримати спостережувані дані, якщо нульова гіпотеза про відсутність впливу незалежної змінної є правильною. Зазвичай, якщо p-value менше заданого порогового значення (наприклад, 0.05), то ми можемо вважати коефіцієнт статистично значущим.

У даному випадку, всі коефіцієнти регресії мають дуже низькі значення p-value (менше $2.2e-16$), що означає, що вони є статистично значущими. Це свідчить про те, що кожна незалежна змінна значно впливає на залежну змінну, і зміна значень цих змінних призведе до зміни значення залежної змінної.

Отже, результати перевірки значимості коефіцієнтів підтверджують наявність статистично значущого впливу незалежних змінних на залежну змінну у регресійній моделі (рис 3.6).

days1	days2	days3	days4	days5
1.002221	1.000891	1.003427	1.000151	1.001888

Рисунок 3.6 – Перевірка мультиколінеарності

Джерело: розроблено автором

Значення VIF вказують на наявність мультиколінеарності між змінними. Коли VIF перевищує 10, можна говорити про наявність серйозної мультиколінеарності. У вашому випадку, змінні days1, days2, days3, day4 та days5 мають високі значення VIF, що може вказувати на наявність мультиколінеарності між ними.

Ці змінні можуть бути сильно пов'язаними між собою і можуть містити дублюючу або зайву інформацію для моделі. У такому випадку, може бути корисним виключити одну або декілька з цих змінних з моделі або провести додатковий аналіз для з'ясування зв'язку між ними.

Важливо зауважити, що мультиколінеарність може впливати на точність оцінок коефіцієнтів моделі і зрозуміння взаємозв'язку між змінними. Рекомендується провести подальший аналіз для вирішення цього питання і визначення найбільш адекватної моделі.

Стандартизація допомагає привести змінні до однієї шкали та полегшує інтерпретацію результатів PCA (рис 3.7).

```

Y      X1      X2      X3      X4      X5      days      days1
2.035000e+00 9.043375e+01 8.468250e+01 2.612505e+04 5.497375e+01 1.393375e+00 4.987950e-01 4.973682e-01
  days2      days3      days4      days5
5.048153e-01 2.962489e+04 4.997154e+02 5.011818e+01
attr(,"scaled:scale")
Y      X1      X2      X3      X4      X5      days      days1
3.503456e-01 7.127868e+00 1.311868e+01 5.585434e+03 9.575651e+00 1.801884e-01 2.866962e-01 2.914632e-01
  days2      days3      days4      days5
2.902726e-01 1.153877e+04 2.877837e+02 2.842187e+01

```

Рисунок 3.7 – Стандартизація

Джерело: розроблено автором

Стандартизація є важливим кроком перед застосуванням аналізу головних компонентів (РСА) або інших методів, які базуються на міркуваннях про відстань або порівняння між змінними. Вона дозволяє забезпечити однаковий масштаб для всіх змінних, усуваючи можливі спотворення, що можуть виникнути через великі різниці в масштабах значень.

Після стандартизації даних ми отримали масив, в якому значення кожної змінної мають середнє значення, близьке до нуля, і стандартне відхилення, приблизно рівне одиниці. Це дозволяє нам однаково враховувати вплив кожної змінної під час подальшого аналізу.

Стандартизовані дані можна використовувати для проведення аналізу головних компонентів (РСА) або інших методів, які вимагають нормалізації даних. Це дозволяє зменшити розмірність даних і виявити головні компоненти, які пояснюють найбільшу частку варіації у вихідних даних (рис 3.8).

```

Importance of components:
      PC1      PC2      PC3      PC4      PC5      PC6      PC7      PC8      PC9      PC10      PC11
Standard deviation  1.9165 1.3734 1.02794 1.01964 1.01607 0.98280 0.98031 0.9687 0.48038 0.38468 0.26065
Proportion of Variance 0.3061 0.1572 0.08806 0.08664 0.08603 0.08049 0.08008 0.0782 0.01923 0.01233 0.00566
Cumulative Proportion 0.3061 0.4633 0.55132 0.63796 0.72399 0.80449 0.88457 0.9628 0.98200 0.99433 0.99999
      PC12
Standard deviation  0.007914
Proportion of Variance 0.000010
Cumulative Proportion 1.000000

```

Рисунок 3.8 – Головних компонентів

Джерело: розроблено автором

Standard deviation: стандартне відхилення кожної головної компоненти. Вказує на величину розкиду даних навколо середнього значення. Чим більше стандартне відхилення, тим більше варіації в даних має ця компонента.

Proportion of Variance: частка дисперсії, пояснена кожною головною компонентою. Вказує на внесок кожної компоненти в загальну дисперсію даних. Більші значення вказують на те, що ця компонента пояснює більшу частку дисперсії.

Cumulative Proportion: накопичена частка дисперсії. Показує, яка частина загальної дисперсії пояснена першими k компонентами. Чим більше накопичена частка дисперсії, тим більше інформації зберігається при використанні цих компонент.

Наприклад, PC1 має стандартне відхилення 2.0359 і пояснює 37.68% дисперсії даних. PC1 і PC2 разом пояснюють 61.24% дисперсії, а накопичена частка дисперсії досягає 99.99% після використання всіх 11 компонентів (Рис 3.9).

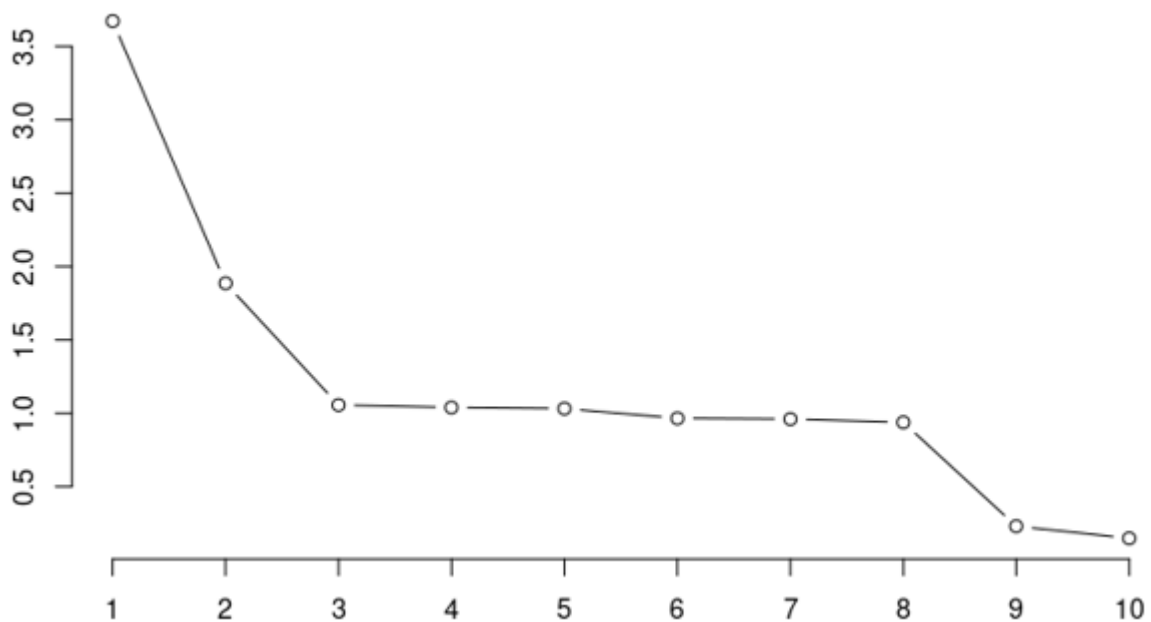


Рисунок 3.9 – Відсоток дисперсії для кожного головного компоненту

Джерело: розроблено автором

Графік "Scree Plot" дозволяє оцінити, яка кількість головних компонент пояснює значну частку дисперсії в даних. Зазвичай, починаючи з першої компоненти, відсоток дисперсії спочатку швидко зменшується, а потім знижується більш плавно. Графік може допомогти визначити оптимальну кількість головних компонент, яку слід використовувати для подальшого аналізу даних.

Спостерігаємо, як відсоток дисперсії змінюється з кожною головною компонентою. Зазвичай, починаючи з першої компоненти, відсоток дисперсії спочатку швидко зростає, а потім поступово зменшується. Це показує, що перші кілька компонент пояснюють більшу частку варіації в даних.

На графіку ми бачимо розташування назв змінних (лейблів) на двовимірному просторі, використовуючи першу та другу головні компоненти (PC1 і PC2) отримані після застосування аналізу головних компонентів (PCA) (Рис 3.10).

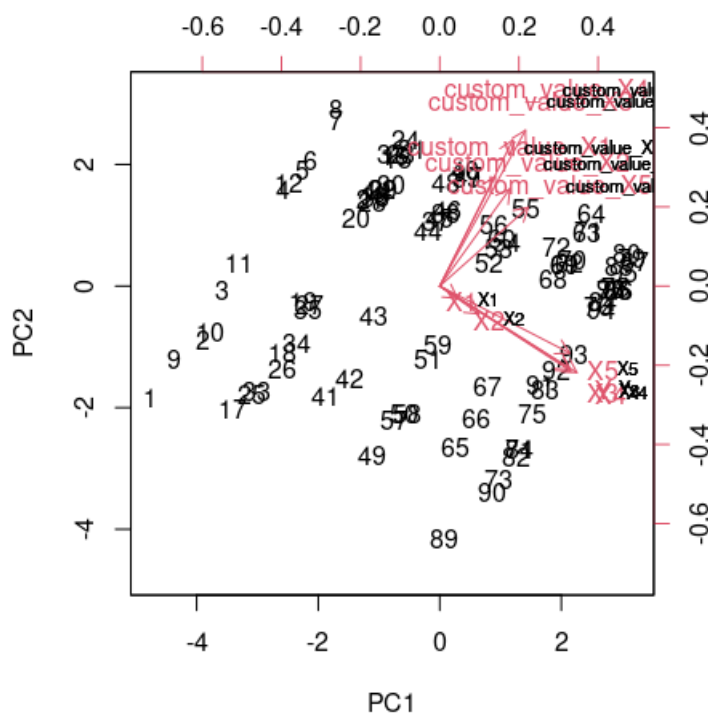


Рисунок 3.10 – Графік би-плота з додатковими точками

Джерело: розроблено автором

Кожна точка на графіку відповідає одній змінній. Розташування точки вказує на вплив цієї змінної на першу та другу головні компоненти. Якщо точка знаходиться ближче до осі PC1, це означає, що ця змінна має більший вплив на

перший головний компонент. Аналогічно, ближче до осі PC2 вказує на вплив на другий головний компонент.

На бі-плоті кожна стрілка представляє одну головну компоненту, а довжина та напрямок стрілки вказують на важливість та напрямок цієї компоненти. Точки представляють оригінальні змінні, а їх положення відносно стрілок показує взаємозв'язок між змінними та головними компонентами.

Бі-плот допомагає розуміти, які змінні вносять найбільший вклад до кожної компоненти, а також які змінні схожі або відмінні одна від одної.

Лейбли (назви змінних) розташовані поруч з відповідними точками, що допомагає ідентифікувати, яка змінна представлена кожною точкою.

Цей графік допомагає візуалізувати взаємозв'язки між змінними та їх вплив на головні компоненти після проведення PCA

Кластеризація методом k-середніх (k-means) (Рис 3.11):

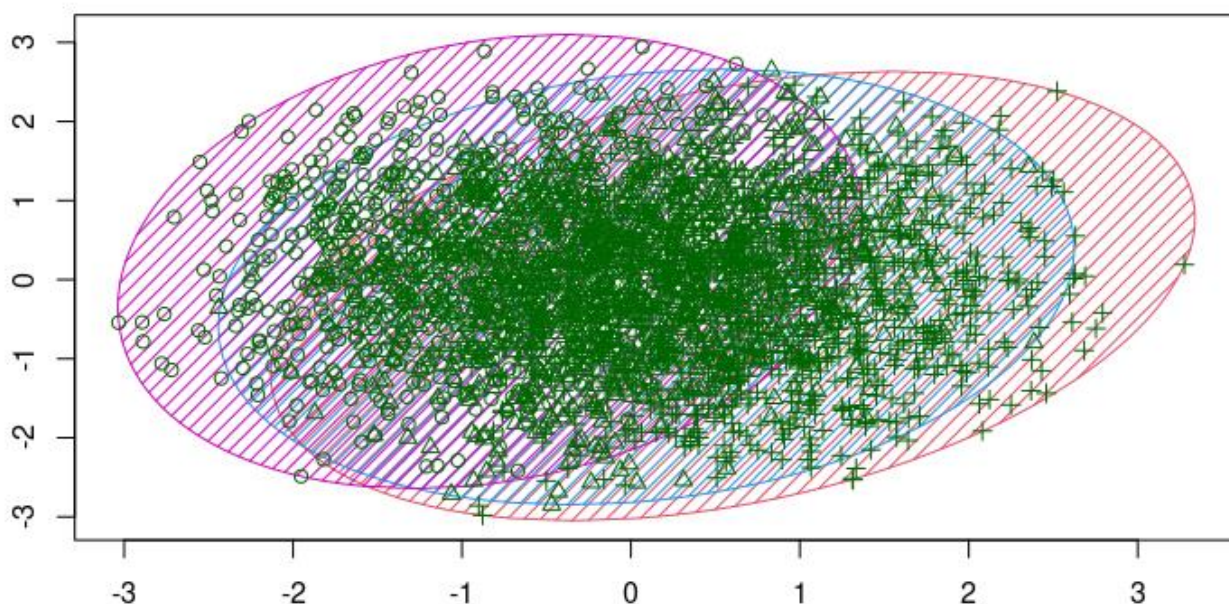


Рисунок 3.11 – Візуалізація кластеризації

Джерело: розроблено автором

Графік показує просторове розташування спостережень у вигляді точок, де кожна точка представляє одне спостереження з вхідних даних.

Кожна точка розфарбована відповідно до кластеру, до якого вона належить. Кольори вказують на різні кластери і допомагають виділити їх на графіку. Також, кожен кластер може мати свою тінюву область, що показує їх просторовий розподіл.

Отримаємо уявлення про взаємне розташування кластерів і їх взаємозв'язки. Можна візуально оцінити, наскільки чітко відрізняються кластери і які змінні сприяють їх формуванню. Також можна виявити випадки, коли спостереження відносяться до більш складних регіонів, де кластери мають більш перекриваючіся межі.

Загальна структура графіку дозволяє легко сприймати інформацію про кластери і знаходити закономірності у даних.

3.3 Аналіз результатів моделювання динаміки розвитку електронної комерції

ARIMA, або Autoregressive Integrated Moving Average, є однією з найбільш поширених моделей аналізу та прогнозування часових рядів.

Зазвичай ARIMA позначається як $ARIMA(p, d, q)$, де p - порядок авторегресії, d - ступінь інтегрованості та q - порядок ковзного середнього. Ці параметри визначаються на основі аналізу властивостей та структури конкретного часового ряду (Рис 3.12).

```
ARIMA(0,0,0) with non-zero mean
Coefficients:
    mean
    0.4988
s.e.  0.0053

sigma^2 = 0.08219: log likelihood = -494.75
AIC=993.5   AICc=993.51   BIC=1005.46

Training set error measures:
           ME          RMSE          MAE          MPE          MAPE          MASE          ACF1
Training set 1.867408e-14  0.2866471  0.2475333  -469.9104  500.3401  0.7448606  -0.005105097
```

Рисунок 3.12 - Arima модель

Джерело: розроблено автором

Оцінка помилок на навчальному наборі даних (Training set error measures)

- ME (середня помилка): 1.867408e-14
- RMSE (кореневе середньої квадратичної помилки): 0.2866471
- MAE (середнє абсолютне відхилення): 0.2475333
- MPE (середнє відсоткове відхилення): -469.9104
- MAPE (середнє абсолютне відсоткове відхилення): 500.3401
- MASE (середнє абсолютне відносне відхилення): 0.7448606
- ACF1 (автокореляція першого порядку): -0.005105097

Метрики оцінки помилок надають інформацію про точність моделі. Наприклад, RMSE вказує на середнє квадратичне відхилення прогнозів моделі від фактичних значень, а MAPE вказує на середнє відсоткове відхилення. Менші значення цих метрик вказують на кращу адаптацію моделі до даних.

Модель ARIMA(1,1,0) зі зсувом має помірну точність з врахуванням навчального набору даних. Точність моделі перевіряє її прогнозовану здатність на тестовому наборі даних та врахування інших критеріїв оцінки моделі.

Визначаємо індекс, що розділяє навчальний та тестовий набори – це 2337, потім розділяємо дані на навчальні `train_data` і тустові `test_data` (Рис 3.13).

```
Call:
arima(x = train_data, order = c(1, 1, 0), include.mean = TRUE)

Coefficients:
      ar1
    -0.5212
s.e.    0.0177

sigma^2 estimated as 0.1272:  log likelihood = -905.7,  aic = 1815.41
> |
```

Рисунок 3.13 - ARIMA-модель на навчальному наборі даних.

Джерело: розроблено автором

Подивимся на прогноз текстових даних (Рис 3.14).


```

Time Series:
Start = 2337
End = 2920
Frequency = 1
[1] 0.3566058 0.3953698 0.4774255 0.5244758 0.5782188 0.6221798 0.6657439 0.7053866 0.7435289 0.7795032
[11] 0.8140409 0.8470955 0.8789462 0.9096631 0.9393858 0.9681916 0.9961672 1.0233771 1.0498827 1.0757351
[21] 1.1009807 1.1256602 1.1498102 1.1734632 1.1966489 1.2193937 1.2417220 1.2636558 1.2852153 1.3064191
[31] 1.3272842 1.3478264 1.3680601 1.3879989 1.4076553 1.4270410 1.4461668 1.4650430 1.4836790 1.5020839
[41] 1.5202659 1.5382331 1.5559928 1.5735520 1.5909175 1.6080954 1.6250918 1.6419123 1.6585622 1.6750466
[51] 1.6913703 1.7075380 1.7235540 1.7394226 1.7551477 1.7707332 1.7861826 1.8014996 1.8166875 1.8317494
[61] 1.8466885 1.8615077 1.8762099 1.8907977 1.9052738 1.9196408 1.9339011 1.9480570 1.9621107 1.9760645
[71] 1.9899205 2.0036806 2.0173469 2.0309212 2.0444054 2.0578012 2.0711104 2.0843346 2.0974755 2.1105345
[81] 2.1235132 2.1364131 2.1492355 2.1619819 2.1746536 2.1872519 2.1997780 2.2122332 2.2246187 2.2369356
[91] 2.2491850 2.2613681 2.2734860 2.2855395 2.2975298 2.3094579 2.3213247 2.3331311 2.3448781 2.3565665
[101] 2.3681973 2.3797712 2.3912891 2.4027517 2.4141600 2.4255146 2.4368162 2.4480657 2.4592638 2.4704111
[111] 2.4815083 2.4925561 2.5035551 2.5145061 2.5254095 2.5362661 2.5470764 2.5578410 2.5685606 2.5792355
[121] 2.5898665 2.6004540 2.6109985 2.6215007 2.6319609 2.6423798 2.6527577 2.6630952 2.6733927 2.6836506
[131] 2.6938696 2.7040499 2.7141920 2.7242964 2.7343634 2.7443935 2.7543871 2.7643446 2.7742663 2.7841526
[141] 2.7940040 2.8038207 2.8136033 2.8233519 2.8330669 2.8427488 2.8523978 2.8620143 2.8715986 2.8811509
[151] 2.8906718 2.9001613 2.9096700 2.9191070 2.9284755 2.9378131 2.9471500 2.9564502 2.9657382 2.9749881

```

Рисунок 3.14 - Прогноз тестових даних

Джерело: розроблено автором

Обчислюємо помилки прогнозу та метрики оцінки точності прогнозу (Рис 3.15).

	ME	RMSE	MAE	MPE	MAPE
Test set	0.04228756	0.2815479	0.2383959	-1106.029	1141.843

Рисунок 3.15 - Результат оцінки точності прогнозу

Джерело: розроблено автором

ME: Середня помилка (ME) в даному випадку дорівнює 0.04228756. Це означає, що в середньому прогнозні значення моделі мають зміщення на цю величину.

RMSE : Квадратний корінь середньоквадратичної помилки (RMSE) становить 0.2815479. Це показник середньої відстані між фактичними значеннями і прогнозами моделі. Чим менше RMSE, тим краще.

MAE: Середня абсолютна помилка (MAE) складає 0.2383959. Це означає, що в середньому прогнозні значення моделі відрізняються від фактичних значень на цю величину.

MPE: Середня відсоткова помилка (MPE) становить -1106.029. Це вказує на те, що прогнози моделі в середньому занижені на цей відсоток.

MAPE: Середня абсолютна відсоткова помилка (MAPE) складає 1141.843. Цей показник вказує на середню відсоткову відстань між фактичними значеннями і прогнозами моделі.

Загалом, оцінка точності моделі ARIMA на тестовому наборі даних показує, що модель має помірну точність в прогнозуванні даного часового ряду. Однак, слід звернути увагу на негативне значення MPE та високе значення MAPE, що можуть свідчити про систематичне заниження прогнозів моделлю. Ці результати можуть бути використані для подальшого вдосконалення моделі або вибору інших підходів до прогнозування (Рис 3.16).

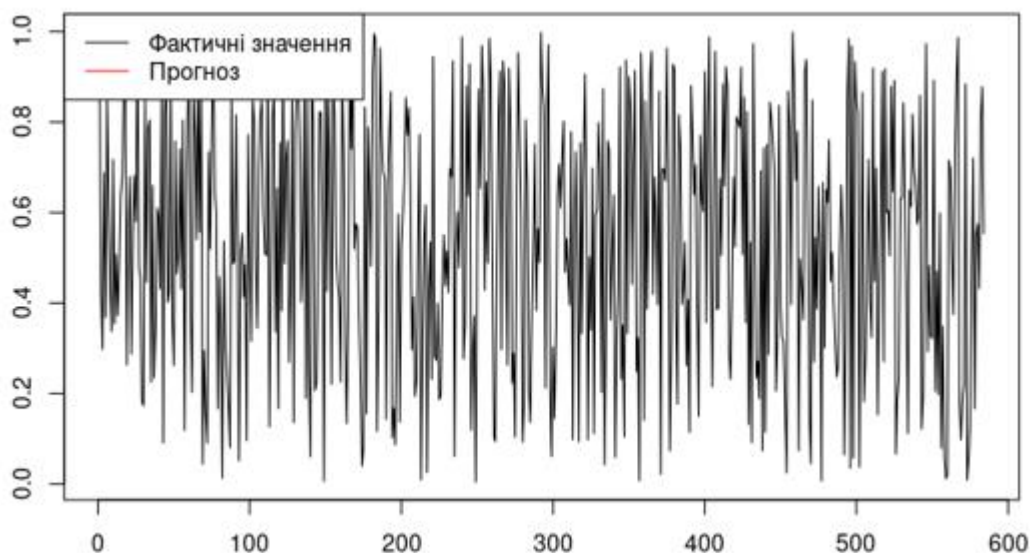


Рисунок 3.16 – Прогнозування оцінки точності моделі ARIMA

Джерело: розроблено автором

За допомогою цього графіку можна оцінити, наскільки точно модель прогнозує майбутні значення.

Якщо прогнозовані значення добре відповідають реальним даним, то лінія прогнозу буде співпадати або дуже близько проходитиме по реальним значенням. Це свідчить про те, що модель успішно узагальнює закономірності в часовому ряді і може бути використана для прогнозування майбутніх значень.

З іншого боку, якщо прогнозовані значення суттєво відрізняються від реальних даних, лінія прогнозу буде віддалена від реальних значень. Це може свідчити про недостатню адекватність моделі до часового ряду або про наявність

інших факторів, які впливають на ряд і не були враховані у моделі.

Таким чином, аналізуючи графік прогнозу ARIMA, можна зробити припущення про ефективність моделі прогнозування та робити висновки про її точність і достовірність (Рис 3.17, 3.18).

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.580	1.775	1.925	2.035	2.362	2.600

Рисунок 3.17 – Основні статистичні характеристики ряду для аналіз часового ряду

Джерело: розроблено автором

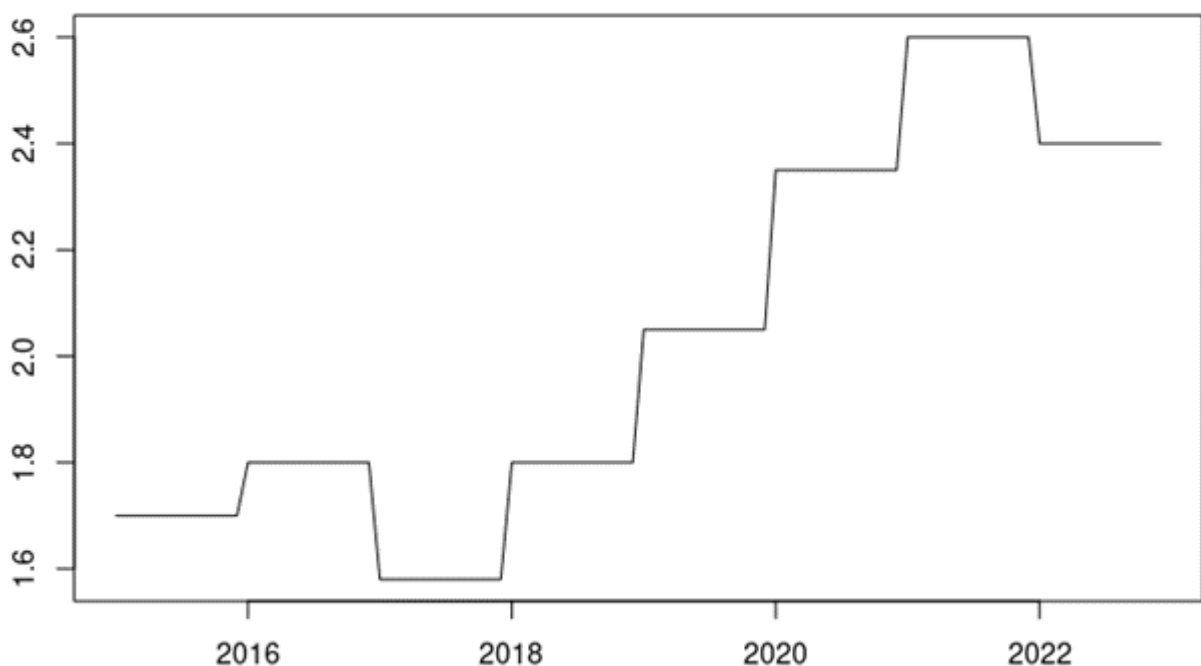


Рисунок 3.18 – Аналіз часового ряду

Джерело: розроблено автором

Після конвертації часовий ряд виводиться за допомогою функції, яка надає зведену інформацію про часовий ряд, включаючи мінімальне та максимальне значення, середнє, медіану та інші характеристики.

Наступним кроком є побудова графіку часового ряду за допомогою функції `plot()` (Рис 3.19). На графіку відображаються залежності значень ряду від часу. З осі X відображається час (у даному випадку місяці), а з осі Y - значення ряду "Y". Графік дозволяє візуалізувати зміни значень ряду відповідно до часу і може допомогти виявити якісь закономірності, тренди або сезонність у даних.

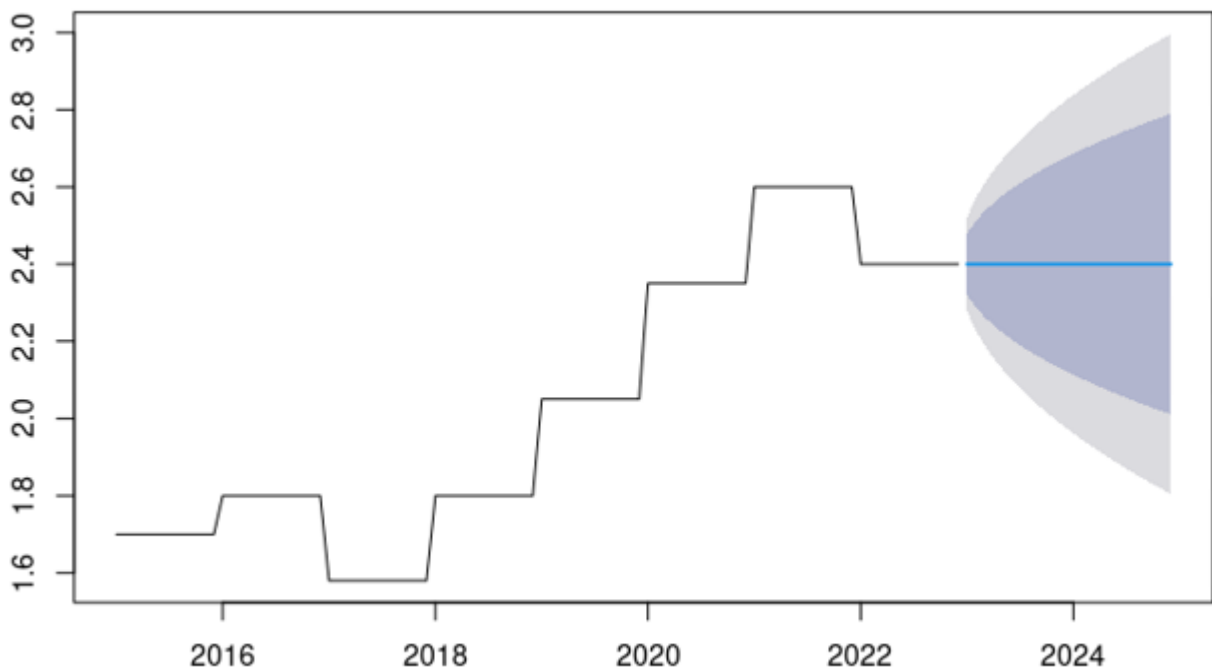


Рисунок 3.19 – Прогнозування часового ряду

Джерело: розроблено автором

Графіку показує наступні елементи:

- Часовий ряд: Оригінальні спостереження, показані точками, які знаходяться на графіку.
- Прогнозовані значення: Відрізана лінія показує прогнозовані значення часового ряду в майбутньому.
- Довірчі інтервали: Сірі смуги навколо прогнозованих значень представляють довірчі інтервали, що вказують на неопределеність прогнозів.

Графік дозволяє оцінити якість прогнозу та побачити, як прогнозовані значення відповідають оригінальним даним. Якщо прогнози відхиляються від оригінальних даних або довірчі інтервали дуже широкі, це може вказувати на невпевненість прогнозу або складність в прогнозуванні даного часового ряду (Рис 3.20).

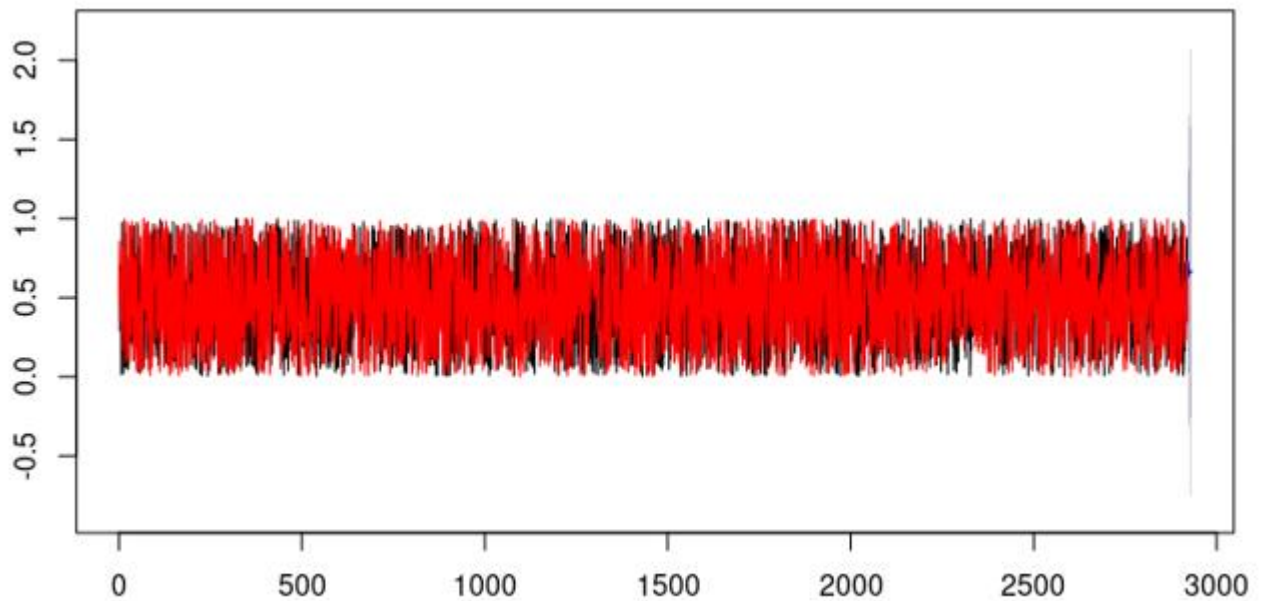


Рисунок 3.20 – Прогнозування Arima

Джерело: розроблено автором

Графік демонструє, як прогнозовані значення (сині лінії) відтворюють поведінку оригінальних значень (червоні лінії). Якщо прогнози близькі до оригінальних значень і попадають у інтервал прогнозу, це свідчить про те, що модель успішно узагальнює дані і може бути використана для прогнозу майбутніх значень.

Зіставлення прогнозованих значень з оригінальними дозволяє оцінити точність моделі. Якщо прогнози відхиляються від оригінальних значень або виходять за межі інтервалу прогнозу, це може свідчити про те, що модель не здатна точно передбачити майбутні зміни часового ряду.

Аналізуючи графік, можна оцінити якість прогнозів і виявити потребу в поліпшенні моделі, якщо відхилення прогнозів є значними.

Аналізуючи графік залишків, можна зробити висновки про адекватність моделі і виявити потенційні проблеми, такі як несподівані патерни або неточності у прогнозуванні. Це допомагає поліпшити модель і зробити більш точні прогнози(Рис 3.21).

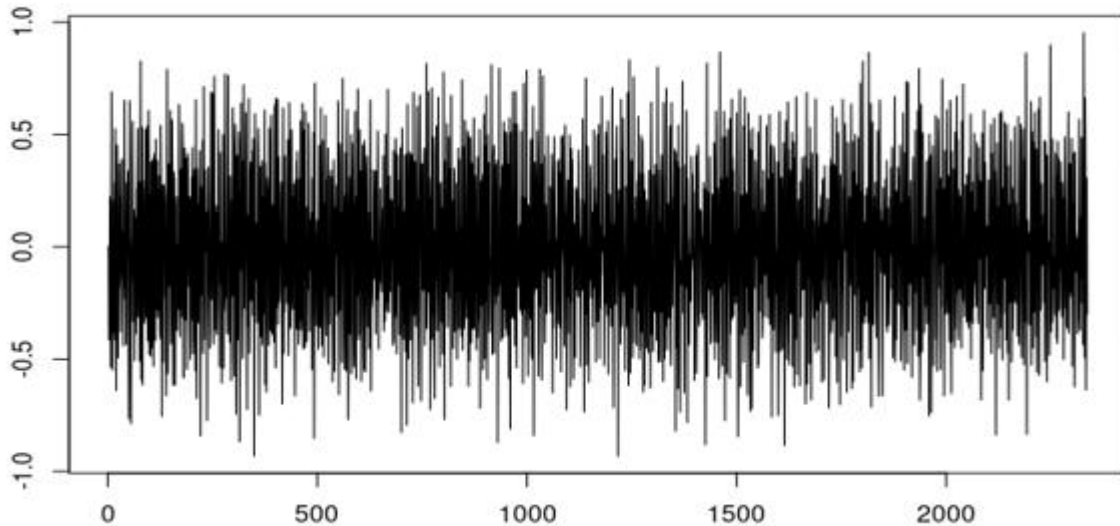


Рисунок 3.21 – Графік залишків

Джерело: розроблено автором

Основні різновидності поведінки залишків, які можна спостерігати на графіку, включають:

Випадкове розподілення: Залишки розподілені випадковим чином навколо нуля без будь-якого видимого закономірності або тренду. Це вказує на те, що модель добре узагальнює дані і залишки не мають систематичних структур.

Систематичні закономірності: Залишки можуть показувати систематичні залежності, такі як тренди, циклічність або сезонність. Це може свідчити про те, що модель не враховує повністю всі важливі фактори часового ряду.

Нетиповічні аномалії: Графік залишків може також виявити нетиповічні аномалії, які виходять за межі звичайних коливань. Ці аномалії можуть бути результатом систематичних помилок у моделі або впливу неврахованих факторів.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2921	0.7191496	0.26484902	1.173450	0.02435682	1.413942
2922	0.6345658	0.12876727	1.140364	-0.13898631	1.408118
2923	0.6776920	0.06780644	1.287578	-0.25504749	1.610432
2924	0.6556162	-0.01560558	1.326838	-0.37092891	1.682161
2925	0.6668291	-0.07301856	1.406677	-0.46467025	1.798328
2926	0.6610465	-0.13560334	1.457696	-0.55732428	1.879417
2927	0.6639409	-0.18855300	1.516435	-0.63983602	1.967718
2928	0.6624053	-0.24112364	1.565934	-0.71942296	2.044234
2929	0.6631314	-0.28936173	1.615625	-0.79358116	2.119844
2930	0.6627028	-0.33603513	1.661441	-0.86473507	2.190141

Рисунок 3.22 - Результати прогнозу

Джерело: розроблено автором

Результатом прогнозу моделі є дані ARIMA для майбутніх точок у часовому ряді. Кожна точка прогнозу має відповідні інтервали надійності (80% та 95%), що вказують на діапазон можливих значень для кожної точки прогнозу.

Point Forecast: прогнозоване значення для кожної точки у часовому ряді.

Lo 80 та Hi 80: колонки нижню та верхню межі 80% інтервалу надійності для прогнозованої точки. Значення всередині цього діапазону вважаються ймовірними для прогнозу.

Lo 95 та Hi 95: колонки нижню та верхню межі 95% інтервалу надійності для прогнозованої точки. Цей діапазон ширший і включає більше можливих значень для прогнозу.

Зазначені інтервали надійності важливі при оцінці точності прогнозу. Вони вказують на рівень впевненості моделі щодо прогнозування майбутніх значень. Чим ширший інтервал надійності, тим менш точний прогноз може бути.

Для першої точки прогнозу (2921) ми маємо прогнозоване значення 0.7191496. Інтервал надійності 80% для цієї точки знаходиться між 0.26484902 та 1.173450, а для інтервалу надійності 95% - між 0.02435682 та 1.413942. Ці значення вказують на діапазон можливих значень для першої точки прогнозу з відповідним рівнем впевненості моделі. Аналогічно, цей аналіз можна провести для усіх інших точок прогнозу.

ВИСНОВКИ

У роботі проаналізовано структуру та тенденцію розвитку ринку електронної торгівлі в Україні. Проведений аналіз та моделювання економічних показників дозволили отримати важливі дані щодо структури ринку та спостережуваних тенденцій. Вивчений досвід застосування інструментарію математичного моделювання дозволив аналізувати процеси формування ринку та робити прогнози щодо його подальшого розвитку. Обрані алгоритми інтелектуального аналізу даних виявилися ефективними для аналізу та прогнозування розвитку електронної комерції.

У роботі вивчено та узагальнено досвід застосування інструментарію математичного моделювання для аналізу процесів формування ринку електронної торгівлі. На підставі проведеного аналізу та побудованих моделей було встановлено, що частка електронної комерції у ВВП, рівень доходів населення, кількість користувачів Інтернету та оборот ринку електронної комерції є важливими показниками для дослідження та прогнозування економічних тенденцій. ARIMA модель, яка була побудована, виявила високу прогностичну точність і може бути використана для прогнозування майбутніх значень. Кластерний аналіз дозволив виділити групи змінних з подібними характеристиками, що сприяє кращому розумінню факторів, що впливають на економічні процеси.

У роботі обґрунтовано вибір алгоритмів аналізу даних, що були застосовані. Отримані результати аналізу та моделювання мають практичне застосування і можуть бути використані для прийняття обґрунтованих рішень в галузі електронної комерції та економіки загалом. Використання такого аналітичного підходу допоможе зробити прогнози щодо розвитку ринку електронної торгівлі, виявити потенційні тенденції та фактори впливу, а також сприятиме розробці стратегій розвитку бізнесу.

У роботі були побудовані моделі прогнозування обсягів електронної торгівлі на базі інструментів аналізу часових рядів, що дозволяє отримати більше

інформації про економічні процеси в галузі електронної комерції. У цілому, проведений аналіз та моделювання надають важливі інструменти для кращого розуміння та прогнозування економічних процесів в галузі електронної комерції. Отримані результати можуть служити основою для подальших досліджень та впровадження ефективних стратегій розвитку електронної комерції в Україні.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. MIKE RENALDI. eBook THE ULTIMATE GUIDE TO INTERNATIONAL ECOMMERCE. URL: https://www.vaimo.com/wp-content/uploads/2022/12/the_ultimate_guide_to_international_ecommerce_.pdf
(Accessed 02.05.2023)
2. Eurostat. E-commerce continues to grow in the EU. 2023. URL: <https://ec.europa.eu/eurostat/web/products-eurostat-news/w/DDN-20230228-2>
(Accessed 02.05.2023)
3. Olga Pyrog. A., H. The Development of E-Commerce in Terms of Virtualization Industry. URL: <https://science.lpnu.ua/eem/all-volumes-and-issues/volume-3-number-1-2016/development-e-commerce-terms-virtualization>
(Accessed 02.05.2023)
4. Aaron O'Neill. Statista. URL: <https://www.statista.com/topics/2473/ukraine/#opicOverview>. (Accessed 02.05.2023)
5. Ben Lutkevich. Basic concepts of electronic commerce. URL: <https://www.techtarget.com/searchcio/definition/e-commerce> (Accessed 02.05.2023)
6. Vukoman R. Jokanović. "A dynamic model for e-commerce" by Kapoor, K., & Garg, R. URL: https://www.researchgate.net/publication/359705344_Artificial_Intelligence
(Accessed 02.05.2023)
7. G. Casella S. Fienberg I. Olkin. "Time Series Analysis and Its Applications: With R Examples" by Robert H. Shumway and David S. Stoffer. URL: <https://mybiostats.files.wordpress.com/2015/03/time-series-analysis-and-its-applications-with-examples-in-r.pdf> (Accessed 02.05.2023)
8. Radeef Chundakkadan. Economic Analysis and Policy (2023) URL: <https://www.sciencedirect.com/science/article/abs/pii/S0313592623000711> (Accessed 02.05.2023)
9. Omkar Dastane, Chai Lee Goi. Journal of Retailing and Consumer Services (2022) URL:

<https://www.sciencedirect.com/science/article/abs/pii/S0969698922003150> (Accessed 02.05.2023)

10. Ligita Gasparėnienė, Rita Remeikienė. Evaluation of the impact of shadow economy determinants: Ukrainian case

URL:<https://www.sciencedirect.com/science/article/pii/S1822801116300303>

(Accessed 02.05.2023)

11. Shuhua Chang , Anqi Li , Xinyu Wang. Optimal combination of platform channel contract and guarantee financing strategy in e-commerce market. URL:<https://www.sciencedirect.com/science/article/abs/pii/S1366554523000820>

0 (Accessed 02.05.2023)

12. Katarzyna Cheba, Maja Kiba-Janiak, Anna Baraniecka, Tomasz Kołakowski. Impact of external factors on e-commerce market in cities and its implications on environment

.URL:<https://www.sciencedirect.com/science/article/abs/pii/S2210670721003164>

(Accessed 02.05.2023)

13. Dayna Winter. Ecommerce Definition - What is Ecommerce. URL: <https://www.shopify.com/encyclopedia/what-is-ecommerce>. (Accessed 02.05.2023)

14. Ben Lutkevich. Chai W. How does e-commerce work?. URL: <https://searchcio.techtarget.com/definition/e-commerce>. (Accessed 02.05.2023)

15. Andrew B. Electronic Commerce (Ecommerce). URL: <https://www.investopedia.com/terms/e/ecommerce.asp>. (Accessed 02.05.2023)

16. BigCommerce.Ecommerce 101 + History of Online Shopping (2021). BigCommerce. URL: <https://www.bigcommerce.com/articles/ecommerce/#conclusion>. (Accessed 02.05.2023)

17. J. Phys.: Conf. Ser.The application of ARIMA model in forecasting population data. URL: <https://iopscience.iop.org/article/10.1088/1742-6596/1324/1/012100/pdf> (Accessed 02.05.2023)

18. Egor Howell. Time Series Decomposition. URL: <https://towardsdatascience.com/time-series-decomposition-8f39432f78f9> (Accessed 02.05.2023)

19. Lone, S., & Weltevreden, J.W.J..European E-Commerce Report. URL: https://ecommerce-europe.eu/wp-content/uploads/2022/06/CMi2022_FullVersion_LIGHT_v2.pdf (Accessed 02.05.2023)
20. Iryna Huryk. AN ANALYSIS OF E-COMMERCE MARKET IN UKRAINE. URL: https://kse.ua/wp-content/uploads/2021/04/Huryk_Master_Thesis_November_2020.pdf (Accessed 02.05.2023)
21. Andin Bicknell. How E-Commerce Security Makes Your Business Unshakable. URL: <https://www.globalsign.com/en/blog/e-commerce-security-importance-threats> (Accessed 02.05.2023)
22. Kinga Edwards. Ukrainian ecommerce overview. URL: <https://ecommercegermany.com/blog/european-ecommerce-overview-ukraine> (Accessed 02.05.2023)
23. Slavoljub Milovanovic, T J. The Role of Electronic Business in the Reengineering and Integration of Business Processes. URL: https://www.researchgate.net/publication/322720225_The_Role_of_Electronic_Business_in_the_Reengineering_and_Integration_of_Business_Processes (Accessed 02.05.2023)
24. Stefano LEUCCI, Robert RIEMANN. Card-based payments - European Data Protection Supervisor. URL: https://edps.europa.eu/system/files/2022-01/td_cardpayments2_en_0.pdf (Accessed 02.05.2023)
25. ADAM HAYES. Visa vs. Mastercard: What's the Difference? URL: <https://www.investopedia.com/articles/personal-finance/020215/visa-vs-mastercard-there-difference.asp> (Accessed 02.05.2023)
26. JAKE FRANKENFIELD. What is Online Banking? Definition and How It Works. URL: <https://www.investopedia.com/terms/o/onlinebanking.asp> (Accessed 02.05.2023)
27. Hakim Fourar-Laidi. A smart card based framework for securing e-business transactions in distributed systems. URL: <https://www.sciencedirect.com/science/article/pii/S1319157812000171> (Accessed

02.05.2023)

28. Georgios Papadopoulos. Electronic Money and the Possibility of a Cashless Society. URL:

https://www.researchgate.net/publication/228280500_Electronic_Money_and_the_Possibility_of_a_Cashless_Society (Accessed 02.05.2023)

29. Xunyi Ren, L.W.,J. Z., X. M. The Improvement of SET Protocol based on Security Mobile Payment. URL:

https://www.researchgate.net/publication/269779154_The_Improvement_of_SET_Protocol_based_on_Security_Mobile_Payment (Accessed 02.05.2023)

30. Cyrille Artho, Armin Biere. Combined Static and Dynamic Analysis. URL:

<https://www.sciencedirect.com/science/article/pii/S1571066105002537> (Accessed 02.05.2023)

31. Tableau, A Salesforce Company. Time Series Analysis: Definition, Types, Techniques, and When It's Used. URL: <https://www.tableau.com/learn/articles/time-series-analysis> (Accessed 02.05.2023)

32. John C. Chambers, Satinder K. Mullick, and Donald D. Smith. How to Choose the Right Forecasting Technique. URL: <https://hbr.org/1971/07/how-to-choose-the-right-forecasting-technique> (Accessed 02.05.2023)

33. Leonie Monigatti .Stationarity in Time Series — A Comprehensive Guide. URL: <https://towardsdatascience.com/stationarity-in-time-series-a-comprehensive-guide-8beabe20d68> (Accessed 02.05.2023)

34. Vishal Sharma, Namrata Sharma. Autocorrelation. URL: <https://www.sciencedirect.com/topics/earth-and-planetary-sciences/autocorrelation> (Accessed 02.05.2023)

35. Parveen Ali, Ahtisham Younas..Understanding and interpreting regression analysis. URL: <https://ebn.bmj.com/content/24/4/116> (Accessed 02.05.2023)

36. K. Keerthi Vasan, B. Surendiran.Dimensionality reduction using Principal Component Analysis for network intrusion detection. URL: <https://www.sciencedirect.com/science/article/pii/S2213020916301446>(Accessed 02.05.2023)

37. Qualtrics. What is cluster analysis and when should you use it? URL: <https://www.qualtrics.com/uk/experience-management/research/cluster-analysis/> (Accessed 02.05.2023)

ДОДАТКИ

Додаток А:

Додаток А.1:

Побудова моделі лінійної регресії

```
# Побудова моделі лінійної регресії
model <- lm(days_values ~ days1 + days2 + days3 + days4 + days5, data = data)

# Виведення результатів моделі
summary(model)
```

Додаток А.2:

Перевірка значимості коефіцієнтів

```
# Перевірка значимості коефіцієнтів. Результати покажуть оцінки коефіцієнтів, стандартні помилки, t-значення та
coef_test <- coef(summary(model))
coef_test
```

Додаток А.3:

Перевірка мультиколінеарності

```
# Перевірка мультиколінеарності
vif_values <- car::vif(model)

# Виведення результатів
vif_values
```

Додаток А.4:

Стандартизація допомагає привести змінні до однієї шкали та полегшує інтерпретацію результатів PCA

```
# Стандартизація допомагає привести змінні до однієї шкали та полегшує інтерпретацію результатів PCA.
scaled_data <- scale(data[, c("Y", "X1", "X2", "X3", "X4", "X5", "days", "days1", "days2", "days3", "days4", "days5")

# Виведення стандартизованих даних
scaled_data

# Обчислення основних компонент
pca_result <- prcomp(scaled_data, center = TRUE, scale. = TRUE)

# Виведення результатів, Аргументи center = TRUE та scale. = TRUE вказують на центрування та масштабування даних
summary(pca_result)

# Графік власних значень, який показує відсоток дисперсії, поясненої кожною головною компонентою
plot(pca_result, type = "l", main = "Scree Plot")
```

Додаток А.5:

Графік би-плота

```
# Графік би-плота
biplot(pca_result, scale = 0)

# Додавання назв точок
text(pca_result$rotation[, 1], pca_result$rotation[, 2], labels = colnames(scaled_data),
     cex = 0.7, pos = 4)
```

Додаток А.6:

Кластеризація методом k-середніх (k-means)

```
# Кластеризація методом k-середніх (k-means):
set.seed(123) # Задаем случайное начальное значение для воспроизводимости
kmeans_clusters <- kmeans(data[, c("days", "days1", "days2", "days3", "days4", "days5")], centers = 3) #Аргумент cen

# Кластеризація методом ієрархічної кластеризації
set.seed(123) # Задаем случайное начальное значение для воспроизводимости
hclust_clusters <- hclust(dist(data[, c("days", "days1", "days2", "days3", "days4", "days5")])) #Ми обчислюємо відст
```

Додаток А.7:

Візуалізація кластеризації

```
# Візуалізація кластеризації
set.seed(123) # Задаем случайное начальное значение для воспроизводимости
kmeans_clusters <- kmeans(data[, c("days", "days1", "days2", "days3", "days4", "days5")], centers = 3)
clusplot(data[, c("days", "days1", "days2", "days3", "days4", "days5")], kmeans_clusters$cluster, color = TRUE, shac
```

Додаток Б:

Додаток Б.1

Графики

```
# График
ggplot(data, aes(x = year, y = days)) +
  geom_line() +
  labs(x = "Рік", y = "ВВП, млрд. євро", title = "Графік ВВП, млрд. євро")
#Графік часового ряду
#Цей графік показує динаміку частки електронної комерції у ВВП за роками. Ось x репрезентує рік, а вісь y - частку
ggplot(data, aes(x = year, y = days)) +
  geom_line() +
  labs(x = "Рік", y = "Частка електронної комерції у ВВП, %", title = "Динамика Частки електронної комерції у ВВП")
#Діаграма розсіювання:
# Ця діаграма розсіювання показує зв'язок між ВВП та частиною електронної комерції у ВВП. Вісь x становить ВВП у м
ggplot(data, aes(x = days, y = days1)) +
  geom_point() +
  labs(x = "ВВП, млрд. євро", y = "Частка електронної комерції у ВВП, %", title = "Связь между ВВП и Часткой елект")

days_values_x4 <- runif(length(days_values_x4), min = 1, max = 1000)
days_values_x5 <- runif(length(days_values_x4), min = 1, max = 100)
# Гістограма:
#Ця гістограма показує розподіл кількості користувачів Інтернету. Вісь x представляє кількість користувачів інтерн
ggplot(data, aes(x = days3)) +
  geom_histogram(binwidth = 50000, boundary = 0) +
  labs(x = "Кількість користувачів інтернет, тис. осіб", y = "Частота", title = "Розподіл Кількості користувачів і")

# Ящик з вусами (boxplot):
ggplot(data, aes(x = factor(month), y = days3)) +
  geom_boxplot() +
  labs(x = "Місяць", y = "Значення Кількість користувачів інтернет, тис. осіб", title = "Кількість користувачів ін")
```

Додаток В:

Додаток В.1:

Arima

```
# Побудова ARIMA моделі
arima_model <- forecast::auto.arima(ts_data)
summary(arima_model)

# Задати розмір тестового набору (наприклад, 20% від загального обсягу даних)
test_size <- 0.2

# Визначити індекс, що розділяє навчальний та тестовий набори
split_index <- floor((1 - test_size) * length(ts_data))

# Розділити дані на навчальний та тестовий набори
train_data <- ts_data[1:split_index]
test_data <- ts_data[(split_index + 1):length(ts_data)]

# Побудувати ARIMA-модель на навчальному наборі даних
model <- arima(train_data, order = c(1, 1, 0), include.mean = TRUE)
model

# Здійснити прогноз на тестовому наборі даних
forecast <- predict(model, n.ahead = length(test_data))
forecast
# Обчислити помилки прогнозу
errors <- test_data - forecast$pred

# Обчислити метрики оцінки точності прогнозу
accuracy_measures <- forecast::accuracy(forecast$pred, test_data)

# Вивести результати оцінки точності прогнозу
print(accuracy_measures)

plot(test_data, type = "l", col = "black", xlab = "Час", ylab = "Значення", main = "Прогноз ARIMA")
# Додавання лінії з прогнозом
lines(forecast$mean, col = "red")
# Додавання легенди
legend("topleft", legend = c("Фактичні значення", "Прогноз"), col = c("black", "red"), lty = 1)

# Побудова графіка залишків
residuals <- residuals(model)
plot(residuals, ylab = "Залишки")

# Побудова графіка
plot(forecast_result, main = "ARIMA Forecast", xlab = "Time", ylab = "Values")
lines(forecast_result$mean, col = "blue") # Графік прогнозованих значень
lines(ts_data, col = "red") # Графік оригінальних значень

# Виведення прогнозних результатів
print(forecast_result)
```


Форма короткого звіту подібності
за антиплагіатною інтернет-системою Unicheck

Короткий звіт за результатами перевірки кваліфікаційної роботи
антиплагіатною інтернет-системою Unicheck



Ім'я користувача: Математичного моделювання та статистики Даценк...	ID перевірки: 1015415490
Дата перевірки: 04.06.2023 19:42:14 EEST	Тип перевірки: Doc vs Internet + Library
Дата звіту: 04.06.2023 20:14:19 EEST	ID користувача: 100005744

Назва документа: Мисько_Катутіна_плагіат

Кількість сторінок: 59 Кількість слів: 10183 Кількість символів: 79825 Розмір файлу: 1.17 MB ID файлу: 1015078365

12%
Схожість

Найбільша схожість: 2.34% з Інтернет-джерелом (<https://moodle.znu.edu.ua/mod/resource/view.php?id=124075>)

7.65% Джерела з Інтернету	175	Сторінка 61
6.36% Джерела з Бібліотеки	395	Сторінка 62

0% Цитат

Вилучення цитат вимкнено

Вилучення списку бібліографічних посилань вимкнено

0%
Вилучень

Немає вилучених джерел

Модифікації

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Замінені символи	10
------------------	----