

**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

Галузь знань

Спеціальність

Економічна кібернетика та Дата Сайнс

05 соціальні та поведінкові науки

051 економіка

Форма навчання: очна (денна)

КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА

на тему «Аналіз та моделювання ринку праці України»

здобувача Горай Руслана Сергійовича _____

(підпис)

Науковий керівник: кандидат економічних наук,
доцент Катуніна Ольга Сергіївна _____

(підпис)

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

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

(підпис)

Київ 2023

ЗМІСТ

ВСТУП.....	3
РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ РИНКУ ПРАЦІ УКРАЇНИ.....	7
1.1 Сутність, структура та особливості ринку праці.....	7
1.2 Напрями дослідження динаміки зайнятості та плинності кадрів на підприємствах.....	15
1.3 Методологічні підходи до дослідження та прогнозування динаміки ринку праці на базі математичного моделювання.....	21
РОЗДІЛ 2. МАТЕМАТИЧНІ МОДЕЛІ АНАЛІЗУ ДИНАМІЧНИХ ПРОЦЕСІВ НА РИНКУ ПРАЦІ.....	29
2.1 Інструментарій розвідувального аналізу часових рядів факторів ринку праці.....	29
2.2 Моделювання динаміки ринку праці України та зарубіжних країн.....	38
2.3 Особливості використання методів машинного навчання для аналізу та моделювання плинності працівників.....	50
РОЗДІЛ 3. ПРАКТИЧНЕ ЗАСТОСУВАННЯ МОДЕЛЕЙ АНАЛІЗУ ДИНАМІКИ ТА ПРОГНОЗУВАННЯ РИНКУ ПРАЦІ УКРАЇНИ.....	58
3.1 Аналіз результатів моделювання ринку праці України на базі одновимірних часових рядів.....	58
3.2 Аналіз результатів багатofакторного моделювання динаміки ринку праці України.....	68
3.3 Аналіз результатів моделювання динаміки плинності працівників на базі методів машинного навчання.....	75
ВИСНОВКИ.....	84
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	88
ДОДАТКИ.....	94

ВСТУП

Ринок праці є одним з найважливіших елементів національної економіки. Через негативні тенденції на ринку праці виникають серйозні економічна та соціальні проблеми, що загрожує соціально-економічному розвитку держави. Український ринок праці характеризується значними негативними тенденціями у безробітті та зайнятості, а також у попиті та пропозиції робочої сили. Тому дослідження та моделювання причин виникнення негативних явищ на ринку праці є актуальною темою дослідження.

Сучасний стан ринку праці можна охарактеризувати як нестабільний. На жаль початок повномасштабного вторгнення приніс Україні великі проблеми в усіх сферах життя. Ринок праці України зазнав важких втрат попиту та пропозиції робочої сили, що вплинуло на загострення проблеми зростання безробіття та зниження рівня економічного активного населення. Україна потребує трансформації та розвитку ринку праці. За умов, які виникли, набуває актуальності дослідження та прогнозування тенденцій на ринку праці для розробки відповідних заходів покращення.

Чисельні праці українських вчених присвячуються дослідженням ринку праці, ключових тенденцій та динаміки, а також моделювання та прогнозування основних показників ринку праці. Дану тематику досліджували такі вчені: А. Колот, Ю. Коновалов, О. Кваша, А. Калина, О. Грішнова, О. Герасименко, І. Лук'яненко, М. Оліскевич, А. Камінський, М. Скрипниченко, І. Андрейцева та ін. У своїх роботах вчені описували теоретичні засади формування ринку праці України, його ключові особливості функціонування та структуру; аналізували динаміку та тенденції основних кількісних показників ринку праці та порівнювали їх з іншими країнами; описували методи статистичного аналізу та моделювання різноманітних процесів на ринку праці. Науковці виділяють такі ключові проблеми

функціонування ринку праці України як: нерівномірність попиту та пропозиції; низька продуктивність та якість праці, що виникають через низький дохід, погані умови праці, відсутність заохочень та навчання; розширення неформальної зайнятості, яку викликає не контрольованість ринку праці. Важливими напрямками дослідження ринку праці є аналіз та моделювання тенденцій основних індикаторів ринку праці таких як: рівень безробіття, зайнятості, економічно активного населення, дохідності. Окрім цього учені досліджують залежність між безробіттям, зайнятістю та іншими економічними факторами такими як: рівень інфляції, прямі іноземні інвестиції, рівень заробітної плати, курс валют, та інші.

Дослідження ринку праці України, в сучасних умовах нестабільності та структурних зрушень є важливою проблемою. Після початку повномасштабного вторгнення проблеми із зайнятістю та безробіттям, з рівновагою попиту та пропозиції лише загострилися. Тому дослідження стану ринку праці України після початку повномасштабного вторгнення є та буде ключовою темою наукових досліджень багатьох фахівців. Поглиблення наукових досліджень в даному напрямку дасть змогу доповнити аналіз саме сучасного стану ринку праці України та допоможе точніше формувати державні заходи, що будуть спрямовані на подальший розвиток ринку праці.

Метою кваліфікаційної роботи є аналіз сучасного стану ринку праці України та розробка моделей прогнозування зайнятості населення та плинності кадрів на підприємстві.

Для досягнення поставленої мети потрібно вирішити такі завдання:

- 1) дослідити сутність ринку праці та особливості його функціонування;
- 2) проаналізувати проблему плинності кадрів на підприємстві;
- 3) оцінити стан ринку праці України та зарубіжних країн;
- 4) побудувати комплекс математичних моделей прогнозування рівня зайнятості в Україні та плинності кадрів на підприємстві на основі визначених методів;

5) оцінити, проаналізувати, порівняти результати побудованих моделей та надати рекомендації щодо їх використання.

Об'єктом дослідження є ринок праці України.

Предметом дослідження є математичні моделі та методи аналізу та прогнозування динамічних процесів на ринку праці України.

Теоретичну основу дослідження складають фундаментальні положення про ринок праці України. Застосовуються загальнонаукові та спеціальні методи дослідження. Метод системного та просторового підходу застосовувався для дослідження об'єкта кваліфікаційної роботи з різних сторін. Метод аналізу та синтезу застосовувався для детального розгляду динамічних процесів ринку праці та плинності кадрів на підприємстві. Методи логічного узагальнення, дедукції та індукції для отримання на основі фактів, загальних висновків до розділів. Крім того, використано спеціальні методи а саме: класичні та векторні авторегресійні моделі, експоненціальне згладжування та метод Prophet для моделювання та прогнозування рівня зайнятості; метод наївного Байеса, k-найближчих сусідів, випадковий ліс для моделювання та прогнозування рівня плинності кадрів на підприємстві.

Робота має теоретичну, методичну та практичну значущість. У роботі детально описано структуру ринку праці та сутність плинності кадрів на підприємстві, також застосовано методи машинного навчання для моделювання рівня зайнятості та плинності кадрів, це відображає теоретичну значущість. Методи та підходи, що використовуються в роботі допоможуть у подальших дослідженнях ринку праці та плинності кадрів, це відображає методичну значущість. В роботі розроблено моделі прогнозування рівня зайнятості, а також автоматизовано процес прогнозування плинності кадрів, що допоможе різним зацікавленим особам покращити роботу підприємства, що є відображенням практичної значущості роботи.

Інформаційною базою дослідження є законодавство України, праці вітчизняних та іноземних науковців, статистичні дані державних органів України та спеціалізованих інтернет ресурсів, науково-методична література.

Кваліфікаційна робота має таку структуру: вступ, три розділи, список використаних джерел та додатки.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ РИНКУ ПРАЦІ УКРАЇНИ

1.1 Сутність, структура та особливості ринку праці

Ринок праці є важливою складовою економічної системи, що характеризує майже всі соціально-економічні процеси, які відбуваються в країні. У процесі діяльності робітники та роботодавці визначають вартість праці, умови функціонування, обов'язки та права [1].

Ринок праці, як і будь-який інший ринок, є механізмом взаємодії покупців та продавців товару чи послуги. Відповідно відносини між продавцем та покупцем регулюються законами ринкової економіки. Але, ринок праці має свою специфіку та відмінності від інших видів ринку. Так охарактеризувала ринок праці у своїй науковій статті О. С. Кваша [1]. Також вона відмічала важливість ринку праці, як фактор впливу на функціонування національної економіки.

Для опису сутності ринку, потрібно визначити саме поняття «ринок». Існує багато напрямів щодо опису сутності терміну ринок. О. Курно, А. Маршал, Дж. М. Кейнс визначали ринок як «будь-який район, де економічні зв'язки покупців і продавців до такого рівня вільні, що ціни на однакові товари досить легко і швидко вирівнюються». Представники новоавстрійської школи, керувалися положеннями А. Сміта, та визначали ринок як «природний процес, породжений діями індивідуумів, які намагаються досягти своєї мети і із різноманітності їх дій формується економічний порядок» [2]. За неокласичним напрямом ринкової теорії – ринок є «сукупність продавців і покупців, у результаті дії яких об'єкти обміну отримують ціну». За марксистською теорією, ринок це «відносини між продавцями та покупцями у сфері обігу за властивими йому законами і формами організації» За

концепцією уряду України ринок це «з механізмом вільного ціноутворення, з вільним підприємництвом на основі економічної самостійності, рівноправності та конкуренції суб'єктів господарювання у боротьбі за споживача» [2].

Поняття «ринок праці» розглядається вченими та науковцями з різних сторін. За одним з підходів ринок праці є сукупністю відносин між людьми щодо відтворення робочої сили. Цей підхід базується на теорії К. Маркса про зв'язок між виробництвом та рівнем розвитку суспільства. Роль виробництва за цим методом є ключовим у процесі відтворення трудової вартості. Згідно економічної теорії західних науковців ринок праці є механізмом відтворення трудового потенціалу, розвитку людського капіталу [3].

У праці [2] Колот А. М. відзначав, що людина та праця є основою організації економіки держави. Праця є фактором суспільного прогресу, соціально-економічного розвитку та умовою задоволення власних потреб людини. Формування і розвиток ринку праці є орієнтиром кожної соціальної держави, а рівновага попиту та пропозиції на ринку говорить про стабільність економіки. Колот А.М. ще раз підкреслив важливість ринку праці для формування усієї економічної системи держави.

Відповідно до Закону України «Про зайнятість населення» «ринок праці - система правових, соціально-трудова, економічних та організаційних відносин, що виникають між особами, які шукають роботу, працівниками, професійними спілками, роботодавцями та їх організаціями, органами державної влади у сфері задоволення потреби працівників у зайнятості, а роботодавців - у найманні працівників відповідно до законодавства» [4].

Об'єктом купівлі-продажу на ринку праці є послуги праці. Вони є ключовим елементом функціонування ринку. Важливо, що реальний вклад працівника у виробництво продукції, а не потенційні здібності працівника до праці, оцінюються ринком праці. Роботодавець при укладенні трудової угоди звертає увагу на професійні, а також моральні якості людей. Часто в сучасному світі, особисті якості

людини переважають над його професійними навичками. Проте, зрозуміло, що моральні якості не є предметом ринкового обміну, а лише можуть впливати на нього [5].

Суб'єктами ринку праці — є учасники процесу праці. До них відносяться:

- працівники, які є продавцем послуг праці;
- підприємці, що є власниками капіталу і покупцями послуг праці;
- різноманітні організації найманих працівників;
- держава, що займається регулюванням ринку [5].

Ринок праці має виконувати такі функції:

- розподільчу – виокремлює працівника та роботодавця, розподіляє працівників за галуззю, регіоном, професією, кваліфікацією, тощо;
- інформаційну – надає учасникам ринку інформацію, щодо умов найму на посаду, рівня заробітної плати, якості робочої сили, пропозиції робочих місць;
- посередницьку – допомагає встановити взаємозв'язок на ринку між роботодавцями та найманими працівниками, які функціонують на ринку;
- ціноутворювальну – встановлює рівновагу між попитом та пропозицією робочої сили та є однією з найважливіших функцій;
- стимулюючу – конкуренція на ринку праці є стимулом для покращення професійних та моральних якостей працівника, а також заохочує роботодавців до ефективного використання трудових ресурсів;
- регулюючу – сприяє розподілу робочої сили по галузям, регіонам для регулювання ситуації на ринку
- оздоровчу – стимулює в умовах конкуренції підприємства розвиватися, щоб не бути нежиттєздатним, а працівників мати

найкращі показники якості робочої сили для отримання або утримання робочого місця [3].

Для оцінки стану ринку праці в державі існують спеціальні державні органи. В Україні збором даних та обчисленням ключових показників оцінки ринку праці займається Державна служба статистики. Вони обчислюють показники за методикою МОП за допомогою проведення вибірових досліджень в державі. За стандартами МОП населення віком 15-70 років ділиться на економічно активне та економічно неактивне. Економічно активне населення(робоча сила) за методикою МОП — це «населення обох статей віком від 15 до 70 років включно, яке впродовж певного періоду часу забезпечує пропозицію робочої сили для ринкового виробництва товарів та послуг, а також деяких видів неринкового виробництва, що входять до Системи національних рахунків і є економічною діяльністю»[6]. До складу ЕАН(економічно активне населення) відносяться зайняті та безробітні. Економічно неактивне населення (населення поза робочою силою) за методикою МОП – це «особи, які впродовж періоду спостереження не входять до складу робочої сили, тобто не є зайнятими або безробітними (у визначенні МОП) та існують за рахунок утримання сім'єю та суспільством»[6].

Зайнятість населення - це результат діяльності ринку праці, котрий означає, що потреба певної частини економічно активного населення була задоволена. Зайнятість населення це важливий фактор створення ВВП, оцінки добробуту населення та реалізації їх потреб. Забезпечення населення місцем праці, тобто регулювання зайнятості, є одним з головних пріоритетів держави [1].

Відповідно до Закону України «Про зайнятість населення» «зайнятість - це діяльність громадян, пов'язана із задоволенням особистих та суспільних потреб і така, що, як правило, приносить їм доход у грошовій або іншій формі [4]».

В Україні до зайнятого населення належать громадяни, що проживають на території держави на законних підставах:

- працюючі по найму на умовах повного або неповного робочого дня (тижня) на підприємствах, в установах і організаціях, незалежно від форм власності, у міжнародних та іноземних організаціях в Україні і за кордоном у фізичних осіб;
- громадяни, які самостійно забезпечують себе роботою, включаючи підприємців, осіб, зайнятих індивідуальною трудовою діяльністю, творчою діяльністю, члени кооперативів, фермери та члени їх сімей, що беруть участь у виробництві, а також члени особистих селянських господарств, діяльність яких здійснюється відповідно до Закону України "Про особисте селянське господарство";
- обрані, призначені або затверджені на оплачувану посаду в органах державної влади, управління та громадських об'єднаннях;
- які проходять службу в Збройних Силах України, Службі безпеки України, Державній прикордонній службі України, військах внутрішньої та конвойної охорони і Цивільної оборони України, органах внутрішніх справ України, інших військових формуваннях, створених відповідно до законодавства України, Державній службі спеціального зв'язку та захисту інформації України, альтернативну (невійськову) службу;
- які проходять професійну підготовку, перепідготовку і підвищення кваліфікації з відривом від виробництва; навчаються в денних загальноосвітніх школах і вищих навчальних закладах;
- працюючі громадяни інших країн, які тимчасово перебувають в Україні і виконують функції, не пов'язані із забезпеченням діяльності посольств і місій [4].

Зайнятість населення має різноманітні види та форми. Види зайнятості це розподіл населення, що зайняте за способом, характером та сферами використання праці. Форми зайнятості класифікуються за такими ознаками:

- за формами організації робочого часу: повна та неповна. Повна зайнятість — це зайнятість протягом повного робочого дня (тижня, сезону, року). Неповна (часткова) — зайнятість протягом неповного робочого часу, чи з неповною оплатою.
- За статусом трудової діяльності: первинна (основна) та вторинна (додаткова). Первинна — це зайнятість за основним місцем роботи, вторинна — це додаткова постійна або тимчасова трудова діяльність осіб, які вже мають основну роботу.
- За характером організації робочих місць та робочого часу: стандартна та нестандартна, яка не відповідає стандартним (типовим) правилам і потребує спеціального правового регулювання та організаційно-економічного забезпечення. Стандартна зайнятість характеризується роботою у приміщенні роботодавця, стандартним навантаженням протягом періоду, наявністю стабільного робочого місця, чітко визначеним графіком роботи, законодавчо встановленою тривалістю робочого дня (тижня, року). Відсутність якоїсь з цих ознак можна розглядати як нестандартну форму зайнятості.
- За формами правового регулювання: формальна та неформальна зайнятість. Формальна зайнятість здійснюється на підставі трудових договорів (контрактів) або інших угод. Неформальна зайнятість є незареєстрованою діяльністю, яка здійснюється без укладання трудових договорів чи інших угод [2].

Зайнятість населення це дуже важливий механізм ринку праці та в цілому, економіки. Він може якісно характеризувати соціально-економічний стан суспільства та його розвиток. Існує декілька показників для оцінки зайнятості, основні з них це:

1. Чисельність зайнятих — абсолютний показник зайнятості населення економічною діяльністю.

2. Рівень зайнятості (P_z) — відношення чисельності зайнятих до чисельності населення:

$$P_z = \frac{З}{Н} * 100, \quad (1.1)$$

де $З$ — чисельність зайнятого населення;

$Н$ — загальна чисельність населення.

3. Частка зайнятих в економічно активному населенні ($Ч_z$):

$$Ч_z = \frac{З}{З+Б} * 100, \quad (1.2)$$

де $Б$ — чисельність безробітних;

$З$ — чисельність зайнятих.

Ще однією складовою ринку праці є безробіття. Відповідно до Закону України «Про зайнятість населення» «Безробітними визнаються працездатні громадяни працездатного віку, які через відсутність роботи не мають заробітку або інших передбачених законодавством доходів і зареєстровані у державній службі зайнятості як такі, що шукають роботу, готові та здатні приступити до підходящої роботи. Безробітними визнаються також інваліди, які не досягли пенсійного віку, не працюють та зареєстровані як такі, що шукають роботу» [4].

У праці [3] А. В. Калина описувала безробіття як процес, що здійснюється через постійні переходи від зайнятості до незайнятості, і навпаки. Вони можуть виникати вимушено або добровільно для працівників. Безробітними вважаються ті, які офіційно зареєстровані в службі зайнятості. Проте, основна частина безробітних частіше за все, не проходить реєстрацію в службі зайнятості.

А.В. Калина детально описувала причини виникнення безробіття [3]. Згідно з класичною теорією, безробіття це наслідок занадто високої заробітної плати щодо такої, яка б вирівнювала попит та пропозицію. За цією теорією, якщо вартість праці

вище рівноважної, то виникає надлишок робочої сили, яку роботодавці не можуть собі найняти за цією ціною.

За кейнсіанською теорією безробіття виникає через те, що номінальна заробітна плата та ціни не встигають пристосуватися до врівноваження попиту та пропозиції на ринку. Стабільність номінальної заробітної плати є запорукою існування довгострокових трудових відносин між роботодавцем та працівником [3].

Безробіття має таку класифікацію форм:

- за причиною виникнення буває сезонне, фрикційне, структурне, циклічне, регіональне, технічне.
- За характером виникнення буває вимушене та добровільне.
- За тривалістю – короткострокове, середньо та довгострокове
- За особливостями вияву може бути відкрите та приховане [3].

Фрикційне безробіття це коли людина добровільно змінює роботу та відповідно тимчасово є незайнятою. Цей вид безробіття існує завжди на ринку праці. Структурне безробіття має тривалий та хронічний характер. Основною причиною цього виду безробіття є кваліфікаційна та територіальна невідповідність між вільними місцями. Тобто, глобальні зміни в структурі попиту на працю є причиною виникнення цього виду безробіття. Технічне безробіття це вид, що пов'язаний зі зміною техніки чи технології на робочому місці, відповідно через це може відбуватися звільнення робочої сили. Регіональне безробіття виникає внаслідок історичних, соціально-психологічних, демографічних подій. Сезонне безробіття виникає в галузях, які мають сезонний характер, що впливає на попит та пропозицію в цих сферах [3].

Циклічне безробіття пов'язане зі спадом попиту на працю, що може виникати внаслідок кризових явищ в економіці країни чи негативних коливань в бізнес середовищі. Циклічне безробіття негативно впливає на розвиток країни [3].

М.О. Оліскевич у роботі [7] проводить детальний економіко-статистичний аналіз динаміку розвитку ринку праці. Вчена аналізує демографічні зміни та її

наслідки для українського ринку праці. Демографічні зміни сильно впливають на ситуацію на ринку праці. Також в даній роботі було детально проаналізована динаміку ключових параметрів ринку праці таких як: рівень зайнятості, безробіття, економічної активності населення. Також вчена порівнює ринок праці України з іншими країнами для виділення ключових тенденцій. В результаті аналізу вчена проводить математичне моделювання різними методами для прогнозування майбутніх тенденцій на ринку праці. Проведення такого глибокого аналізу та моделювання процесів ринку праці дозволяє покращити процес планування заходів спрямованих на покращення ринку праці. Як бачимо, велика кількість українських вчених говорить про актуальність та потрібність проведення детальних досліджень ринку праці України. Проте в сучасних реаліях ринок праці зазнав великих змін, тому варто дослідити та оцінити саме сучасний стан ринку праці України.

1.2 Напрями дослідження динаміки зайнятості та плинності кадрів на підприємствах

Проблема плинності кадрів на підприємстві є актуальною у веденні будь-якого бізнесу. Плинність кадрів або як ще називають плинність персоналу відображає кількість співробітників, що залишили компанію. Персонал є ключовою складовою будь-якого підприємства. Тому важливо оцінювати та аналізувати рівень плинності кадрів на підприємстві. Цей показник дає гарну оцінку якості підбору, навчання та адаптації персоналу на підприємстві. Адже якісний підбір персоналу, а в подальшому навчання та допомога з адаптацією на підприємстві впливає на процес плинності кадрів [8].

Плинність кадрів в основному виникає з таких причин:

- незадоволеність заробітною платою та умовами робочого місця;

- невідповідність виконуваних робіт рівню кваліфікації працівника;
- невизначеність кар'єрного зростання та відсутність його плану;
- напружені стосунки з працівника компанії та керівництвом;
- незадоволеність керівництвом компанії роботою працівника;
- порушення трудової дисципліни в компанії [8].

Шишкін В.О. у роботі [9] виділяв такі чинники виникнення плинності як :

- 1) внутрішні чинники, сюди відносяться причини, що формуються в середині підприємства, тобто оплата праці, умови, відносини в колективі та інші;
- 2) особисті чинники: фізіологічні та психологічні якості, кваліфікація, досвід роботи, вік та стать;
- 3) зовнішні чинники, які не залежать від ситуації всередині підприємства, проте можуть впливати на персонал та підприємство, зокрема це економічна, політична ситуація в країні та інші [9].

Усі визначені вище групи чинників можуть впливати як негативно так і позитивно на рівень плинності кадрів.

Для оцінки плинності кадрів можна розраховувати абсолютну кількість звільнених працівників за певний період або коефіцієнт плинності кадрів. Коефіцієнт плинності кадрів розраховується як співвідношення кількості звільнених працівників до середньої чисельності персоналу за певний період. Формула має наступний вигляд [10]:

$$K_{пл} = \frac{K_z}{Ч_{ср}} * 100, \quad (1.3)$$

де $K_{пл}$ — коефіцієнт плинності;

K_z — кількість звільнених працівників;

$Ч_{ср}$ — середня чисельність працівників за певний період часу.

Загалом даний коефіцієнт можна визначати лише для конкретного відділу або для всієї компанії. Також можна рахувати в розрізі статі працівника чи рахувати лише звільнених працівників за власним бажанням та за порушення трудової дисципліни. В цілому цей процес узгоджується керівництвом кожної компанії.

У роботі [10] Процик І.С. відзначала, що рівень плинності кадрів може бути природнім або надмірним. Природній рівень плинності кадрів коливається в межах 3%-5% на рік від чисельності персоналу. Природна плинність кадрів не вимагає впровадження якихось заходів та може виникати по причині виходу працівника на пенсію або за станом здоров'я чи зміною місця проживання. Якщо рівень плинності кадрів значно перевищує 5%, то це може призвести до збільшення економічних витрат на підприємстві та створення кадрових, технологічних та психологічних проблем у роботі персоналу.

Норма рівня плинності кадрів індивідуальна для кожної сфери підприємства. Наприклад на виробничих підприємствах нормальний рівень плинності становить 10-15%, а для готельно-ресторанного бізнесу рівень 80% вважається в межах норми. Також цей показник може сильно коліватися в залежності від категорії персоналу. Наприклад для топ-менеджерів норма плинності кадрів у межах 0-2%, а для персоналу початкової кваліфікації цей показник може коліватися від 30% до 50% [11]. Тому варто детально досліджувати плинність кадрів на підприємстві в розрізі різних відділів, рівня кваліфікації та інших чинників.

Значний вплив на плинність кадрів мають зовнішні чинники. В Україні після початку повномасштабного вторгнення у 2022 р. ринок праці та економіка зазнала критичних змін. Велика кількість підприємств були просто зруйновані та зупинені, а багатьом іншим підприємствам довелося скоротити велику частку персоналу. Це і є однією з великий причин зростання рівня плинності кадрів на підприємствах України.

Загалом для дослідження плинності кадрів на підприємстві можна використовувати наступний алгоритм:

1. збирання та обробка даних для дослідження шляхом проведення опитувань та тестувань працівників;
2. аналіз основних факторів плинності кадрів на основі зібраних даних;
3. оцінка причин виникнення плинності кадрів на можливість впливу на них керівництвом;
4. розроблення та впровадження заходів для зниження плинності кадрів [9].

Для збирання даних можна використати спеціальні онлайн сервіси для проведення опитувань. Після цього можна переглянути коректність заповнених даних та приступати до аналізу основних причин. Для цього одним з методів можуть бути методи машинного навчання. Наприклад можна використати методи класифікації для передбаченні звільнення працівників в залежності від визначених факторів. Побудована модель допоможе прогнозувати рівень плинності кадрів на майбутні періоди в залежності від визначених факторів.

Пропоную дослідити проблему плинності кадрів на прикладі підприємства бази практики ТОВ «МЕДІА АДВАЙЗЕР». Підприємство є медіа агенцією, яка співпрацює з багатьма українськими та закордонними підприємствами. Рекламний медіаринок України через початок повномасштабного вторгнення у 2022 році зазнав значного скорочення доходів. В перші місяці повномасштабної війни демонстрація прямої телевізійної реклами повністю зупинилась. В цілому абсолютно усі види реклами в більшій або меншій мірі зазнали значного падіння бюджетів у порівнянні з 2021 р. Згідно дослідження Всеукраїнської рекламної коаліції рекламний ринок скоротився у 2,7 рази, тобто втратив близько 20,83 млрд. грн. порівняно з 2021 роком [12]. Таке значне скорочення ринку негативно вплинуло на роботу рекламних агенцій. В результаті загострилася проблеми плинності кадрів.

Згідно даних дослідження Всеукраїнської рекламної коаліції у 76% опитаних медіа агенцій скоротилась кількість персоналу в розмірі від 11 до 70% і

більше від кількості працівників до початку повномасштабного вторгнення (рис 1.1). До війни кількість усіх працівників опитуваних компаній становила 6444 людини, а на липень 2022 р. кількість працівників зменшилася до 4924 людини, тобто 24% працівників втратили своє місце роботи. Таке значне скорочення персоналу значно підвищило рівень плинності кадрів в медіа агенціях. Проте це було вимушеним кроком агенцій через значне зниження дохідності та кількості клієнтів та зниження об'єму роботи за рахунок цього [13].

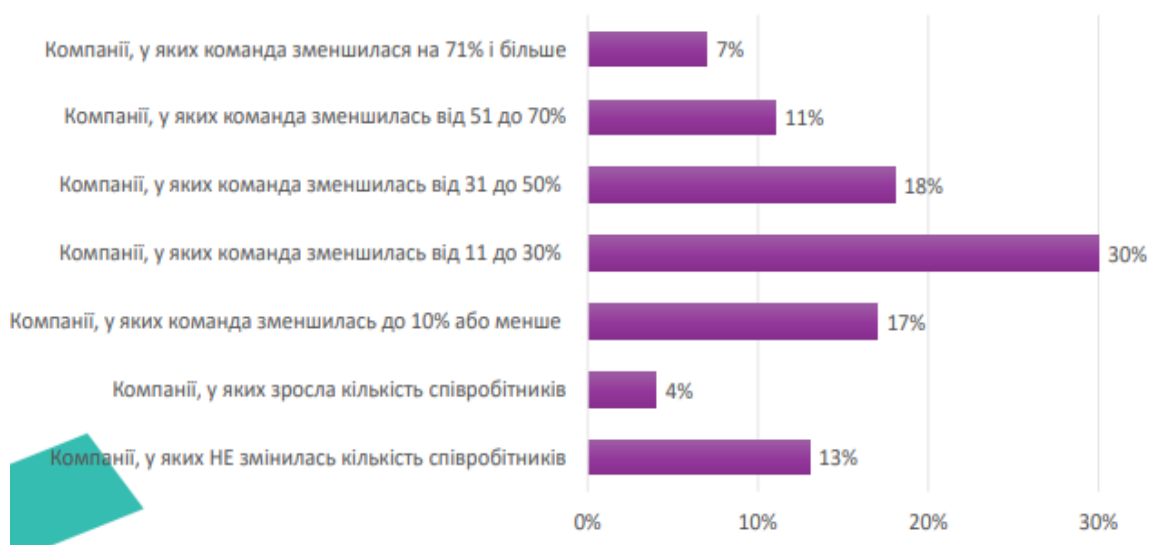


Рисунок 1.1 – Відсоток медіа агенцій, що скоротили персонал після початку повномасштабної війни [13]

Як бачимо зовнішні чинники плинності кадрів дуже сильно можуть вплинути на роботу підприємств та підвищити рівень плинності кадрів. Проте, на даний момент більшість агенцій повернулися до найму персоналу, а за оцінкою Всеукраїнської рекламної коаліції, об'єм ринку рекламних послуг зростає на 30% у порівнянні з 2022 роком та майже досягне рівня 2021р. Для вирішення проблеми плинності кадрів на сучасному підприємстві можна запроваджувати такі заходи:

1. Створення академій для розвитку професійних та особистих якостей.

Для цього можна запустити дистанційну платформу для навчання

працівників. Це допоможе розвивати персонал та вирощувати майбутніх керівників з молодшого персоналу.

2. Надання працівникам додаткових бонусів. Корпоративний психолог, надання корпоративної страховки, знижки для проходження додаткових курсів поза роботою, усі ці заходи позитивно впливають на імідж компанії на ринку праці та задоволеність персоналу місцем роботи.
3. Щорічна оцінка працівників. Дуже важливо проводити щорічну оцінку керівництва, окремих відділів та в цілому працівників. Це дозволить зрозуміти, чи всі задоволені працівники в компанії та наскільки якісно працівники виконують свою роботу [14].

Дані заходив допоможуть покращити умови праці та атмосферу в компанії. Це дозволить дійсно знизити рівень плинності кадрів по внутрішнім причинам. Проте, в умовах загострення зовнішніх чинників таких як війна, досить складно впровадити такі заходи з економічної сторони. Але при стабілізації економічної ситуації впровадження таких заходів покращить роботу компанії та знизить плинність кадрів по причинах незадоволення місця роботи.

Отже, плинність кадрів це важлива проблема для дослідження на підприємствах. В сучасних умовах ця проблема ще більше загострилася. Тому важливо детально досліджувати усі чинники збільшення рівня плинності кадрів та шукати заходи для покращення цієї проблеми. Для цього можна проводити щорічні опитування працівників на задоволеність умов та поточно місця роботи. Також додатково можна побудувати математичну модель для прогнозування плинності кадрів. Впровадження таких методик на підприємстві покращить процес дослідження плинності кадрів, а це в свою чергу дасть змогу провести дієві заходи на підприємстві задля зниження рівня плинності кадрів.

1.3 Методологічні підходи до дослідження та прогнозування динаміки ринку праці на базі математичного моделювання

Моделювання успішно використовуються в прогнозуванні динаміки ринку праці. Для цього застосовуються різноманітні методи математичного моделювання. Зокрема для прогнозування динаміки ринку праці використовують структурні векторні та класичні авторегресійні моделі, регресійні моделі згладженого переходу та сучасні нейронні мережі. Усі ці методи дають змогу проаналізувати динамічні процеси на ринку праці та спрогнозувати майбутні тенденції на базі математичних моделей. Це допомагає точніше та якісніше планувати заходи для покращення стану ринку праці, а також підготовки до майбутніх прогнозованих коливань.

У роботі [15] Єріна А.М відзначала, що суспільно-економічні явища є динамічними і зазнають еволюції протягом певного часу, впливаючи на інші процеси. Закономірності розвитку цих динамічних процесів найкраще виявляються через аналіз часових рядів. У галузі економіки це можуть бути щоденні ціни на акції, обмінні курси, обсяги продажу, рівень зайнятості чи безробіття, інфляція та інші явища. Цим процесам притаманні дві взаємопов'язані характеристики: динамічність та інерційність. Динамічність виявляється у змінах рівнів і варіаціях показників, що описують цей процес, тоді як інерційність виявляється сталістю механізму формування процесу, напрямку та інтенсивності його динаміки протягом конкретного періоду часу.

Отже, послідовність спостережень $y_1, y_2, \dots, y_n$ випадкової величини $\xi(t)$, зібрані в моменти часу $t_1, t_2, \dots, t_n$ називається часовим рядом. Кожен елемент цього ряду є рівнем ряду, а кількість таких рівнів визначає його довжину [16].

Робін Джон Хайдман у своїй роботі [17] відмічав, що моделювання часових рядів є важливим статистичним завданням кожного бізнесу і держави. Воно

допомагає приймати вірні рішення щодо планування виробництва, персоналу на підприємстві, а також планування рівня економіки, зайнятості, безробіття, курсів валют, інфляцій та інших процесів в державі. Також Робін Джон Хайдман пропонує наступну послідовність етапів моделювання:

- 1) Визначення проблеми. Найперший та ключовий етап на якому потрібно визначити ціль моделювання та, який саме результат ми хочемо отримати.
- 2) Збирання даних. Для побудови різноманітних моделей потрібна інформація, тому важливо її якісно зібрати та визначити, які саме дані нам потрібні.
- 3) Дослідницький аналіз. На цьому етапі потрібно проаналізувати закономірності та тенденції, зв'язки в даних.
- 4) Вибір та оптимізація моделей. Важливим етапом моделювання є вибір потрібних методів. На цьому етапі безпосередньо відбувається побудова моделей різноманітними методами.
- 5) Прогнозування та оцінка моделі. Після вибору методів та побудови моделей відбувається оцінка її точності та побудова прогнозу. Оцінити точність та ефективність моделі можна лише після отримання даних за прогнозований період [17].

Для проведення аналізу часового ряду існує метод декомпозиції. Сутність цього методу полягає у виокремленні складових часового ряду, за рахунок чого можна проаналізувати вплив кожної складової на часовий ряд. Після внески кожної складової поєднуються для отримання прогнозів майбутніх періодів часового ряду. Кожен часовий ряд складається з 4 компонент, таких як:

- Тенденція. Тренд або тенденція це складова, що формує загальний напрямок у зміні часового ряду. Вона спостерігається, коли в даних існує тривале збільшення чи зменшення показника

- Сезонність. Сезонність виникає, коли на часовий ряд впливають сезонні фактори.
- Циклічність. Циклічність виникає в даних, що демонструють спад чи зростання, які не мають фіксованої частоти.
- Залишки. Складова частина часового ряду, що залишається після видалення вище зазначених компонент. Ця компонента є результатом впливу випадкових чинників [18].

Для декомпозиції часового ряду в основному використовують адитивну та мультиплікативну форми моделей. Адитивну модель [див. формулу (2.1)] доцільно використовувати тоді, коли часовий ряд має приблизно сталі сезонні коливання, а тенденція прямує до зростання чи спадання. Мультиплікативну модель [див. формулу (2.2)] доцільно використовувати тоді, коли коливання часового ряду збільшуються або зменшуються протягом певного періоду [17].

$$y(t) = T(t) + C(t) + S(t) + I(t), \quad (1.4)$$

де T — тренд;

C — циклічність;

S — сезонність;

I — залишки.

$$y(t) = T(t) * C(t) * S(t) * I(t), \quad (1.5)$$

де T — тренд;

C — циклічність;

S — сезонність;

I — залишки.

На основі адитивної та мультиплікативної форми існує метод декомпозиції STL. Це метод, який розкладає ряд на тренд, сезонність та залишкову складову. Даний метод згладжує часові ряд з використанням LOESS. Тобто спочатку

алгоритм обчислює сезонну компоненту та видаляє її для подальшого обчислення тренду. Залишкова компонента формується шляхом віднімання сезонної та трендової складової. Також цей метод мінімізує викиди, тому вони не впливатимуть на сезонну та трендову компоненту [17].

Часові ряди бувають стаціонарними та нестаціонарними. Стаціонарний часовий ряд це ряд статистичні параметри якого такі як середнє значення, дисперсія, коваріація та стандартне відхилення є сталими величинами та не змінюються з часом. Тобто, якщо часовий ряд має явну тенденцію чи сезонність, то він є не стаціонарним. Стаціонарність часового ряду можна перевірити за допомогою статистичних тестів на стаціонарність або переглянувши ряд візуально за допомогою графіку [19].

Одним зі статистичних тестів перевірки на стаціонарність є розширений тест Дікі-Фуллера. Цей тест базується на теорії одиничного кореня, який може спричиняти проблеми в статистичних висновках моделей часових рядів. Даний тест проводиться з такими припущенням:

- H_0 : ряд є нестаціонарним, тобто ряд має одиничний корінь.
- H_1 : ряд є стаціонарним, тобто ряд не має одиничного кореня.

Якщо розрахована статистика менша критичного значення та p-value менше обраного порогу вірогідності, то можна відхилити нульову гіпотезу, тобто часовий ряд стаціонарний. Дана властивість часового ряду дуже важлива при використанні різних методів моделювання та прогнозування часового ряду [19]. Проте, існують способи за допомогою яких можна позбутися стаціонарності ряду. Для цього потрібно трансформувати часовий ряд за допомогою логарифмування або застосування кореня квадратного до значень ряду чи диференціювання. Диференціювання є одним з найефективніших методів [20]. Різниця першого порядку часового ряду розраховується за такою формулою:

$$p_t = p_t - p_{t-1}, \quad (1.6)$$

де p_t — значення поточно періоду;

p_{t-1} — значення попереднього періоду.

Може бути таке, що після трансформації ряду за допомогою різниці першого порядку, він все рівно буде нестационарним. Тоді потрібно спробувати добути різницю наступного порядку. Також існує поняття сезонної різниці. Це різниця між поточним спостереженням і спостереженням попереднього такого ж сезону. В результаті трансформацій часового ряду можна отримати стаціонарний ряд [17].

Наступною важливою властивістю часових рядів, яку потрібно досліджувати це автокореляція. Автокореляція вимірює зв'язок між значеннями часового ряд з відставанням. Часовий ряд, який не має автокореляції називається білим шумом. Існують методи прогнозування, що базуються на теорії автокореляції. Ці методи мають назву ARIMA. Для застосування групи цих методів важливо, щоб часовий ряд був стаціонарним та мав статистично значиму автокореляцію. Метод ARIMA включає в себе три компоненти:

- Авторегресія (AR): модель використовує множинну регресію з лаговими тобто попередніми значеннями в ролі предикторів.
- Інтеграція (I): компонента, що представляє порядок різниці необроблених спостережень, для того щоб привести ряд до стаціонарного
- Ковзне середнє (MA): модель використовує множинну регресію з помилками та лаговими значеннями в ролі предикторів [21].

Сезонна авторегресійна модель включає в себе додатково до авторегресійної, сезонну частину моделі. Сезонна компонента в даній моделі складається з сезонних різниць.

Одним з методів, який можна застосовувати до нестационарних часових рядів є методи експоненціального згладжування. Прогнози створені за допомогою цих методів, є середньозваженими значеннями минулих спостережень, ваги зменшуються експоненціально, коли спостереження беруться за більш давній

період. Ключова ідея цього методу полягає в тому, щоб надати більшої ваги останнім спостереженням, ніж давнішим. Цей метод найкраще прогнозує ряди з наявним постійним трендом та сезонністю [22].

Залишкова компонента часових рядах це, те що залишається після підгонки нашої моделі. Дуже важливо після проведення моделювання оцінити залишкову складову на наявність якихось залежностей. Якщо, модель є якісною, то залишкова компоненти не матиме кореляції, а її середнє значення дорівнюватиме нулю. Також важливо, щоб залишки мали постійну дисперсію та нормальний розподіл. В іншому випадку означатиме, що модель є не досконалою. Для залишкової діагностики потрібно побудувати гістограму та корелограму залишків. Також додатково до цього існують статистичні тести для перевірки залишків. Одними з таких тестів є тест Бокса-Пірса, тест Льюнга-Бокса. Дані тести перевіряють часові ряди на наявність автокореляції. Якщо в залишковій компоненті відсутня автокореляція за даними тестами, це означає, що наша модель якісно використала всі дані, тому в залишках відсутні будь які залежності [17].

Для оцінки похибки прогнозу моделей можна розрахувати різні метрики. Одним з важливих показників точності прогнозу є середньоквадратичне відхилення. RMSE розраховується за формулою [23]:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_t - \hat{y})^2}{n}}, \quad (1.7)$$

де y_t — фактичне значення;

$\hat{y}$ — модельоване значення;

n — кількість спостережень.

Цей показник вимірює похибку прогнозу. Тому чим нижче отримане значення, тим точніше модель розраховує прогнозовані значення. Ще одним з видів показників оцінки похибки прогнозу є середня абсолютна похибка. MAE розраховується за такою формулою [23]:

$$MAE = \sum_{i=1}^n \frac{|y_t - \hat{y}|}{n}, \quad (1.8)$$

де y_t — фактичне значення;

$\hat{y}$ — модельоване значення;

n — кількість спостережень.

Для вибору якісної моделі існують спеціальні інформаційні критерії. Критерій Акаїке визначається за такою формулою [24]:

$$AIC = 2k - 2 \ln(L), \quad (1.9)$$

де L — значення функції правдоподібності моделі;

k — кількість використаних предикторів.

Критерій Акаїке оцінює складність моделі. Цей критерій встановлює баланс між правдоподібністю моделі та її складністю у вигляді змінної кількості параметрів моделі. Чим менше значення даного критерію, тим якіснішою побудована наша модель. Якщо, у нас є невелика вибірка даних, то існує скоригований критерій Акаїке. Так як звичайний критерій Акаїке віддає перевагу більш складним моделям, тому коли розмір вибірки достатньо невеликий існує скоригований показник, що коригується на розмір вибірки [24].

Інформаційний критерій Байеса теж штрафує модель на правдоподібність та складність, проте цей критерій відразу враховує розмір вибірки. Баєсівський інформаційний критерій розраховується за наступною формулою [24] :

$$BIC = k * \ln(n) - 2 \ln(L), \quad (1.10)$$

де L — значення функції правдоподібності моделі;

k — кількість використаних предикторів;

n — кількість спостережень у вибірці.

Інформаційні критерії враховують правдоподібність, кількість параметрів та розмір вибірки моделі. Критерій Байеса віддає пріоритет більш простішим моделям так як суворіше штрафує складність ніж критерій Акаїке. Оцінюючи результат моделювання ми можемо оцінити похибку прогнозу, а також складність моделі за допомогою інформаційних критерії та прийняти оптимальне рішення для нас [24].

Методи, які розглядалися до цього є однофакторними методами, проте досить часто важливо переглянути як впливають на незалежну змінну інші фактори. Наприклад ми хочемо оцінити вплив інфляції, рівня ВВП та рівня заробітної плати на безробіття. Для цього можна використовувати метод векторної авторегресії. Модель векторної авторегресії базується на однофакторній моделі авторегресії, проте з додаванням лагових векторів інших змінних. Для застосування даного методу нам потрібні лише стаціонарні часові ряди. Якщо ряди нестаціонарні, то ми беремо різниці даних, для того щоб зробити їх стаціонарними. Параметри для кожного рівняння оцінюються шляхом мінімізації суми квадратів. Даний метод генерує моделі та прогнози для кожної змінної включеної в систему. Для побудови моделі методом VAR необхідно визначити кількість змінних, які будуть використовуватись, а також скільки лагів повинно бути включено в модель [17].

РОЗДІЛ 2

МАТЕМАТИЧНІ МОДЕЛІ АНАЛІЗУ ДИНАМІЧНИХ ПРОЦЕСІВ НА РИНКУ ПРАЦІ

2.1 Інструментарій розвідувального аналізу часових рядів факторів ринку праці

Ринок праці є невід’ємною складовою економіки країни. Основою ринку праці є люди. Тому потрібно проаналізувати демографічну ситуацію в Україні, для розуміння її стану та динаміки. Динаміка чисельності населення України має спадаючий тренд (рис 2.1). На початку, 1992 р. кількість постійного населення України була на рівні 52 млн. осіб, то на початок 2022 р. ця кількість зменшилася до 41 млн. осіб. За останні 20 років, чисельність постійного населення знизилася на 20%, а середній річний темп падіння був на рівні 0,8%. У 2015 році був значний спад чисельності населення. Великий вплив на зниження показнику у 2015 р. мало вторгнення у 2014р. Росії в Крим та на Донбас. З початком повномасштабного вторгнення Російської Федерації в Україну 24 лютого 2022 р. багатьом людям довелося покинути свої домівки. На жаль оцінити рівень населення на даний момент досить складно. Проте за даними ООН від початку повномасштабного вторгнення з України виїхали понад 6,2 млн. українців. Відповідно до цього населення України на початок 2023 р. складало 36,7 млн. осіб, а держава втратила близько 10,9% населення за 1 рік.

На зниження рівня чисельності населення впливає рівень народжуваності в країні. Рівень народжуваності розраховується як відношення кількості народжених за рік на 1 тис. населення до середньої чисельності всього населення. За даними Держстату України на початку 2022 р. рівень народжуваності становив 1,2 проте за

демографічною теорією для відновлення покоління цей показник має бути на рівні 2,1-2,2. На жаль з початком повномасштабного вторгнення кількість народжених дітей значно знизилася. За шість місяців 2023 року в Україні народилося близько 93,5 тис. дітей, це на 28% нижче, ніж в аналогічний період 2021р. В цілому в Україні за останні роки тенденція до зниження народжуваності (рис 2.2). Для балансу населення країни повинна бути рівновага між народжуваністю та смертністю. За останні 20 років смертність сильно перевищує народжуваність в Україні. Проте як бачимо пік смертності був у 2021 р. через пандемію COVID-19 [27]. На жаль поки що немає точних даних смертності за 2023 р. Проте зрозуміло, що військові дії на території України несуть за собою велику кількість жертв серед населення.

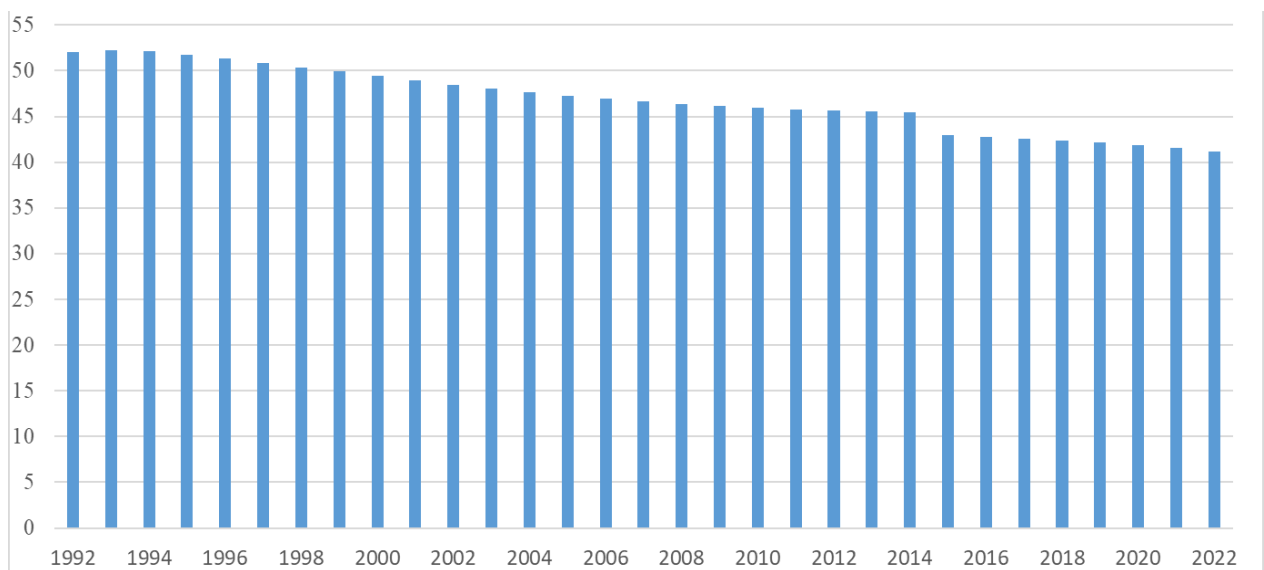


Рисунок 2.1 - Чисельність населення (млн.осіб) України за період 1992-2022 рр.

Джерело: розроблено автором за даними [25]

На даний момент демографічну ситуацію в Україні можна оцінити як критичну. Рівень народжуваності за 2023 р. може бути найнижчим за всю історію незалежної України. Військові дії на території України негативно впливають на чисельність населення. За прогнозами Інституту демографії та соціальних

досліджень імені М.В. Птухи НАН України чисельність населення України до 2030 року може скоротитися до 24-32 млн. осіб [29].

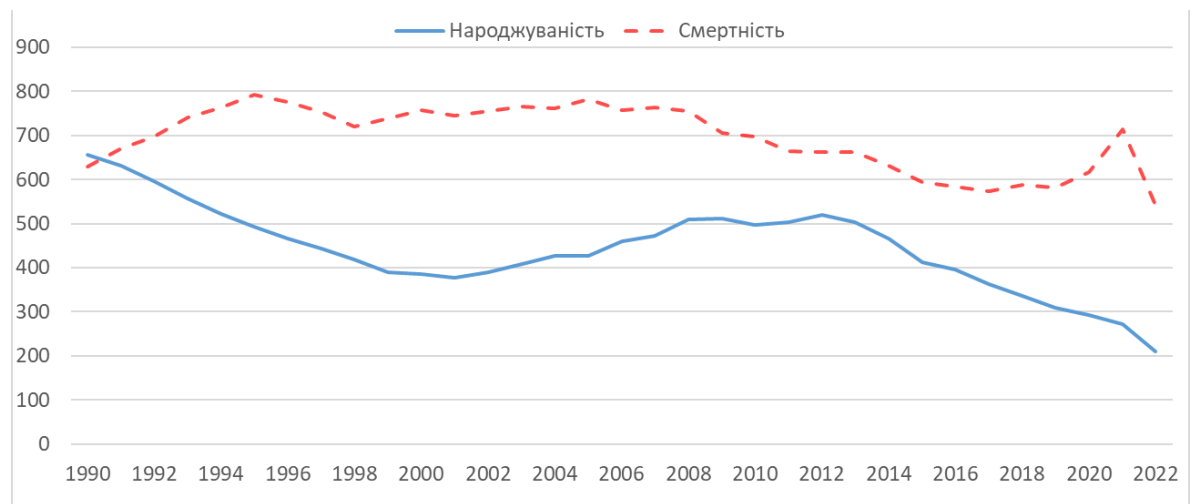


Рисунок 2.2 - Народжуваність та смертність (тис.осіб) в Україні за період 1990-2022 рр.

Джерело: розроблено автором за даними [26][28]

Результатом демографічних змін в Україні є старіння нації. Демографічне старіння це збільшення частки населення старшого віку в структурі населення. Частка людей похилого віку зростає у всьому світі. Проте, в Україні внаслідок демографічних проблем з народжуваністю та смертністю ця проблема загострюється. Частка населення 65 років і старше у 2020 р. була на рівні 17,1% але за останні 20 років, частка зросла на 5% (рис 2.3). В 2023 році проблема з народжуваністю та смертністю через повномасштабне вторгнення стало ще більш критичною, то частка населення віком 65 років і старше теж залишається на високому рівні. Старіння нації впливає не лише на ринок праці, а й в цілому на всю економіку. Так велика частка населення пенсійного віку збільшує навантаження на економіку та бюджет країни [30].

Високий рівень населення пенсійного віку (65 років і старше) створює загрозу для нормалізації процесу відтворення робочої сили. До економічно активного населення відносять осіб чоловічої та жіночої статі віком від 15 до 70

років, які протягом певного періоду забезпечують пропозицію робочої сили на ринку праці [1].

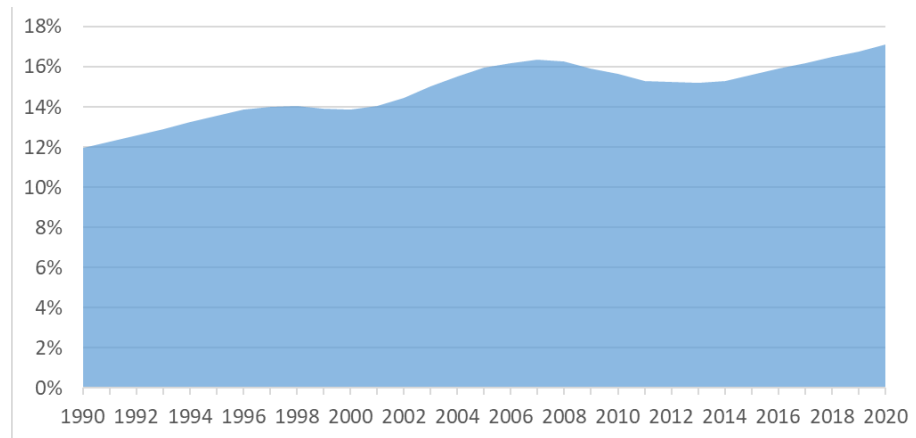


Рисунок 2.3 - Частка населення у віці 65 років та старше за період 1990-2020 рр.

Джерело: розроблено автором за даними [26]

Збором та публікацією даних по економічній активності населення займається Державна служба статистики України. На жаль з початком повномасштабного вторгнення у 2022 р. дослідження робочої сили не проводяться, так як процеси на ринку праці зараз складно оцінити. Проте за даними Українського інституту майбутнього на початок 2023 р. в Україні проживає 12,3 млн. економічно активного населення [31]. В цілому за останні 20 років зберігається негативна тенденція рівня економічно активного населення (рис 2.4). Рівень економічно активного населення за останні 20 років знизився на 45%. Найбільше падіння спостерігається через початок повномасштабного вторгнення 2022 р. На жаль, важко давати якісь прогнози на найближче майбутнє, проте зрозуміло, що нас очікує довготривале відновлення як в економіці так і на ринку праці.

Найкращу оцінку стану ринку праці відображає безробіття та зайнятість. Ці показники в цілому характеризують соціально-економічний стан суспільства. Контроль та максимізація рівня зайнятості є важливим завданням держави [32]. Як бачимо у 2014 р. було падіння рівня зайнятості на 3%, що є одним з найбільших за

період 2003 – 2020 рр. (рис 2.5). Падінню посприяли зміна влади у 2014 році, а також початок військових дій на Донбасі та в Криму. Далі протягом 6 років спостерігалось незначне зростання рівня зайнятості до 58,2% у 2020р. Проте, у 2022 р. через повномасштабне вторгнення рівень зайнятості значно знизився. На жаль, точно оцінити падіння складно, проте за даними НБУ рівень зайнятості знизився на 5,3% порівнюючи з довоєнним 2021 р. та становить 49,7%. За прогнозами НБУ рівень зайнятості поступово стабілізується протягом 2023р.

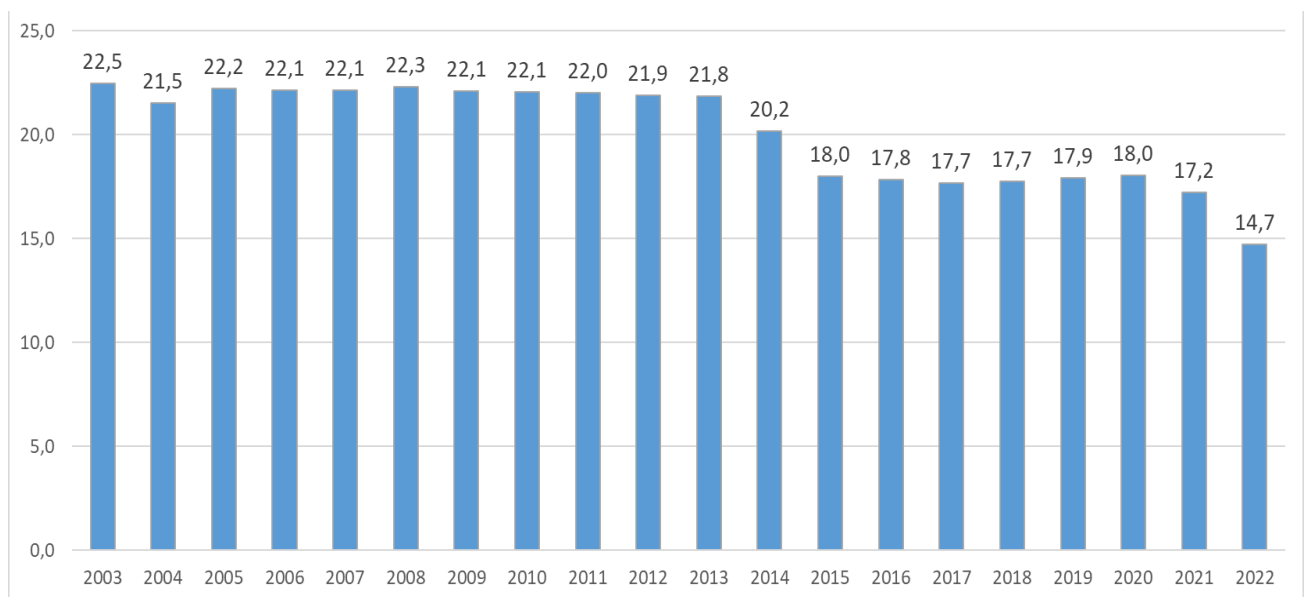


Рисунок 2.4 - Рівень економічно активного населення (млн.осіб) за період 2003-2022 рр.

Джерело: розроблено автором за даними [26]

Ще одним важливим показником ефективності ринку праці є рівень безробіття. Дослідженням рівня безробіття займається Державна служба статистики України за допомогою методології МОП [32]. У 2009 році. було значне зростання рівня безробіття на 2,7% у порівнянні до попереднього року (рис 2.6). В усьому світі відбувалася фінансово-економічна криза через, що відбулося значне зростання кількості безробітного населення в Україні. Протягом наступних 5 років рівень знижувався та досягнув 8,6% у 2013 р.. Проте, на початку 2014 р. відбулася зміна влади в Україні, а також початок військових дій на Донбасі та в Криму, тому

було поступове зростання рівня безробіття до 10% у 2015 р. На даний момент, через початок повномасштабного вторгнення точно оцінити рівень безробіття, досить складно. Проте за даними дослідницької організації Info Sapiens рівень безробіття на кінець 2022 р. становив 22,3% [33]. Рівень безробіття зріс на 11,2% у порівнянні з 2021 р.

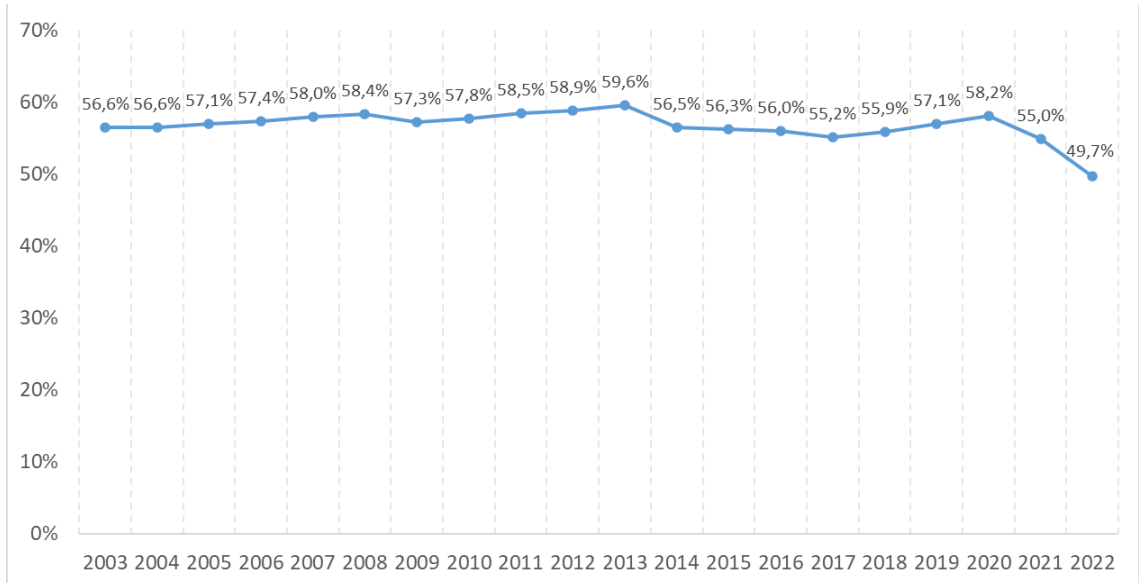


Рисунок 2.5 – Рівень зайнятості в Україні за період 2003 – 2022 рр.

Джерело: розроблено автором за даними [26][33]

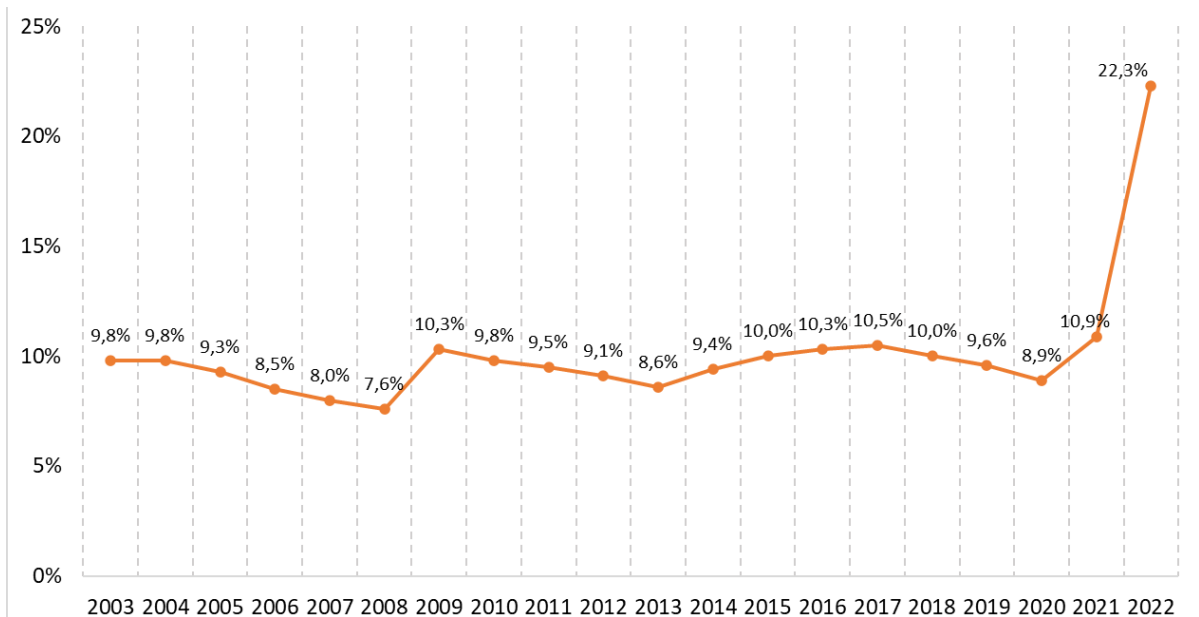


Рисунок 2.6 – Рівень безробіття в Україні за період 2003 – 2022 рр

Джерело: розроблено автором за даними [26][33]

Для оцінки поточно стану ринку праці варто розглянути попит та пропозицію на ринку праці. Попит на ринку праці створює робоча сила, а пропозицію роботодавець, який в пошуку працівника. Для оцінки попиту на ринку праці слід розглянути активність шукачів роботи. На жаль з початком повномасштабного вторгнення у лютому 2022 р. відбулися масштабні зміни як в економіці країни так і на ринку праці. Велика кількість населення покинула свої домівки. Через це активність роботодавців та робочої сили значно знизилася в умовах війни. За дослідженням Центру економічної стратегії у березні 2022 року кількість нових резюме знизилася на 75% у порівнянні до середньорічного рівня 2021р (рис 2.7). Проте поступово рівень активності шукачів роботи відновився та в травні 2022 р. кількість нових резюме досягнула довоєнного рівня. Попит на ринку праці поступово відновився, проте після піку попиту в середині 2022 р. попит в 2023 р. знизився в середньому на 20% у порівнянні з довоєнним рівнем.

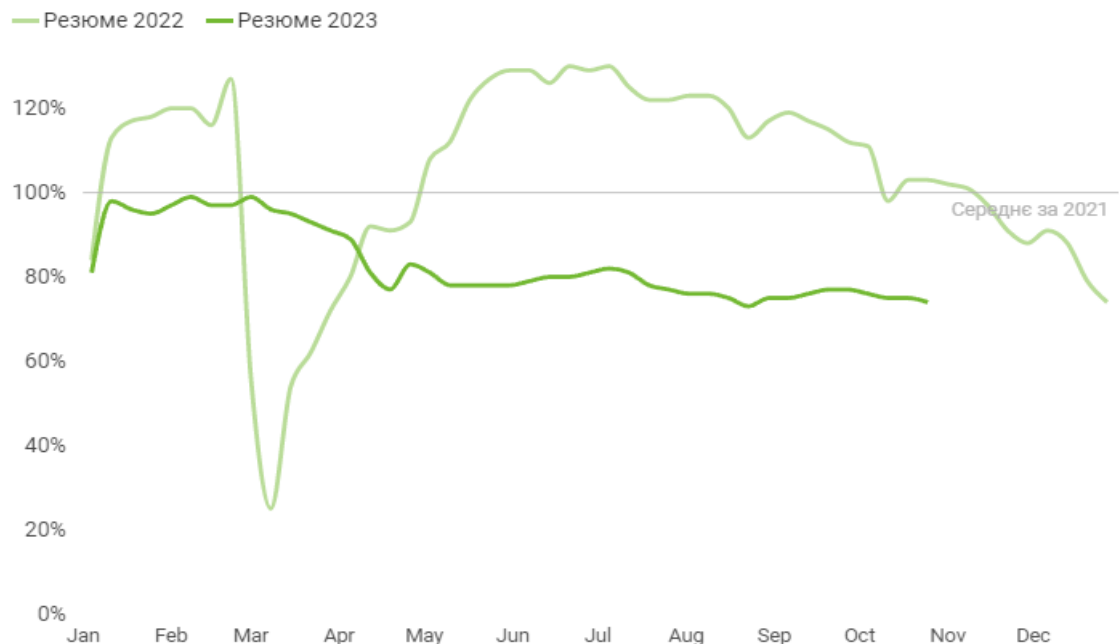


Рисунок 2.7 – Індекс приросту кількості нових резюме у порівнянні з 2021 р.

Джерело: [33]

Проаналізувавши попит на ринку праці, потрібно розглянути пропозицію. Для оцінки пропозиції на ринку праці розглянемо кількість вакансій на ключових

платформах пошуку роботи (robota.ua, work.ua, Jooble) в Україні. Як бачимо з початком повномасштабних військових дій на території України в лютому 2022р. було падіння кількості вакансій на 72% у порівнянні з довоєнним рівнем (рис 2.8). На жаль через військові дії багато підприємств було зруйновано чи припинили свою роботу. Поступово роботодавці почали відновлювати свою активність. Тому на вересень 2023 р. кількість вакансій становить 70% від довоєнного рівня. На жаль пропозиція на ринку праці не відновилася до довоєнного рівня, проте вона стабілізувалася та покращилася у порівнянні з березнем 2022 р. Як бачимо, попит значно перевищує пропозицію на ринку праці України в 2023 р. Проте рівновага попиту та пропозицію на ринку поступово вирівнюється.

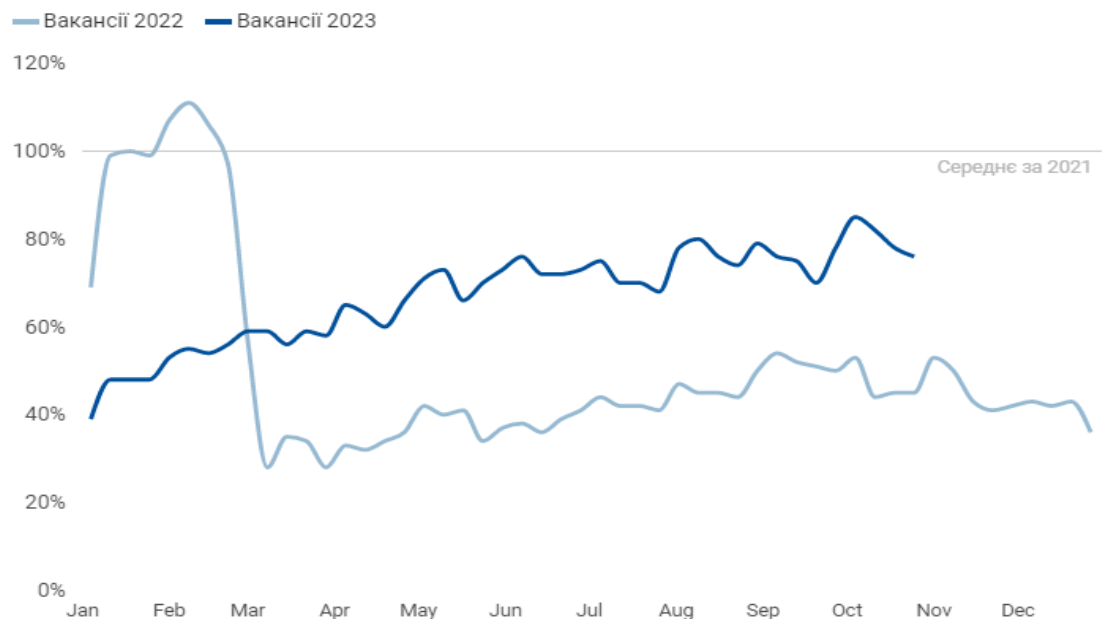


Рисунок 2.8 – Індекс приросту кількості нових вакансій у порівнянні з 2021р.

Джерело: [33]

Рівень середньої заробітної плати є важливим показником стану ринку праці. Заробітна плата є одним з головних індикаторів оцінки рівня соціально-економічного життя. В цілому середня заробітна плата має позитивний тренд (рис 1.9). Проте, зростання заробітної плати, не означає, що рівень життя також підвищився, так як зі зростанням заробітної плати зростала інфляція та ціни на

товари і послуги, курс валют [34]. За даними Пенсійного фонду України на початок 2022 р. середня заробітна плата становила 13735 грн. Як бачимо, з початком повномасштабного вторгнення у лютому 2022 р. заробітна плата підвищилася на 15%, проте на це повпливали авансові платежі та компенсації, які виплачувало багато компаній у якості підтримки своїх працівників. Уже на початку 2023 р. рівень середньої заробітної плати мав незначне падіння до 12954 грн. В цілому рівень заробітної плати підвищився у порівнянні з довоєнним рівнем. Проте, це сильно на вплинуло на покращення життя українців, тому що зі зростанням заробітної плати зростали ціни на ключові товари та послуги.

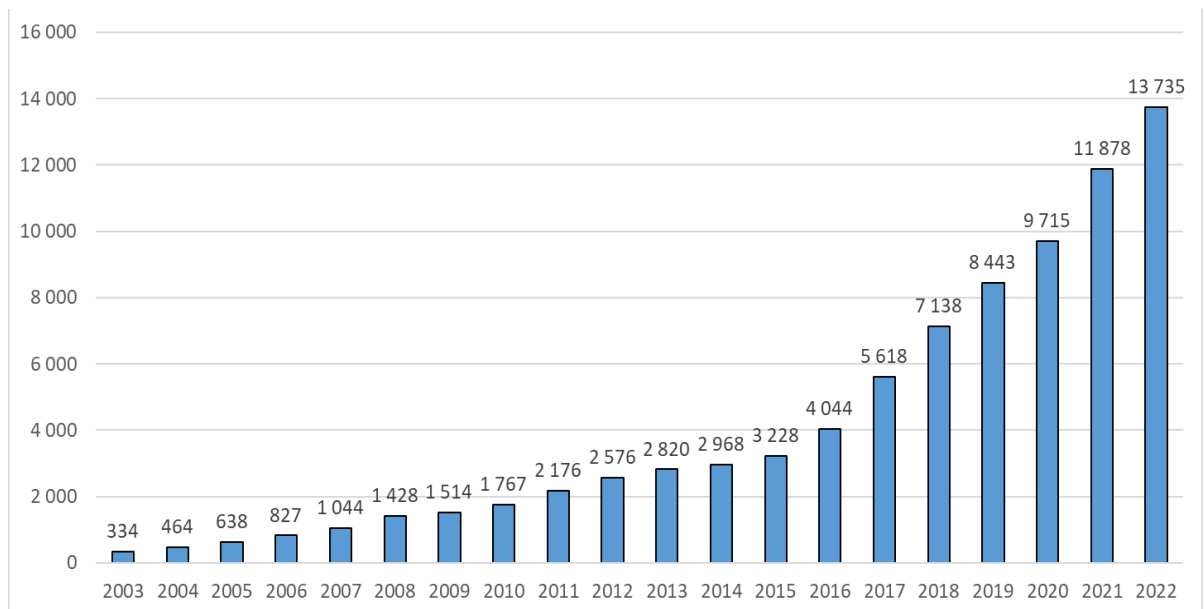


Рисунок 2.9 – Динаміка середньої заробітної плати за період 2003 – 2023 рр.

Джерело: розроблено автором за даними [35]

Отже, поточний стан на ринку праці України є критичним. Чисельність населення за рік скоротилася приблизно на 6,2 млн. осіб, що є найбільшим падінням за останні 20 років. На це вплинув початок повномасштабної війни на території України 24 лютого 2022 р. На жаль рівень народжуваності на кінець року може досягнути найменшого показника за 20 років. Оцінити рівень смертності досить складно, проте зрозуміло, що масштабні військові дії несуть за собою великі втрати

серед населення. Внаслідок цього, в Україні загострюється проблема старіння нації так як частка населення віком 65 і старше дуже висока. Через це чисельність економічно активного населення сильно скоротилася та становить 12,3 млн. на початок 2023р. Зміни в демографічному стані держави негативно вплинули на попит та пропозицію на ринку праці, а також на показники зайнятості та безробіття. За даними НБУ рівень зайнятості на даний момент скоротився на 12% порівнюючи з довоєнним періодом та становить 44% на 2023р. Рівень безробіття теж досягнув свого максимуму за 20 років та становить 22,1% на початок 2023р. Попит на ринку праці відновився до довоєнного рівня, проте пропозиція значно знизилася, через, що на ринку велика кількість потенційних кандидатів, але досить низька кількість вакансій. За прогнозами фахівців, ситуація з зайнятістю та безробіттям поступово відновлюватиметься до довоєнного рівня. Проте для покращення демографічної ситуації потрібно багато років. Нас очікує довге та важке соціально-економічно відновлення, внаслідок якого ми можемо досягнути довоєнного рівня показників оцінки ринку праці та навіть їх перетнути.

2.2 Моделювання динаміки ринку праці України та зарубіжних країн

Ринок праці є ключовою парадигмою економічного зростання у всьому світі. Тому аналіз ринку праці важливо оцінювати з різних сторін. Спочатку ми оцінили ринок праці в Україні, тепер же для розуміння ситуації проведемо порівняльний аналіз з іншими країнами світу. Пропоную для початку порівняти динаміку чисельності постійного населення в Україні та інших країнах. Для порівняння був обраний такий список країн:

- Сполучені Штати Америки так як це країна з найбільшою економікою у світі за даними 2022р [36].

- Велика Британія, Німеччина, Франція та Італія – країни Європейського континенту з найбільшою економікою у 2022 р [36].
- Польща, Румунія, Угорщина, Словаччина – найбільші країни сусіди України за розміром економіки у 2022 р.

Отже, як бачимо Україна має значне скорочення населення на рівні 9,8 млн. осіб у 2022р. США, Велика Британія Франція, Німеччина та Італія відносяться до країн Великої сімки і мають приріст постійного населення (рис 2.10). Населення Польщі, Румунії та Угорщини теж скоротилося за останні 20 років як і в Україні. Проте, на жаль Україна виділяється серед усіх країн великим рівнем скорочення населення. На жаль на це сильно вплинуло початок повномасштабного вторгнення в Україні в лютому 2022р. Адже лише за один рік рівень постійного населення України скоротився приблизно на 6 млн осіб.

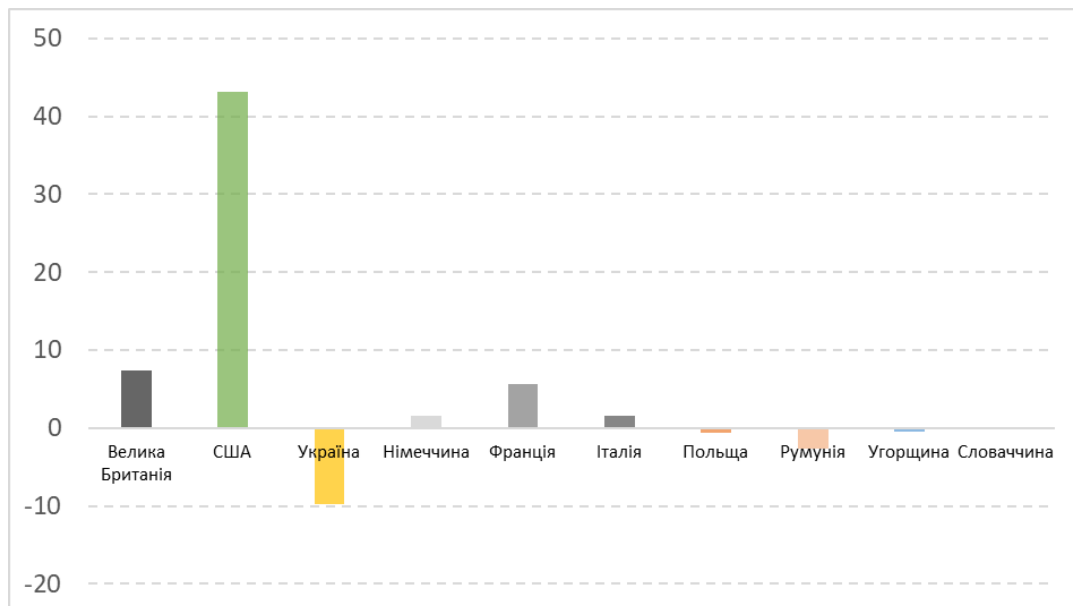


Рисунок 2.10 – Темп зміни постійного населення (млн. осіб) в розрізі країн за період 2003 – 2022 рр.

Джерело: розроблено автором за даними [36]

Також проаналізуємо важливий показник оцінки демографічного стану країни - рівень народжуваності. В Україні на 2021 р. рівень народжуваності становив 1,16, що є найнижчим показником серед визначених раніше нами країн

(рис 2.11). Рівень народжуваності в усьому світі становить 2,3 дитини на одну жінку. Тобто, країни, які наведені на рисунки мають рівень народжуваності нижчий ніж середній по всьому світу. Як бачимо у Франції найвищий рівень народжуваності серед досліджуваних країн. Проте в цілому у багатьох країнах Європи та США існує проблема з відтворенням населення. Наприклад рівень народжуваності у Польщі та Італії теж є низьким на рівні 1,3. Тому проблема з відтворенням населення є актуальною не лише в Україні. Проте на жаль початок повномасштабного вторгнення у 2022р. лише негативно вплинув на рівень народжуваності. Тому, ми можемо говорити, що у 2023р. рівень народжуваності буде ще нижчий ніж у 2021р. Такі події негативно вплинуть на темп зміни постійного населення. Відповідно це негативно впливає на стан ринку праці в Україні.

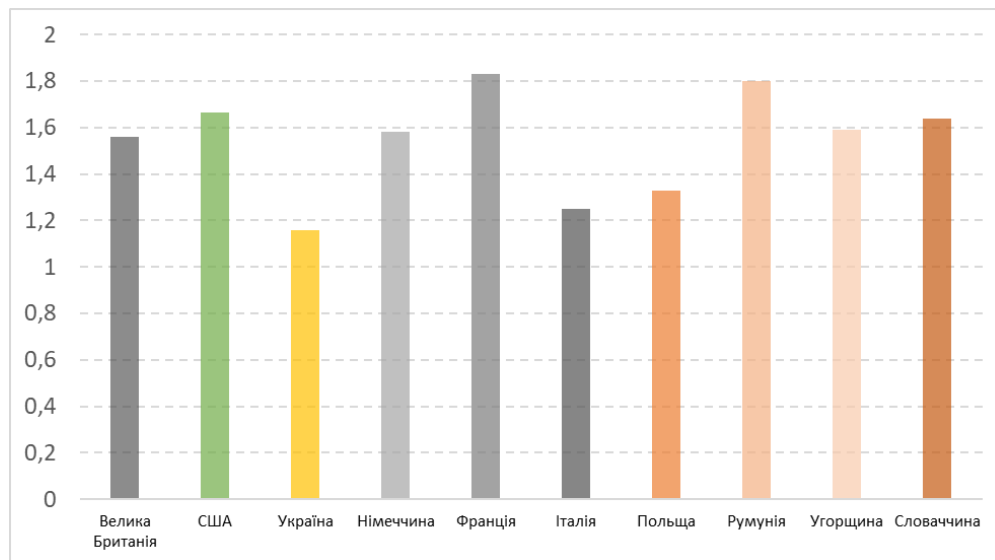


Рисунок 2.11 – Рівень народжуваності в розрізі країн у 2021р.

Джерело: розроблено автором за даними [36]

Далі пропоную проаналізувати вікову структуру населення. Це дасть змогу детальніше оцінити демографічну ситуації в середині країни. Як бачимо, Україна має достатньо не високий відсоток населення віком 65 і старше порівнюючи з іншими країнами (рис 2.12). В Україні на 2022 р. цей показник становив 18,8%.

Також в Україні рівень населення дитячого віку 0-14 років перевищує рівень в Німеччині чи Італії та становить близько 15% (рис 2.13). В Італії, Німеччині та Франції рівень населення пенсійного віку (65 і старше) значно вищий ніж в Україні. В цілому у багатьох країнах Європи спостерігається проблема старіння нації. Так як висока частка людей пенсійного віку в державі впливає на зростання економічного навантаження. Як бачимо, у Франції, Німеччині та Італія дана проблема дійсно є гострою. Проте, Україні дана проблема теж притаманна. Адже зниження рівня народжуваності останніми роками в Україні позитивно впливає на зростанні рівня населення пенсійного віку та зниження населення дитячого віку. Висока частка населення пенсійного віку також негативно впливає на становище ринку праці. Так як зростання рівня населення пенсійного віку негативно впливає на чисельність робочої сили. Початок повномасштабного вторгнення лише загострить проблему старіння нації в Україні. Адже уже зрозуміло, що рівень народжуваності знизився до найнижчого рівня за всю незалежність України, а рівень населення 65 років і старше відповідно зростає за рахунок цього.

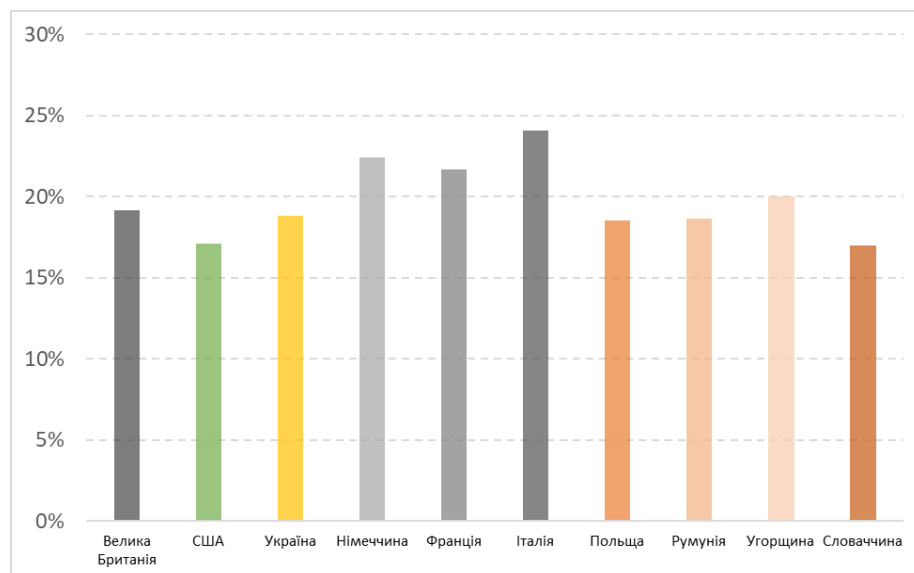


Рисунок 2.12 – Частка населення віком 65 років і старше в розрізі країн у 2022р.

Джерело: розроблено автором за даними [36]

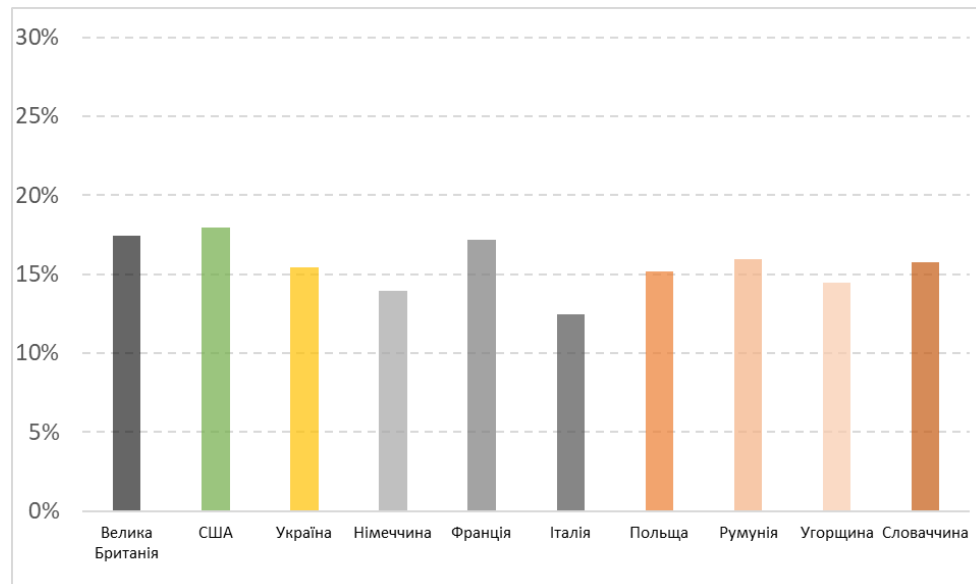


Рисунок 2.13 – Частка населення віком від народження до 14 років в розрізі країн у 2022р.

Джерело: розроблено автором за даними [36]

Далі порівняємо рівень зайнятості та безробіття України та країн Європи, США. Показники зайнятості та безробіття найкраще характеризують ситуацію на ринку праці та в цілому соціальну складову держави. Середній рівень безробіття за даними 2022р. у світі становив 5,8%. Рівень безробіття в Україні на 2022 р. є найвищим серед досліджуваних нами країн (рис 2.14). На жаль, на це вплинули сильні міграційні процеси внаслідок початку повномасштабного вторгнення в Україні. Також через початок військових дій на території України багато підприємств було зруйновано або вимушено закрито. Як бачимо в США, країні з найбільшою економікою у світі, рівень безробіття у 2022 р. був на рівні 3,6%. В країнах Європейського союзу в середньому рівень безробіття становить 6,1%. Проте, у довоєнний період рівень безробіття в Україні був теж один з найвищих в Європі та становив 10,9% на 2021р. Проблема з високим рівнем безробіття актуальна для України досить давно.

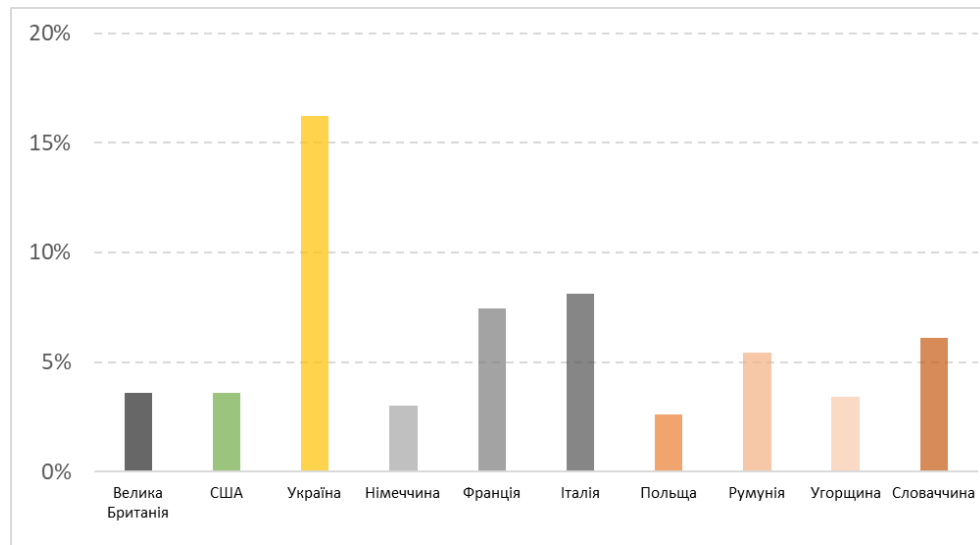


Рисунок 2.14 – Рівень безробіття в розрізі країн у 2022р.

Джерело: розроблено автором за даними [36]

Рівень безробіття тісно пов'язаний із зайнятістю. Як бачимо, у 2022 р. в Україні рівень зайнятості становив 49,7% (рис 2.15). Цей показник є одним з найнижчих в Європі, проте в Італії рівень зайнятості ще нижчий ніж в Україні, а в Румунії на такому ж рівні. В цілому по світу рівень зайнятості становить 56%. Висока частка населення пенсійного віку негативно впливає на зайнятість робочої сили. Наприклад в Італії частка населення пенсійного віку є найвищою серед досліджуваних країн, що вплинуло на низький рівень зайнятості. В Україні частка населення пенсійного віку теж зростає з кожним роком. З початком повномасштабного вторгнення проблема з зайнятістю лише загострилася. Тому, рівень зайнятості значно знизився у 2022 р. На жаль погана демографічна ситуація, нестабільна політична та економічна ситуація, а також військові дії на території України лише загострюють проблему з зайнятістю та безробіттям. На даний момент в Україні одна з найгірших ситуацій з зайнятістю та безробіттям населення в Європі. Тому в найближчому майбутньому потрібно проводити певні дії зі сторони держави для покращення ситуації. Адже це впливає не лише на ринок праці, а і на економіку та суспільство в цілому.

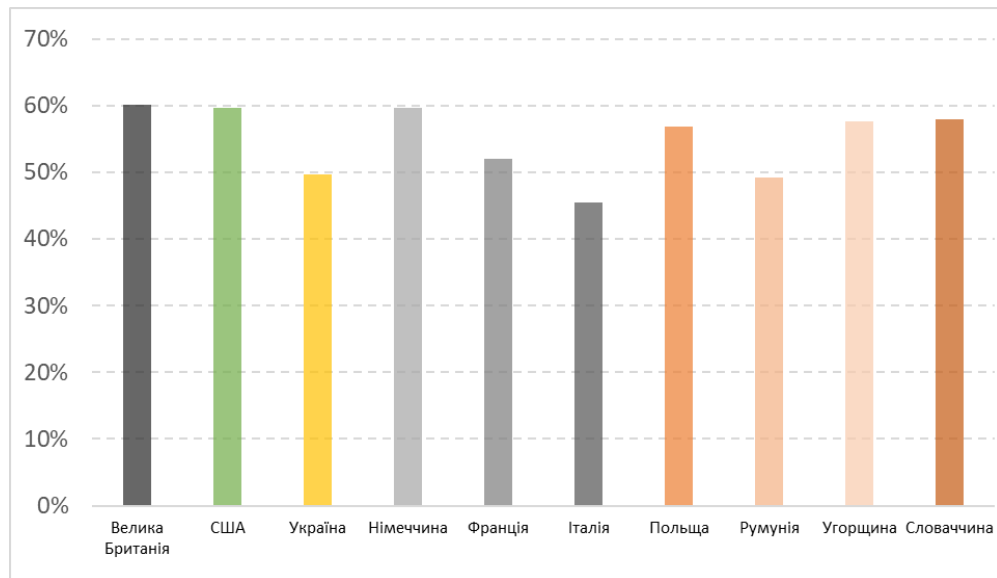


Рисунок 2.15 – Рівень зайнятості в розрізі країн у 2022р.

Джерело: розроблено автором за даними [36]

Оцінка рівня задоволеності своєю роботою якісно оцінює ситуацію в суспільстві та ринку праці, що складається з людей. Таке дослідження досить складно провести, проте у 2021 р. в багатьох країнах незалежна організація Randstad Workmonitor провела дослідження задоволеності праці [37]. В Україні дане дослідження не проводилося. Проте, є місцеві організації, які досить часто проводять такі опитування серед українців. Тому за даними кадрового порталу grc.ua 48% українців в цілому задоволені роботою з них лише 10% повністю задоволені своїм місцем роботи [38]. Якщо спробувати порівняти дану оцінку з оцінками в інших країнах, то Україна має найнижчий бал (рис 2.16). Як бачимо серед досліджуваних країн США мають близько 80% задоволених роботою громадян, що є високим показником у всьому світі. В середньому у Європейських країнах 70% громадян задоволені своєю роботою.

Задоволеність працею тісно пов'язана з рівнем заробітної плати. Тому пропонуємо також розглянути середню заробітну плату по країнам за 2022 р. За даними сформованими Організацією економічного співробітництва та розвитком в Україні середня заробітна плата становила у 2022 р. 414 дол. США (рис 2.17). На жаль як бачимо, це найнижчий показник серед досліджуваних нами країн. Рівень

заробітної плати нижчий від середнього рівня Європейських країн більше ніж в 2 рази. Тому, для покращення задоволеності працею в суспільстві нам потрібно в першу чергу намагатися підвищити рівень заробітної плати до Європейського рівня.

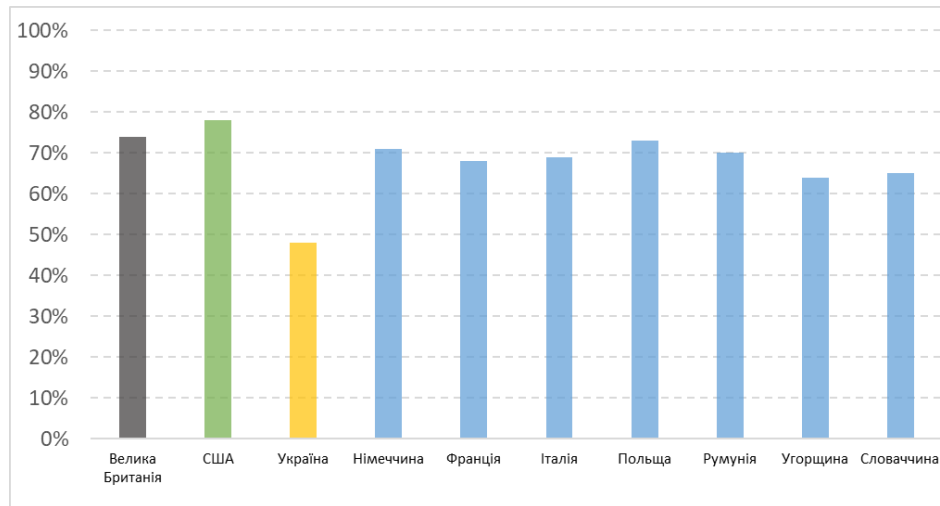


Рисунок 2.16 – Задоволеність працею в розрізі країн у 2021 р.

Джерело: розроблено автором за даними [37][38]

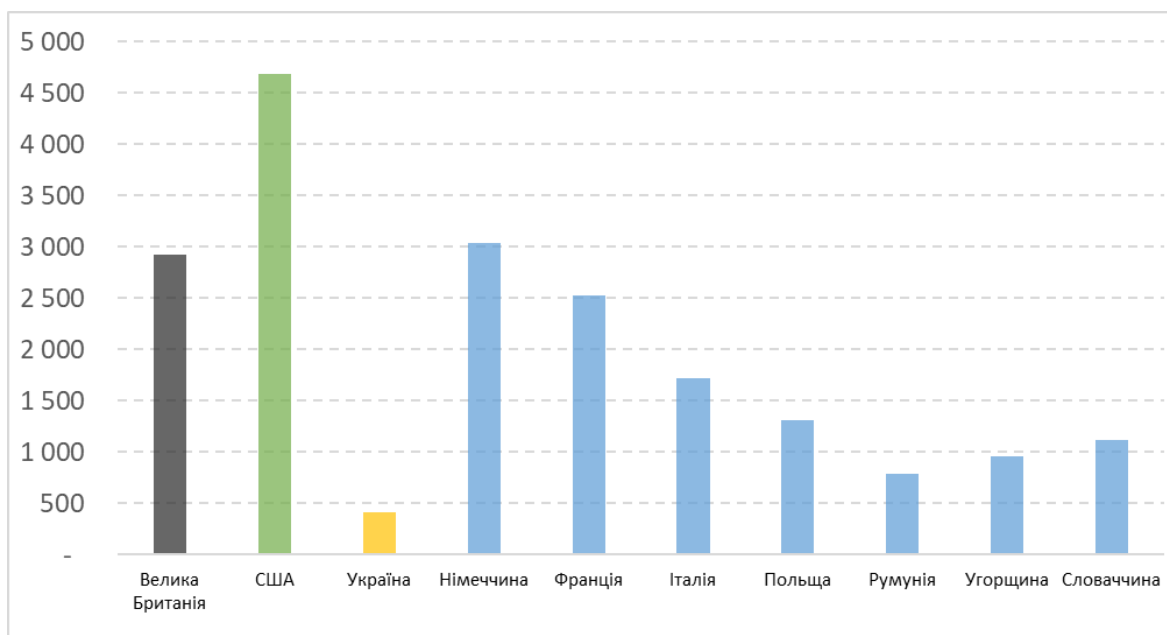


Рисунок 2.17 – Середня заробітна плата (дол. США) в розрізі країн у 2022р.

Джерело: розроблено автором за даними [39]

Далі пропоную проаналізувати також продуктивність праці у світі та в Україні. Продуктивність праці оцінює результативність праці та її витрати як на макро так і на мікро рівнях. Для оцінки продуктивності праці на рівні держави цей показник розраховується як ВВП що припадає на одну відпрацьовану годину часу або ВВП, що припадає на одного зайнятого. Даний показник рекомендований для оцінки стану ринку праці Міжнародною організацією праці. За даними МОП у 2021 р. продуктивність праці в Україні становила 14,5 дол. США на одну відпрацьовану годину (рис 2.18). Це є найнижчим показником серед досліджуваних нами країн. До прикладу в США продуктивність праці оцінювалась на рівні 70,68 дол. США у 2021 р. Нам потрібно підвищувати продуктивність праці, так як це дасть змогу покращити в цілому економічне становище держави, стан суспільства та ринку праці. Для цього потрібно створювати національні програми та стратегії по розвитку новітніх секторів економіки та розробкою якихось нових рішень для старих галузей [40].

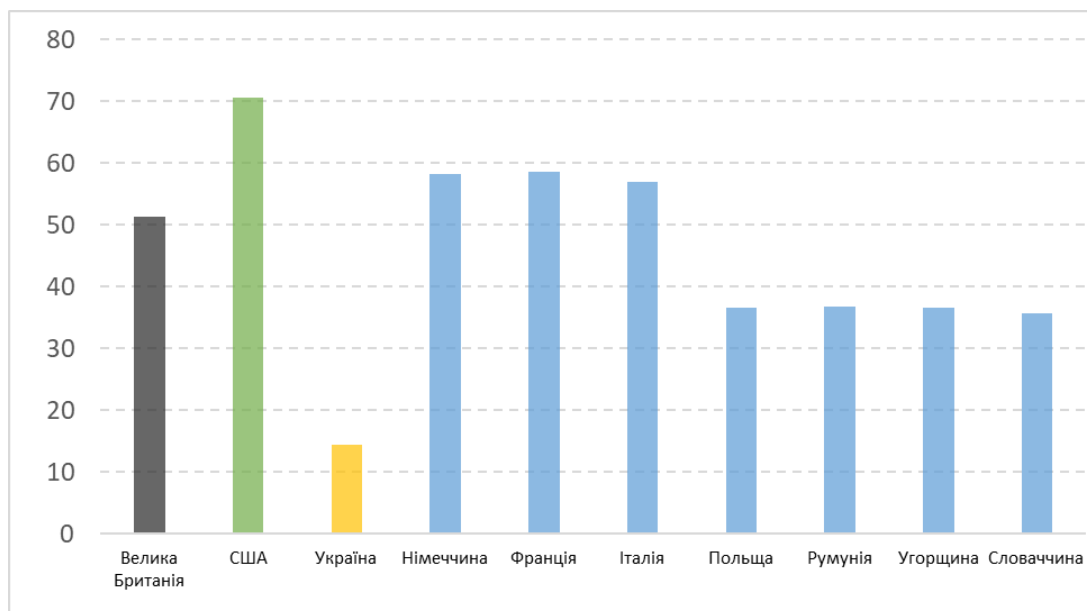


Рисунок 2.18 – Продуктивність праці (дол. США) в розрізі країн у 2021.

Джерело: розроблено автором за даними [41]

Ще одним показником оцінки продуктивності праці є індекс глобальної конкурентоспроможності. Це показник, який публікується у щорічному звіті про

глобальну конкурентоспроможність на Всесвітньому економічному форумі. В цілому цей індекс формується на базі оцінки мікро та макро аспектів бізнесу в країні. Організація оцінює державну як на основі даних з відкритих джерел так і на основі проведеного ними дослідження серед бізнес-лідерів у середині країни. Україна має 57 балів за даним індексом (рис 2.19). Даний результат є найнижчим серед досліджуваних нами країн. Як бачимо в США індекс глобальної конкурентоспроможності оцінюється в 84 бали. Тому нам потрібно в цілому покращувати своє економічне становище, а також впроваджувати рішення для підвищення рівня життя та сприяти розвитку бізнесу в країні [42].

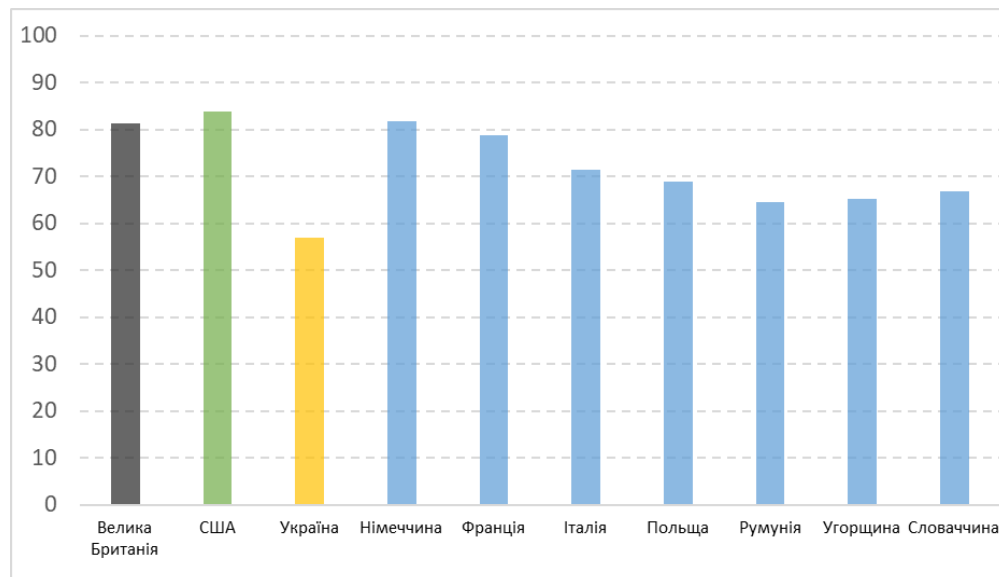


Рисунок 2.19 – Індекс глобальної конкурентоспроможності в розрізі країн у 2019 р.

Джерело: розроблено автором за даними [43]

Для якісного функціонування ринку праці потрібні працівники абсолютно різної класифікації. Проте, досить часто підприємства стикаються з браком робочої сили необхідної кваліфікації. Тому пропоную порівняти Україну та інші країни за індексом інновацій. Цей індекс розраховується Всесвітньою організацією інтелектуальної власності. Вони аналізують близько 80 різних показників, що оцінюють політичну ситуацію, систему освіти, інфраструктуру, науково-

технологічний розвиток, економічну ситуацію кожної країни [44]. За даними 2021 р. Україна мала 36 балів індексу інновацій (рис 2.20). В цілому за цим індексом Україна займає 49 місце у світі . Серед досліджуваних нами країн, Україна знаходиться на одному з останніх місць на рівні з Румунією. В цілому за даним індексом ми маємо непоганий результат. Проте потрібно далі розвивати свою інфраструктуру та впроваджувати інноваційні рішення для розвитку економіки та держави в цілому.

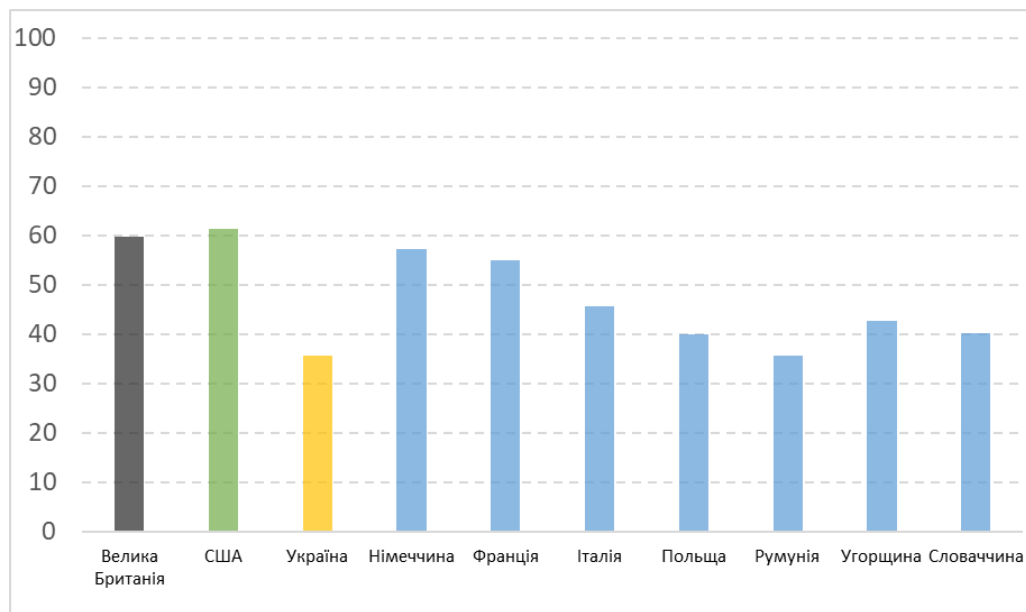


Рисунок 2.20 – Індекс інновацій в розрізі країн у 2021 р.

Джерело: розроблено автором за даними [44]

Також можемо розглянути частку робочої сили з вищою освітою по країнам. Це дасть змогу трохи краще, зрозуміти кваліфікацію працівників кожної країни. Як бачимо, в Україні за даними 2021р. було 66,5% людей, що відносяться до робочої сили з вищою освітою (рис 2.21). Проте, у Європейських країнах, США та Великій Британії цей показник вищий ніж в Україні. Як бачимо у Великій Британії показник найвищий серед досліджуваних країн та становить 84,5%. Тому можна сказати, що нам потрібно розвивати освітню систему задля того, щоб випускати кваліфікованих працівників на ринок праці.

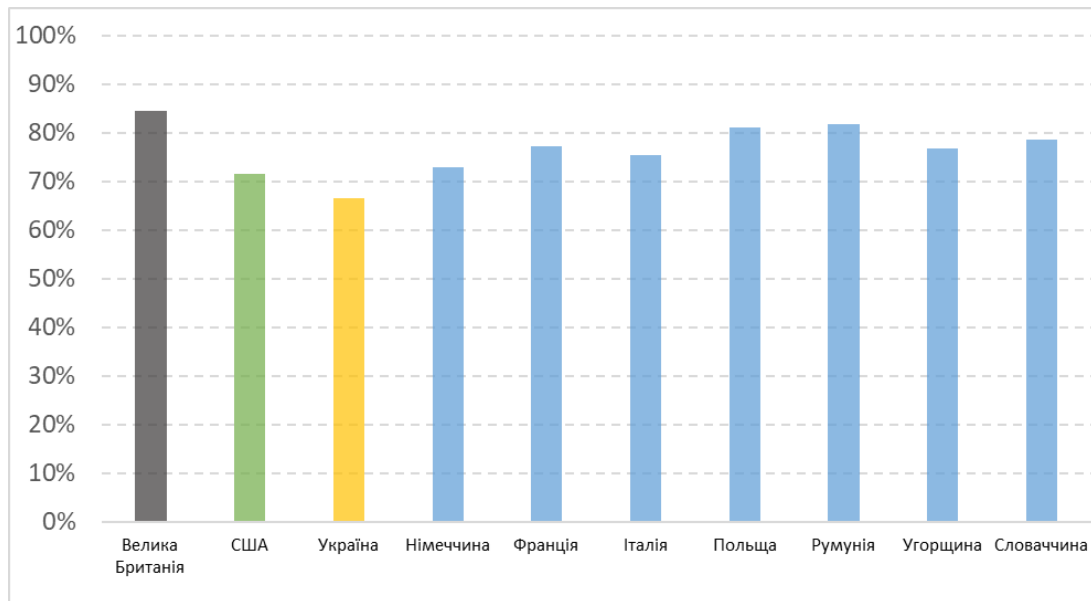


Рисунок 2.21 – Частка зайнятих з вищою освітою в розрізі країн у 2021р.

Джерело: розроблено автором за даними [36]

Отже, ми порівняли ринок праці України з країнами Європи, що є нашими сусідами, США та країнами з найбільшою економікою у світі з європейського континенту. Україна має негативний темп зміни постійного населення, що є найгіршим результатом серед досліджуваних країн. Також Україна характеризується низьким рівнем народжуваності та високою часткою населення віком 65 і старше. Внаслідок цього гостро виникає проблема старіння нації, що негативно впливає на рівень економічно активного населення. Це призводить до погіршення стану ринку праці та економіки в цілому. За ключовими показниками ринку праці, а саме рівнем зайнятості та рівнем безробіття Україна теж посідає останні місця серед досліджуваних країн та характеризується високим рівнем безробіття та низьким рівнем зайнятості. На жаль на це сильно вплинув початок повномасштабного вторгнення у 2022р. Як бачимо, в цілому українське суспільство не задоволене роботою, в основному через низьку заробітну плату, так як середня заробітна плата значно відстає від інших досліджуваних країн. На жаль, продуктивність праці України можна оцінити як низьку, а однією з причин можна назвати недостатність кваліфікованої робочої сили. В цілому нам потрібно

впроваджувати дуже багато заходів для покращення нашої освітньої система, для випуску кваліфікованої праці, економічного розвитку для задоволення потреб учасників ринку праці та впровадженню інновацій для технологічного розвитку держави. Як бачимо, український ринок праці на даний момент характеризується негативним станом ніж ринок праці Європейських країн та США. Тому нам в найближчому майбутньому нам потрібно стабілізувати та покращити стан економіки, демографії та ринку праці.

2.3 Напрями використання методів машинного навчання для аналізу та моделювання плинності працівників

Проблема плинності персоналу актуальна для кожного бізнесу. Плинність кадрів це природний процес на ринку праці. Важливо детально аналізувати цей процес на підприємстві. Адже від персоналу залежить робота будь-якого бізнесу. Одним з способів аналізу плинності кадрів є його моделювання. Для цього можна використати класифікаційні методи машинного навчання.

Ідея класифікації досить просто, це алгоритм, що прогнозує цільовий клас, аналізуючи наші вхідні дані. Наприклад ми хочемо побудувати модель, що буде прогнозувати те, чи може працівник звільнитися з підприємства в залежності від вхідних факторів. Для цього ми можемо використати різноманітні методи класифікації такі як: логістична регресія, метод наївного Байеса, дерева рішень, нейронні мережі та інші. Для побудови моделей методами машинного навчання вхідні дані зазвичай діляться на навчальну та тестову вибірку. Під час побудови модель повністю навчається на навчальних даних, а потім оцінюється на нових тестових даних [45].

Наївний класифікатор Байеса це метод створений на основі теорії Байеса. Ця теорія припускає, що наявність чи відсутність певної ознаки не пов'язана з наявністю чи відсутністю іншої ознаки та використовується для обчислень умовних ймовірностей. Умовна ймовірність розраховується за такою формулою [45] :

$$P(A|B) = \frac{P(B|A)*P(A)}{P(B)}, \quad (2.1)$$

де $P(A|B)$ — апостеріорна ймовірність класу A ;

$P(A)$ — апріорна ймовірність класу A ;

$P(B)$ — апріорна ймовірність класу B ;

$P(B|A)$ — правдоподібність, що є ймовірністю класу B .

Тобто, ми розраховуємо ймовірність події A , враховуючи, що відбувається подія B . Для цього нам потрібна ймовірність події A та ймовірність події B , а також ймовірність події B , враховуючи, що відбувається подія A . Знайшовши умовну ймовірність модель в залежності від значення відносить спостереження до класу A чи B . Цей метод можна використовувати для бінарної класифікації, тобто коли в нас лише 2 прогнозовані класи [46].

Наступним методом, який часто використовуються для задачі класифікації є метод k -найближчих сусідів. Суть цього методі полягає у виборі класу нової точки в залежності від класу найближчих сусідів. Щоб визначити, які точки є найближчими сусідами обчислюється відстань між вибраною точкою та іншими точками. Важливим попереднім етапом перед стартом моделювання є нормалізація даних, тому що метод вимірює відстань між точками, тому нам потрібні дані, що матимуть єдину шкалу вимірювання, для побудови точної моделі. Частіше всього для розрахунку відстані між точками обирають евклідову відстань. Евклідову відстань розраховують за такою формулою [47] :

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}, \quad (2.2)$$

де y_i — значення вектору цільової змінної;

x_i — значення вектору незалежної змінної.

Відповідно точки з найменшою відстанню враховуються як найближчі сусіди. Тому одним з ключових параметрів є кількість найближчих сусідів, які будуть братися в урахування. Загалом якогось спеціального алгоритму для визначення кількості найближчих сусідів немає. Для даних з більшою кількістю викидів та шуму, рекомендується брати більше число сусідів. В цілому краще всього побудувати різні моделі з різною кількістю сусідів та перевірити, яка модель буде найточнішою. Відповідно модель, що буде мати найвищу точність має оптимальну кількість найближчих сусідів [47].

Наступним методом машинного навчання, який можна використовувати при розв’язанні задачі класифікації є дерева рішень. Дерево рішень – це алгоритм що має ієрархічну структуру дерева, яке складається з кореня, гілок, внутрішніх вузлів або як ще називаються вузлів прийняття рішень та листкових вузлів (рис. 2.22). Корінь є початком дерева, відповідно він не має ніяких вхідних гілок. Далі гілки від кореня надходять у вузли прийняття рішень. Листкові вузли є закінченням дерева та представляють усі можливі результати в наборі даних. Дана ієрархія має дуже зручну інтерпретацію, що є перевагою використання цього методу. В цілому дерево рішень може мати дуже багато вузлів прийняття рішень та листкових вузлів. Проте, дуже великі дерева рішення погіршують точність та інтерпретацію. Щоб зменшити складність дерева і запобігти перенавчанню існує обрізка дерев, яка видаляє непотрібні гілки [48].

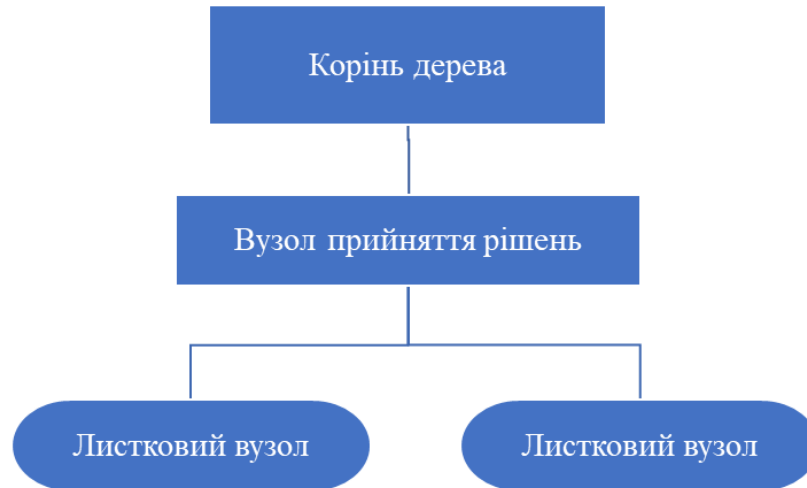


Рисунок 2.22 – Ієрархічна структура дерева рішень

Джерело: розроблено автором за даними [41]

Щоб побудувати дерево рішень спочатку використовується рекурсивне двійкове розбиття. Цей алгоритм застосовується на навчальних даних, для визначення оптимальних точок поділу всередині дерева. Для цього потрібно обрати критерій за яким дерево рішення буде здійснювати розбиття на гілки. Одним з критеріїв поділу дерева рішень є ентропія. Це критерій, що вимірює міру ступеня випадковості або невизначеності в даних. Ентропія розраховується за формулою [48] :

$$E = \sum_{i=1}^c -p_i(c) \log_2 p_i(c), \quad (2.3)$$

де c - класи в наборі обраних даних;

$p_i(c)$ – частка точок даних, що належать до класу c , до загальної кількості точок даних у наборі.

Значення ентропії коливаються від 0 до 1. Якщо в наборі даних присутній лише один клас, то ентропія буде дорівнювати 0. Щоб вибрати найкращий атрибут для розділення нашого дерева, потрібно вирахувати ентропію до поділу даних за вибраним атрибутом та після поділу даних за вибраним атрибутом. Атрибут за яким буде найвищий приріст ентропії, буде найкращим атрибутом для поділу [48].

Ще одним критерієм поділу дерева на вузли є індекс Джинні. Цей індекс є ймовірністю неправильної класифікації випадкових точок даних у нашому наборі. Індекс Джинні розраховується за такою формулою [48] :

$$G = 1 - \sum_{i=1}^c (p_i(c))^2, \quad (2.4)$$

де p_i – частка точок даних, що належать i -го класу.

Після побудови моделей за різними методами машинного навчання, важливо правильно оцінити точність моделей. Існує багато різних показників оцінки точності моделей. Для оцінки моделей класифікації важливо побудувати матрицю відповідностей. Це таблиця з комбінаціями правильних та помилкових прогнозованих і фактичних значень (табл 2.1). За цією матрицею можна оцінити скільки правильних та помилкових прогнозів було зроблено моделлю за тестовою вибіркою [49].

Таблиця 2.1 – Матриця відповідностей класифікаційної моделі [49]

Фактичні / Прогнозовані	Так	Ні
Так	Правильні Позитивні (TP)	Помилкові Негативні (FN)
Ні	Помилкові Позитивні (FP)	Правильні Негативні (TN)

На основі матриці відповідностей можна розрахувати додатково різні показники точності. Точність класифікація розраховується за формулою [50]:

$$\text{Асс} = \frac{TP+TN}{TP+TN+FP+FN}, \quad (2.5)$$

де TP – кількість правильно позитивних результатів;

TN – кількість правильно негативних результатів;

FP – кількість помилко позитивних результатів;

FN – кількість помилково негативних результатів.

Точність класифікації є часткою правильних прогнозів серед усіх спостережень. Цей показник корисний, коли у нас класи мають приблизно однакову кількість спостережень. Наприклад в наборі даних 70% спостережень належать класу А і 30 % спостережень класу Б. Тоді ми можемо отримати 70% точності правильно передбачивши лише один клас. Хоча нам важливо, щоб модель була максимально точною при прогнозуванні обох класів. Додатково існують різні показники точності [50].

Точність прогнозу позитивного класу є часткою правильно передбачених випадків, які належали до позитивного класу від усіх спостережень позитивного класу. Цей показник важливий, коли хибно-позитивний результат більше впливає ніж хибно-негативний результат. Наприклад якщо класифікаційна модель віднесла клієнта банку, до класу, що розраховалися за кредит вчасно, хоча насправді цей клієнт не сплатив заборгованість. Така помилка моделі може завдати шкоди бізнесу. Тому в даному випадку нам краще перевіряти показник точності прогнозу позитивного класу. Цей показник розраховується за формулою [50] :

$$Precision = \frac{TP}{TP+FP} \quad (2.6)$$

де TP – кількість правильно позитивних результатів

FP – кількість помилково позитивних результатів.

Чутливість або точність відображає відсоток правильно передбачених фактичних позитивних випадків. Цей показник важливий, коли хибно-негативний результат більше впливає ніж хибно-позитивний результат. Наприклад якщо класифікаційна модель віднесла пацієнта з смертельним захворюванням, до класу здорових пацієнтів. Для цього нам потрібно перевірити чутливість моделі, яка показує кількість таких прогнозованих випадків. Цей показник розраховується за формулою [50] :

$$Recall = \frac{TP}{TP+TN}, \quad (2.7)$$

де TP – кількість правильно позитивних результатів;

TN – кількість правильно негативних результатів.

Наступним показником оцінки точності моделі є показник F1. Цей показник є середнім гармонійним показників точності правильних позитивних результатів та чутливості. Він показує наскільки класифікатор є точним, а також надійним, тобто не пропускає значну кількість неправильних випадків. Чим вищий даний показник, тим надійнішою та кращою є наша побудована модель. Показник F1 розраховується за наступною формулою [51] :

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (2.8)$$

де Precision – частка правильно передбачених випадків з усіх випадків позитивного класу;

Recall – точність передбачення позитивних випадків.

Далі розглянемо показник площі під кривою помилок (AUC). Крива помилок відображає частоту правильного позитивного результату проти частоти помилкового позитивного результату. Ці показники розраховуються при різних порогових значеннях та будується графік. Площа під цією кривою є нашим показником AUC. Чим більше значення площі, тим вища частота правильних позитивних прогнозів. Відповідно наша модель правильно розрізняє обидва класи та є точною [50].

Отже, побудова моделей часового ряду є важливим в дослідженні динамічних економічних процесів. Ринок праці представляє одну з задач, яку можна прогнозувати за допомогою часових рядів. Для моделювання процесів на підприємстві доцільно використовувати методи машинного навчання. Однією з частих задач підприємства є дослідження плинності персоналу. Для цього можна

використовувати методи класифікації такі як: дерева рішення, метод Наївного Байеса, k -найближчих сусідів та інші. Моделювання дає можливість глибоко проаналізувати процеси як на підприємстві так і в державі.

РОЗДІЛ 3

ПРАКТИЧНЕ ЗАСТОСУВАННЯ МОДЕЛЕЙ АНАЛІЗУ ДИНАМІКИ ТА ПРОГНОЗУВАННЯ РИНКУ ПРАЦІ УКРАЇНИ

3.1 Аналіз результатів моделювання ринку праці України на базі одномірних часових рядів

Спостереження послідовні в часі є часовими рядами. Для того, щоб оцінити майбутній стан ринку праці застосуємо методи прогнозування часових рядів. Для моделювання ринку праці ми будемо використовувати дані по зайнятості, безробіттю, економічній активності, валового внутрішнього продукту, інвестицій, інфляції, заробітної плати. Цільовою змінною для моделювання буде виступати рівень зайнятості. Цей показник якісно характеризує стан ринку праці. Тому, буде доцільно отримати прогноз даного показника на найближче майбутнє.

Першим етапом є збирання та обробка даних. Державна служба статистики займається збором даних ринку праці, стану економіки, демографії та інших важливих аспектів існування держави. Тому, з сайту Державної служби статистики були отримані щоквартальні дані по оцінці рівня зайнятості за період 2003 та 2021 рр. Для отримання актуальних прогнозів, потрібно було отримати дані за 2022р. Через початок повномасштабного вторгнення у 2022 р. Держстат не проводив оцінку рівня зайнятості. Тому з альтернативних джерел оцінки ринку праці (НБУ, дослідницька організація Центр економічної стратегії) були отримані поквартальні дані по рівню зайнятості за 2022 р. Факторами, що можуть впливати на рівень зайнятості було визначено рівень безробіття, середня заробітна плата, індекс споживчих цін, номінальний ВВП, прямі інвестиції та рівень економічно активного населення. Дані по незалежним змінним були відібрані за період 2003-2022 р. з

поквартальною періодичністю. Основним джерелом даних був Держстат України, джерела оцінки стану ринку в 2022 р. Центр економічної стратегії, Український інститут майбутнього.

Для проведення моделювання ринку праці буде використовуватись мова програмування R. Після отримання вхідних даних по рівню зайнятості потрібно їх завантажити в наше середовище програмування. Для цього ми використаємо набір пакетів Tidyverse, який складається з бібліотек, для моделювання, перетворення та обробки, візуалізації даних. Дані завантажилися у форматі tibble, що є фактично звичайною таблицею, яка складається зі стовбчиків та рядків. Так як наші дані є часовим рядом, тобто послідовні в часі, нам потрібно перетворити їх у форматі tsibble. Це об'єкт даних, що зберігає часові ряди в мові програмування R. Для цього потрібно використати набір пакетів frr3, що складається з бібліотек для роботи з часовими рядами.

Спочатку проведемо однофакторне моделювання рівня зайнятості. Завантаживши наші дані у форматі tsibble ми можемо приступити до етапу аналізу часових рядів. За допомогою функції autoplot() побудуємо візуалізацію часового ряду рівень зайнятості. За допомогою такого графіку можна оцінити тренд та сезонність, а також в цілому динаміку ряду. Тренд та сезонність це ключові компоненти часового ряду. Тренд або тенденція спостерігається, коли відбувається тривале зниження чи зростання. Сезонність виникає, коли на часовий ряд впливає, якийсь фактор часу, тобто пора року, день тижня, місяць, квартал. Сезонність повинна мати фіксований період. На даному графіку, є певні коливання в залежності від кварталу, проте вони не постійні, тому сезонності немає. В цілому були 3 періоди часу, коли рівень зайнятості мав значне падіння (рис 3.1). Це початок 2009 р. де рівень зайнятості знизився на 2%. Наступним періодом спаду був початок 2014 р. де рівень зайнятості знизився на 4%, а також початок 2022 р. де рівень знизився майже на 10%. В цілому часовий ряд має спадаючий тренд з незначним зростанням в деякі періоди.

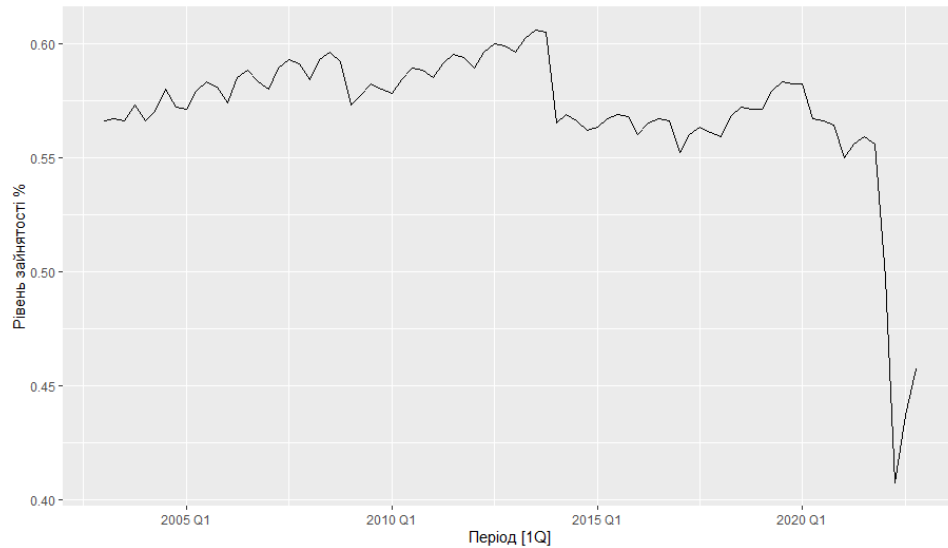


Рисунок 3.1 – Рівень зайнятості по кварталам за період 2003 – 2022 рр.

Джерело: розроблено автором за даними [26]/[33]

Для перевірки сезонності за допомогою команди `gg_season` можна побудувати графік в розрізі сезонів. Як бачимо, в цілому на кінець року рівень зайнятості зазвичай трохи вище ніж на початок (рис 3.2). Також, можна зробити висновок, що 2022 р. має найнижчий рівень зайнятості. 2012-2013 рр. характеризуються найвищими показниками рівня зайнятості в Україні.

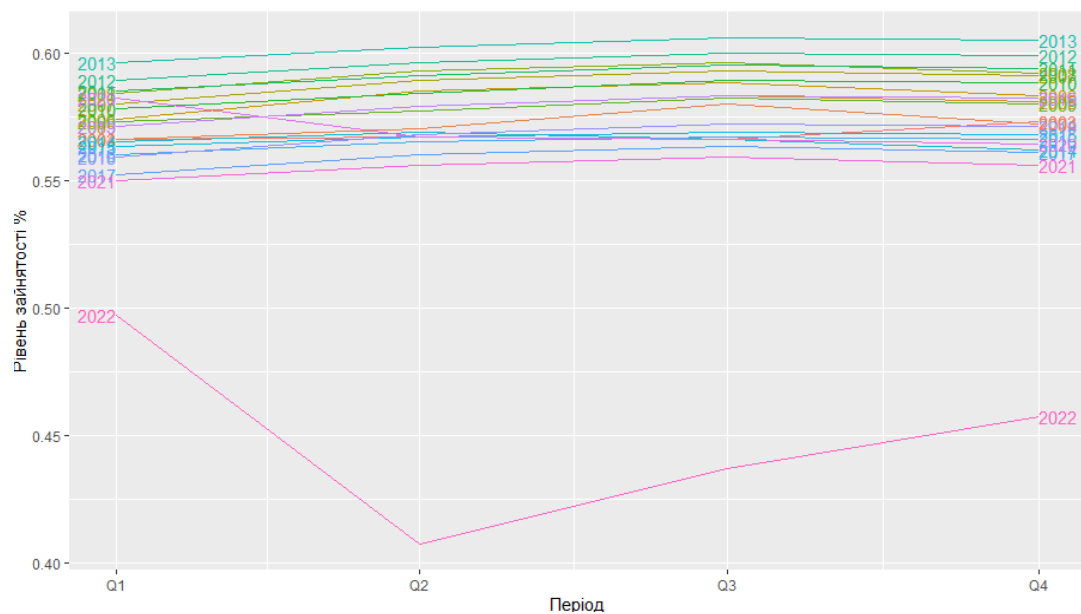


Рисунок 3.2 – Рівень зайнятості в розрізі кварталів за період 2003 – 2022 рр.

Джерело: розроблено автором за даними [26]/[33]

Для того, щоб зрозуміти основні закономірності часового ряду проведемо його декомпозицію. В даному випадку застосуємо метод декомпозиції STL. Цей метод базується на адитивній декомпозиції. Тобто має такий вигляд:

$$\text{Рівень зайнятості} = \text{Тренд} + \text{Сезонність} + \text{Залишки}, \quad (3.1)$$

За результатом декомпозиції з графіку компонент можна сказати, що часовий ряд має сильний спадаючий тренд (рис 3.3). Також часовий ряд має сильну сезонність. Залишки розподіляється на рівні 0, а також не мають ніякої залежності, що говорить про те, що декомпозиція проведена якісно.

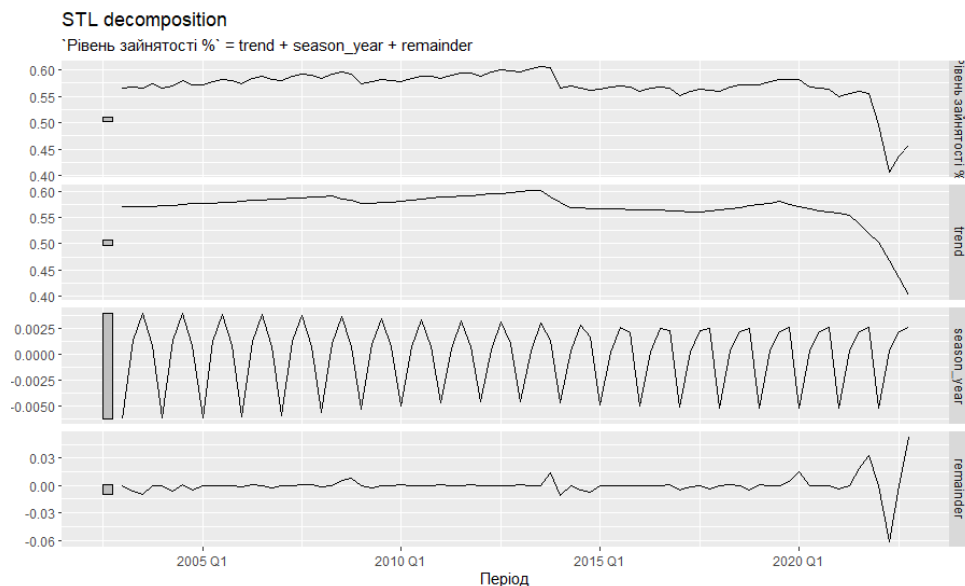


Рисунок 3.3 – Декомпозиція часового ряду рівень зайнятості методом STL

Джерело: розроблено автором за даними [26]/[33]

Важливим дослідженням часового ряду, є вивчення автокореляції. Автокореляція вимірює лінійний зв'язок всередині часового ряду. Для оцінки автокореляції побудуємо її графік. За графіком автокореляції можна визначити сильний тренд. Часовий ряд має статистично значиму автокореляцію на перших 4 лагах (рис 3.4). Тому ми можемо застосувати авторегресійні методи моделювання для прогнозування нашого часового ряду.

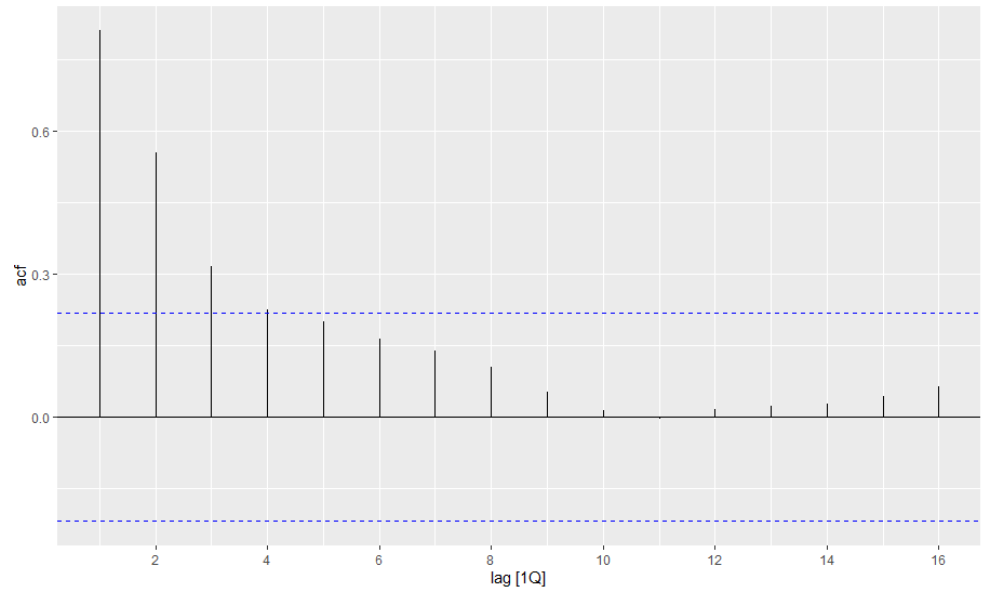


Рисунок 3.4— Графік автокореляції часового ряду рівень зайнятості

Джерело: розроблено автором за даними [26]/[33]

Одним з важливих припущень є перевірка часового ряду на стаціонарність. Це робиться для того, щоб наші висновки були надійними. Стаціонарний часовий ряд має постійну дисперсію, середнє значення та автокореляцію. Для перевірки рядку на стаціонарність існують статистичні тести. Ми використаємо тест Дікі-Фуллера та тест KPSS. За результатом тесту Дікі-Фуллера та тесту KPSS часовий ряд не стаціонарний (рис 3.5). Нам потрібно зробити перетворення часового ряду, щоб отримати стаціонарний часовий ряд. Перетворенням може бути логарифмування часового ряду або диференціювання.

```
Augmented Dickey-Fuller Test
data: data$`Рівень зайнятості %`
Dickey-Fuller = 0.074101, Lag order = 4, p-value = 0.99
alternative hypothesis: stationary

kpss_stat kpss_pvalue
<dbl>      <dbl>
1      0.734      0.0104
```

Рисунок 3.5 – Результат тестів Дікі-Фуллера та KPSS на стаціонарність часового ряду рівень зайнятості

Джерело: розроблено автором за даними [26]/[33]

Використавши функцію `unitroot_ndiffs()` можна визначити, який порядок різниць у часовому ряді потрібно взяти, для приведення його до стаціонарного. В результаті, якщо взяти та порахувати різницю 1 порядку в нашому часовому ряді, то він стане стаціонарним (рис 3.6). Перевіривши за тестом KPSS ряд з різницею першого порядку є стаціонарним. Тому тепер ми можемо його використати для побудови авторегресійних моделей.

```
ndiffs
<int>
1      1
kpss_stat kpss_pvalue
<dbl>    <dbl>
1      0.296      0.1
```

Рисунок 3.6 – Результат тесту на оптимальний порядок різниць та тесту KPSS

Джерело: розроблено автором за даними [26][33]

Спочатку розділимо наш часовий ряд на тестову та навчальну вибірку. Дані за період 2003-2018 рр будуть навчальною вибіркою, а ряд з періодичністю 2019-2022 рр є нашою тестовою вибіркою. Для моделювання часового ряду рівень зайнятості використаємо методи авторегресії, експоненціальне згладжування та методологію Prophet. ARIMA або як ще називаються авторегресійні методи використовують лінійну комбінацію з минулих значень нашого ряду. Модель включає 3 компоненти а саме авторегресія AR, інтегрування I, ковзне середнє MA. Для визначення параметрів компонент моделі існує 2 ключові способи. Можна побудувати графік часткової та звичайної автокореляції або побудувати модель з автоматичним вибором параметрів за алгоритмом Хайндмана-Кхандакара. Як бачимо з лагом 2 автокореляція статистично значима, а також з лагом 2 часткова автокореляція є статистично значимою (рис 3.7). Тому побудуємо модель AR (2,1,0), тобто спочатку були взяті різниці першого порядку з нашого часового ряду, щоб привести його до стаціонарності з порядком затримки 2 спостереження. Ще додатково побудуємо модель MA (0,1,2) тобто модель ковзного середнього з

порядком відхилення 2 спостереження. Також додатково використаємо метод автоматичного підбору параметрів.

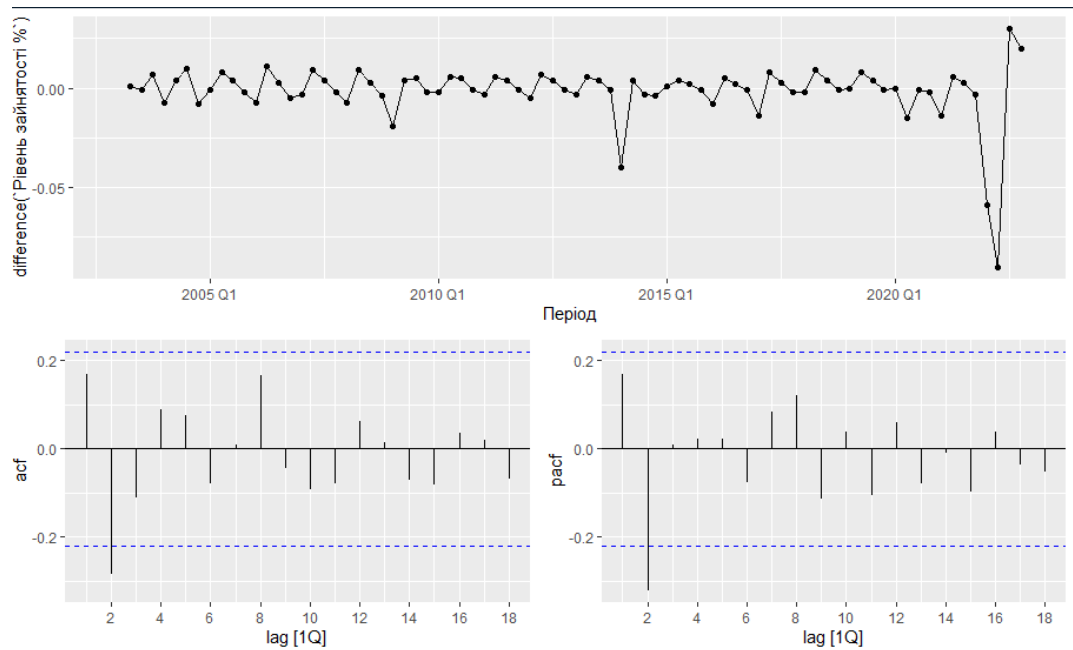


Рисунок 3.7 – Графік часткової та повної автокореляції часового ряду «рівень зайнятості»

Джерело: розроблено автором за даними [26][33]

Для побудови моделі використаємо функцію `model()` з бібліотеки `fabletools`. В результаті були побудовані 3 моделі AR (2,1,0) MA(0,1,2) SAR (1,0,0)(2,1,0). Для вибору моделі потрібно порівняти між собою інформаційні критерії Акаїке та Байеса. Як бачимо модель з автоматичним підбором параметрів SARIMA має нижчі значення критерію Акаїке та Байеса, тому вона є кращою для прогнозування нашого ряду (рис 3.8).

	MA	AR	auto_season
	<model>	<model>	<model>
1	<ARIMA(0,1,2)>	<ARIMA(2,1,0)>	<ARIMA(1,0,0)(2,1,0)[4]>
	.model	sigma2 log_lik	AIC AICc BIC
	<chr>	<dbl>	<dbl> <dbl> <dbl>
1	MA	0.0000554	220. -435. -434. -428.
2	AR	0.0000539	221. -436. -436. -430.
3	auto_season	0.0000398	219. -429. -428. -421.

Рисунок 3.8 – Конфігурація побудованих авторегресійних моделей та їх інформаційні критерії

Джерело: розроблено автором за даними [26][33]

Далі потрібно перевірити нашу найякіснішу модель SARIMA на незалежність залишків. Якщо залишки є незалежними та відповідно мають вигляд білого шуму, це означає, що модель дійсно якісна та може застосовуватись для прогнозу. Для цього проведемо тест Льюнга-Бокса та побудуємо корелограму залишків. За результатом тесту $p\text{-value} = 0,444$, тобто можна відхилити альтернативну гіпотезу, про те, що залишки є залежними. За корелограмою в цілому можна сказати, що залишки схожі на білий шум. Тобто наша модель якісно використали дані для побудови.

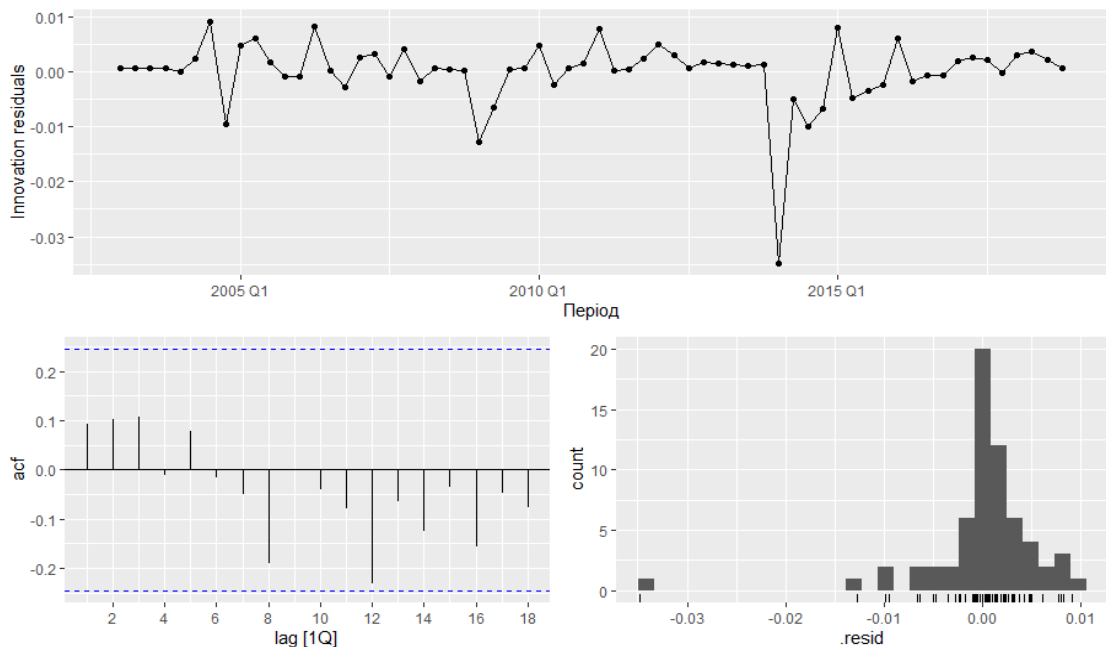


Рисунок 3.9 – Автокореляція та розподіл залишків моделі SARIMA часового ряду рівень зайнятості

Джерело: розроблено автором за даними [26][33]

Побудуємо модель ETS це метод, що оцінює тренд, сезонність та залишки, та порівняємо з іншими моделями. Також додатково побудуємо модель за методологією Prophet та порівняємо результати усіх моделей. Модель ETS була побудована з мультиплікативними залишками та сезонністю. В результаті моделювання модель за методом Prophet виявилась найточнішою з RMSE 0,0615 (рис 3.10).

.model	.type	ME	RMSE	MAE
<chr>	<chr>	<dbl>	<dbl>	<dbl>
1 ETS	Test	-0.0307	0.0631	0.0380
2 auto_season	Test	-0.0290	0.0627	0.0379
3 prophet	Test	-0.0278	0.0623	0.0367

Рисунок 3.10 – Результат оцінки точності моделей побудованих методом ETS, SARIMA, Prophet

Джерело: розроблено автором за даними [26]/[33]

Потрібно перевірити нашу модель методом Prophet, яка має найкращу точність на незалежність залишків. Для цього проведемо тест Льюнга-Бокса та побудуємо корелограму залишків. За результатом тесту $p\text{-value} = 0,0533$, доволі низький результат, проте можна прийняти нульову гіпотезу, про те, що залишки є незалежними. Як бачимо з корелограми залишки мають статистично значиму автокореляцію з лагом 4, що є негативним для якості нашої моделі (рис 3.11). Проте тест Льюнга Бокса говорить про незалежність залишків.

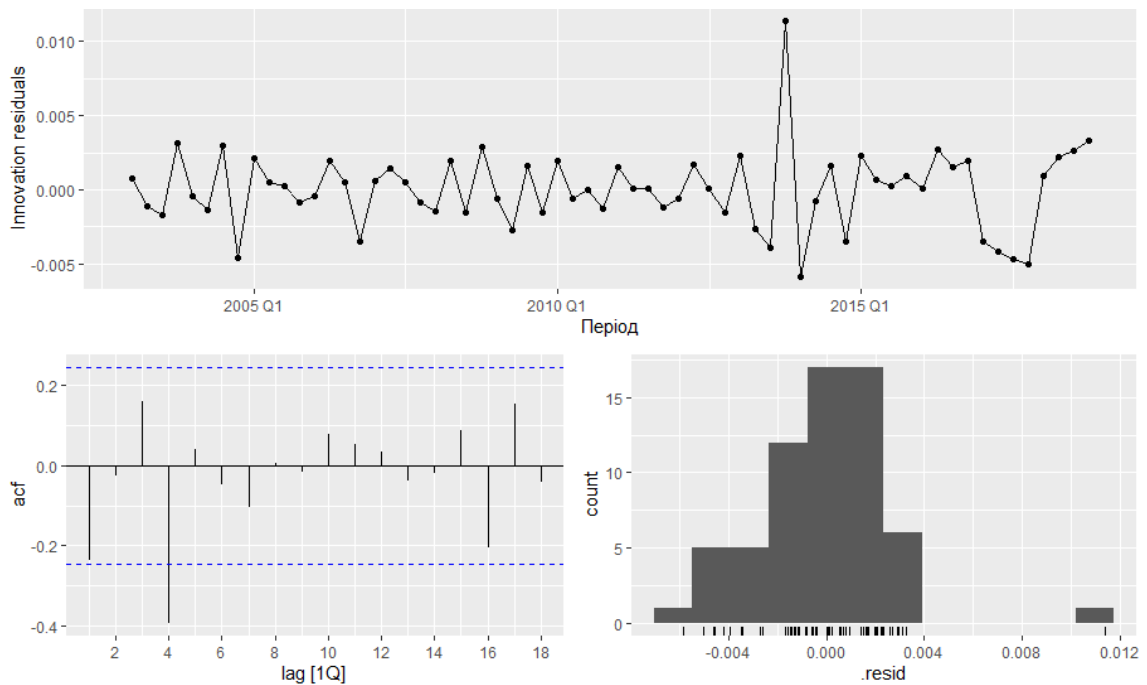


Рисунок 3.11 – Графік розподілу залишків побудованої моделі методом Prophet

Джерело: розроблено автором за даними [26]/[33]

В результаті моделювання моделі за методами SARIMA, ETS, Prophet показали доволі схожу точність. Як бачимо з графіку прогноз за усіма методами, має подібну структуру (рис 3.12). Проте, модель не прогнозує значний спад у 2022 р. через початок повномасштабного вторгнення. На жаль оцінити такий фактор досить складно. Проте до 2022 р. наші моделі мають достатньо низьке відхилення у порівнянні з фактичними значеннями. За показниками MAE, RMSE все ж модель Prophet має незначну перевагу. Тому ця модель є найкращою серед побудованих нами.

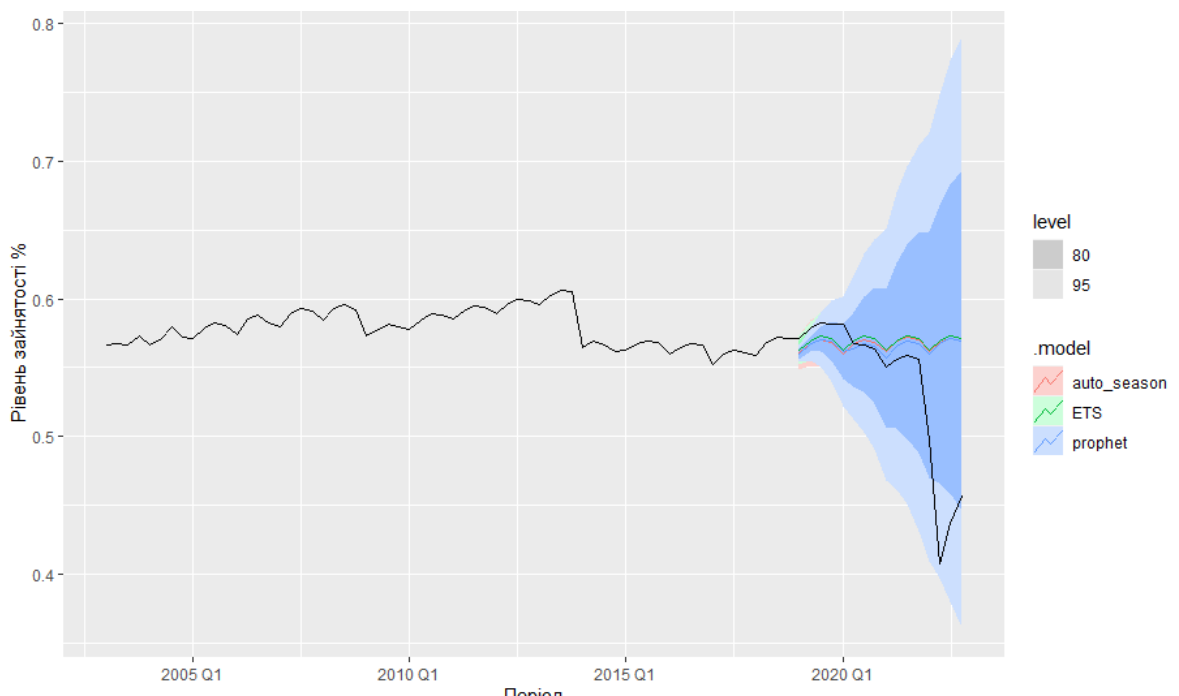


Рисунок 3.12 – Графік фактичних та прогнозованих значень за методами ETS, SARIMA, Prophet

Джерело: розроблено автором за даними [26][33]

3.2 Аналіз результатів багатфакторного моделювання динаміки ринку праці України

Наступним етапом є моделювання рівня зайнятості на основі залежності від інших факторів. Пропоную провести моделювання рівня зайнятості в залежності від наступних факторів: рівень безробіття, середня заробітна плата, індекс споживчих цін, прямі іноземні інвестиції, рівень економічно активного населення та номінальний ВВП. Для цього дані по вказаним факторам були зібрані за таким же періодом, як і рівень зайнятості. Після цього варто переглянути ступінь зв'язку між незалежною змінною та вказаними факторами. Для цього побудуємо корелограму часових рядів. З корелограми часових рядів можна зробити висновок, що рівень зайнятості має тісний негативний зв'язок з рівнем безробіття, середньою заробітною платою, номінальним ВВП а також тісний позитивний зв'язок з показником рівень економічно активного населення рис(3.13). З показниками індекс споживчих цін та прямі іноземні інвестиції, рівень зайнятості має низьку кореляцію. Тому в першу чергу побудуємо моделі з факторами, з якими існує найтісніший зв'язок.

Для побудови багатфакторної моделі часових рядів будемо використовувати метод векторної авторегресії (VAR). Цей метод базується на дослідженні авторегресії в даних проте з додаванням лагових векторів інших факторів. Для цього нам потрібні лише стаціонарні часові ряди. Тому нам потрібно визначити оптимальну кількість лагових затримок для нашої моделі. Для цього існує функція, яка розраховує інформаційні критерії такі як: критерій Акаїке (AIC), Ханна-Квінна (HQ), Байеса (SC або BIC) та скоригований критерій Акаїке (FPE). Спочатку будемо будувати модель рівня зайнятості в залежності від рівня безробіття. Для цієї моделі оптимально буде обрати вектори з лаговою затримкою

1 так як з цією затримкою маємо найменші значення інформаційних критеріїв (рис 3.14).

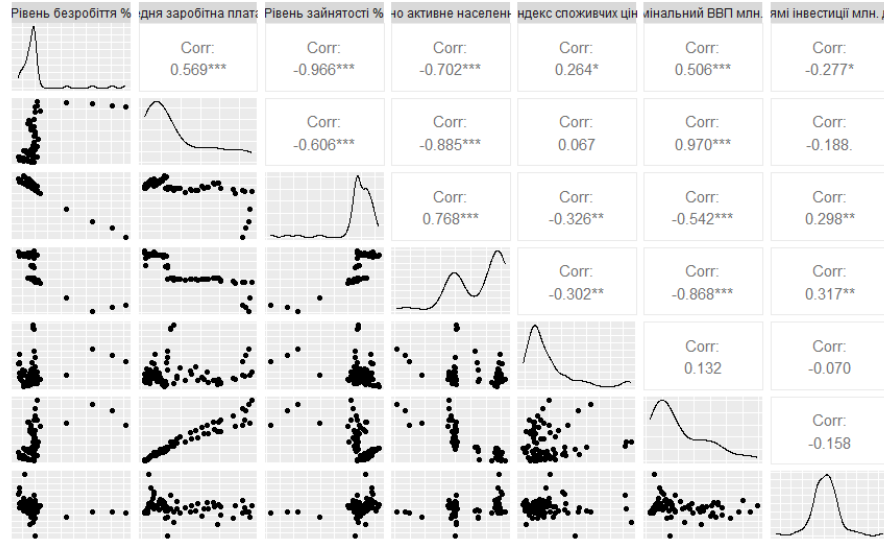


Рисунок 3.13 – Матриця кореляцій часових рядів

Джерело: розроблено автором за даними [26][33][35][52]

\$selection						
AIC(n)	HQ(n)	SC(n)	FPE(n)			
1	1	1	1			
\$criteria						
	1	2	3	4	5	6
AIC(n)	-2.086551e+01	-2.073445e+01	-2.071513e+01	-2.059000e+01	-2.079213e+01	-2.064954e+01
HQ(n)	-2.078028e+01	-2.059240e+01	-2.051626e+01	-2.033431e+01	-2.047962e+01	-2.028021e+01
SC(n)	-2.064451e+01	-2.036612e+01	-2.019947e+01	-1.992700e+01	-1.998181e+01	-1.969188e+01
FPE(n)	8.676109e-10	9.899338e-10	1.011131e-09	1.149750e-09	9.443040e-10	1.097462e-09
	7	8	9	10		
AIC(n)	-2.053212e+01	-2.039901e+01	-2.047190e+01	-2.034164e+01		
HQ(n)	-2.010597e+01	-1.991603e+01	-1.993211e+01	-1.974503e+01		
SC(n)	-1.942713e+01	-1.914668e+01	-1.907225e+01	-1.879465e+01		
FPE(n)	1.247482e-09	1.445599e-09	1.369131e-09	1.596850e-09		

Рисунок 3.14 – Інформаційні критерії по лаговим затримкам

Джерело: розроблено автором за даними [26][33][35][52]

В результат побудовано модель залежності рівня зайнятості від рівня безробіття методом векторної авторегресії з лаговими затримками першого порядку. Модель є статистично значимою за критерієм Фішера, так як p-value значно менше 0,05 (рис 3.15). Коефіцієнт детермінації даної моделі складає 0,7, що говорить про доволі високу точність прогнозу. Середньо квадратичне відхилення моделі складає 0,045, що говорить про не високу похибку прогнозу моделі.

```

Call:
lm(formula = y ~ -1 + ., data = datamat)

Residuals:
    Min       1Q   Median       3Q      Max
-0.036334 -0.002945  0.000664  0.004590  0.011215

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
Рівень.зайнятості...l1  0.92355    0.10941   8.441 8.7e-12 ***
Рівень.безробіття...l1  0.15465    0.13832   1.118  0.268
const                 0.03068    0.07309   0.420  0.676
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.007377 on 60 degrees of freedom
Multiple R-squared:  0.7011,    Adjusted R-squared:  0.6912
F-statistic: 70.38 on 2 and 60 DF,  p-value: < 2.2e-16

```

Рисунок 3.15 – Результат моделювання методом VAR

Джерело: розроблено автором за даними [26]/[33]/[35]/[52]

Наступним етапом є перевірка залишків побудованої моделі. Для цього проведемо тест Портманто на автокореляцію залишків. За тестом Портманто можна відхили нульову гіпотезу про те, що залишки мають статистично значиму автокореляцію. Також ми можемо оцінити розподіл залишків з графіку розподілу. Як бачимо в цілому залишки мають середнє значення на рівні 0, а також виглядають як білий шум (рис 3.16). Також пропонуємо оцінити причинний зв'язок Грейнджера у нашій моделі. Цей тест допоможе нам оцінити чи впливає рівень безробіття на рівень зайнятості на певному лазі. Отже за результатом теста причинності Грейнджера, не можна відхили нульову гіпотезу про те, що рівень безробіття має значний вплив на прогнозування рівня зайнятості.

В цілому оцінка точності моделі, говорить про те, що вона має доволі низьке відхилення, а за критерієм Фішера модель статистично значима. Тому побудуємо прогноз за даною моделлю та переглянемо його графік. Як бачимо прогнозовані значення доволі сильно відхиляються на графіку від фактичних значень (рис 3.17). Проте, все ж до 2022 р. фактичні значення знаходяться в межах довірчих інтервалів прогнозу, що є позитивним чинником. Побудуємо моделі з іншими факторами та перевіримо їх вплив на рівень зайнятості. Для цього пропонуємо додавати кожен

фактор до початкової моделі та оцінювати зміну точності прогнозу та статистичної значимості моделі.

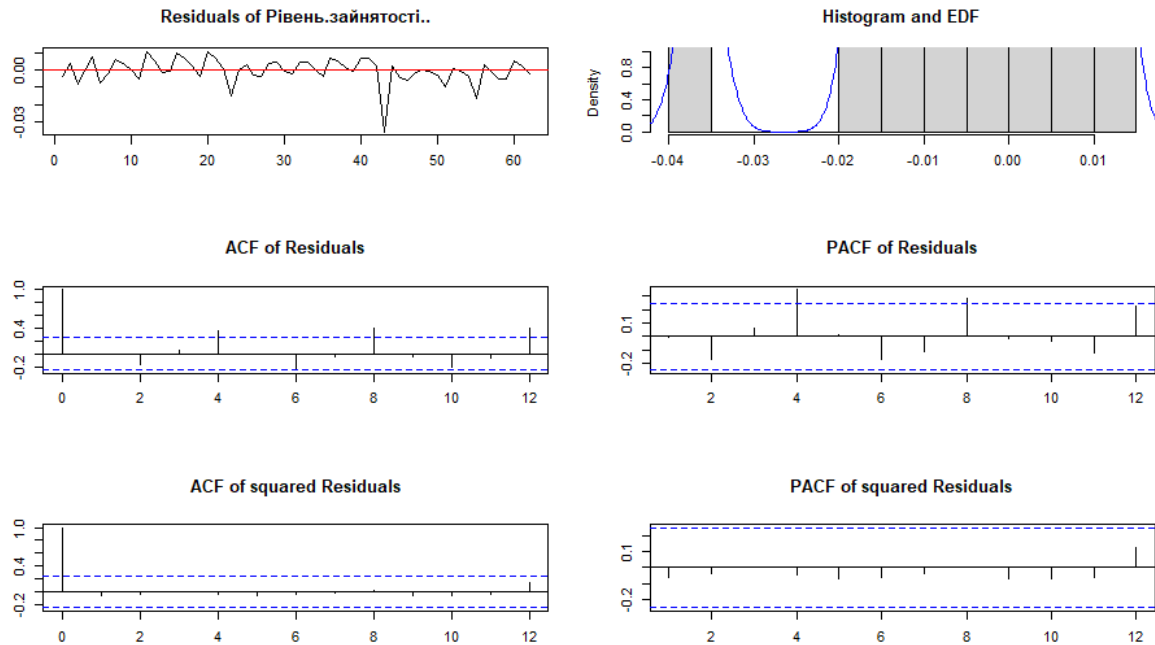


Рисунок 3.16 – Графік автокореляцій та розподілу залишків моделі

Джерело: розроблено автором за даними [26][33][35][52]

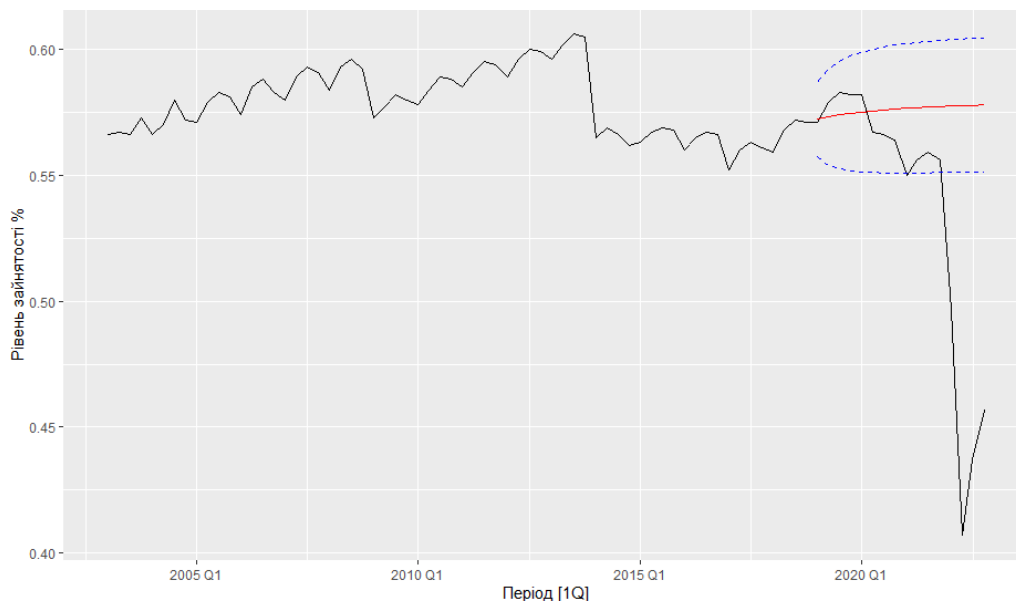


Рисунок 3.17 – Графік фактичних та прогнозованих значень рівня зайнятості

Джерело: розроблено автором за даними [26][33][35][52]

Отже, побудовано 6 моделей в які поступово додавалося по одному фактору. В цілому усі моделі статистично значимі (табл 3.2). Проте, додавання нових факторів лише позитивно вплинули на точність моделей. Як бачимо, модель залежності рівня зайнятості від рівня безробіття, економічно активного населення, середньої заробітної плати, індексу споживчих цін, прямих інвестицій та номінального ВВП. Отже ця модель має коефіцієнт детермінації на рівні 92%, а також середнє квадратичне відхилення 0,044, що є найменшим серед усіх побудованих моделей.

Таблиця 3.2. Результати та модифікації побудованих моделей

Модель	Кількість лагів	RMSE	Значимість за F критерієм
Рівень зайнятості ~ Рівень безробіття	1	0,045	Статистично значима
Рівень зайнятості ~ Рівень безробіття + Економічно активне населення	5	0,058	Статистично значима
Рівень зайнятості ~ Рівень безробіття + Економічно активне населення + Середня заробітна плата	5	0,046	Статистично значима
Рівень зайнятості ~ Рівень безробіття + Економічно активне населення + Середня заробітна плата + Індекс споживчих цін	5	0,044	Статистично значима

Продовження табл 3.2

Рівень зайнятості ~ Рівень безробіття + Економічно активне населення + Середня заробітна плата + Індекс споживчих цін + Прямі інвестиції	5	0,045	Статистично значима
Рівень зайнятості ~ Рівень безробіття + Економічно активне населення + Середня заробітна плата + Індекс споживчих цін + Прямі інвестиції + Номінальний ВВП	6	0,044	Статистично значима

Джерело: розроблено автором

Далі потрібно перевірити залишки на автокореляцію та розподіл. За результатом тесту Портманто, можна відхили нульову гіпотезу про те, що залишки мають статистично значиму автокореляцію. Також можна сказати, що розподіл залишків має структуру білого шуму, а середнє значення залишків на рівні 0 (рис 3.18). Це говорить про те, що наша модель є якісно побудованою.

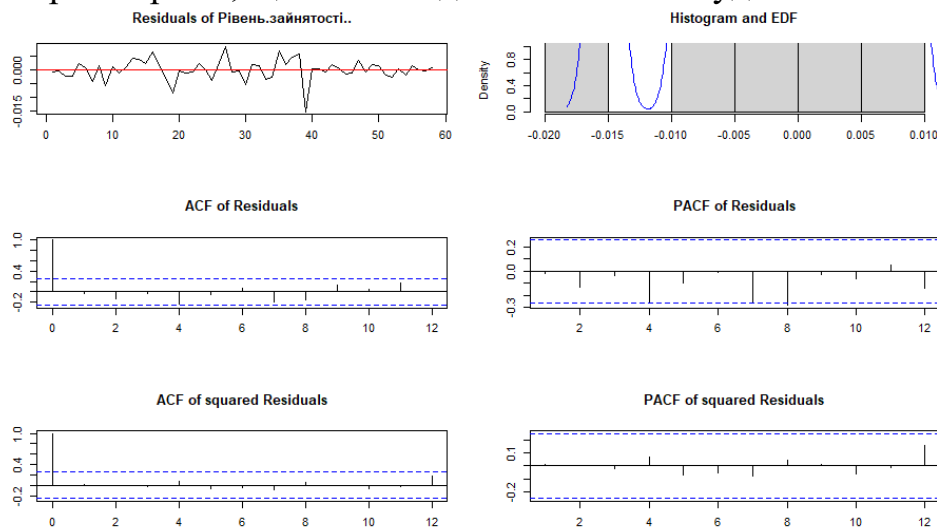


Рисунок 3.18 – Графік автокореляцій та розподілу залишків моделі

Джерело: розроблено автором за даними [26][33][35][52]

Далі нанесемо прогнозовані значення та довірчі інтервали на графік та подивимось на їх структуру та різницю. З рисунку 3.19 видно, що прогнозовані значення мають доволі схожу тенденцію з фактичними значеннями. Також дана модель набагато краще спрогнозувала значне падіння рівня зайнятості в 2022р. Отже в результаті проведених моделювань модель з усіма визначеними раніше факторами є найточнішою та найоптимальнішою. Дану модель можна використовувати для побудови прогнозу на наступні періоди. Як бачимо модель прогнозує рівень зайнятості на 1 квартал 2023 р. на рівні 41,7%.

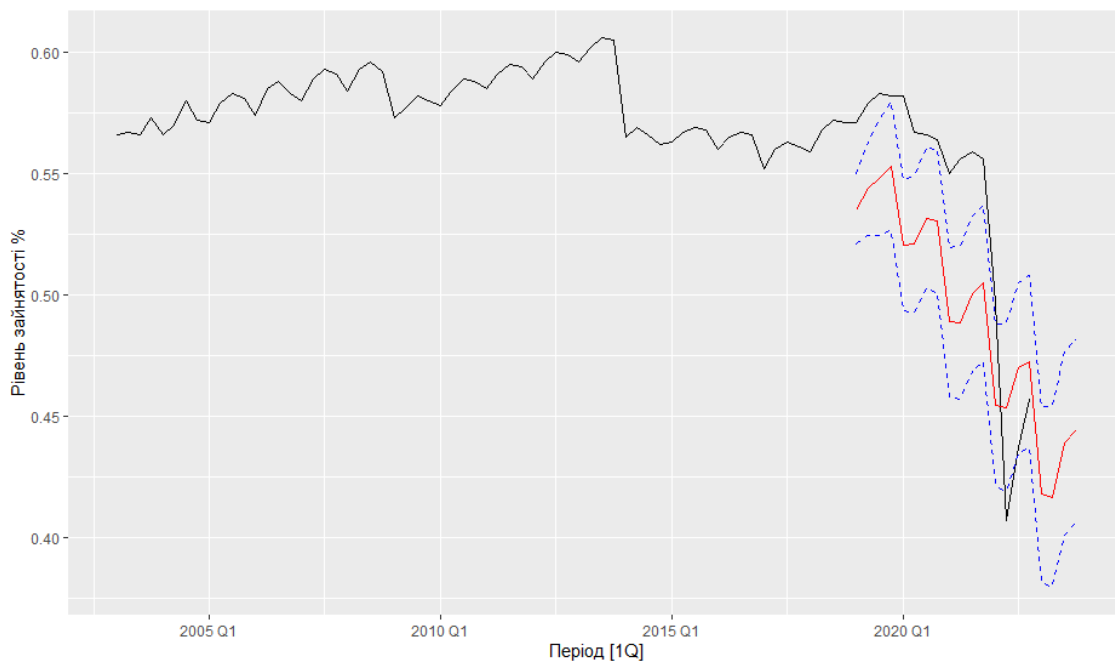


Рисунок 3.19 – Графік фактичних та прогнозованих значень рівня зайнятості

Джерело: розроблено автором за даними [26][33][35][52]

3.3 Аналіз результатів моделювання динаміки плинності працівників на базі методів машинного навчання

Важливим завданням бізнесу є контроль плинності персоналу на підприємстві. Персонал це головна складова будь-якого бізнесу. Для поглибленого дослідження плинності кадрів на підприємства можна використовувати методи машинного навчання. Однією з груп методів, що підходять для задачі моделювання плинності кадрів на підприємстві є методи класифікації. Завданням класифікації є прогноз дискретної цільової змінної. В нашому випадку цільовою змінною моделі буде виступати змінна, що відповідає на питання чи звільниться працівник з підприємства та має дві відповіді так чи ні. Побудова моделі такого типу допоможе підприємству прогнозувати плинність кадрів в залежності від визначених факторів.

Початковим етапом проведення будь-якого моделювання є збирання та обробка даних. Для побудови даної моделі отримані дані від бази даних підприємства на якому проводилась переддипломна практика. Вхідні дані моделі мають наступну структуру:

Таблиця 3.3. Структура вхідних даних

Коротка назва	Інформація про змінну	Формат даних
Attrition	Чи звільниться працівник?	Факторний
Age	Вік працівника	Числовий
Job_pos	Посада, що займає працівник	Факторний
Job_title	Ступінь посади, яку займає працівник	Факторний

Продовження таблиці 3.3

Salary_satisfaction	Задоволеність заробітною платою	Факторний
Fin_status	Оцінка фінансового стану на думку працівника	Факторний
gender	Стать працівника	Факторний
education	Рівень освіти працівника	Факторний
Work_experience	Досвід роботи	Числовий

Джерело: розроблено автором

Після отримання даних обов'язково потрібно перевірити їх на наявність дублікатів та пустих значень. В даному випадку в даних не було виявлено дублікатів або пустих значень. Також після завантаження даних в систему програмування RStudio , потрібно перевірити типи даних. Для наших факторних змінних автоматично було присвоєно текстовий тип даних. Нам потрібно спочатку перетворити змінні текстового типу на тип даних фактор. Факторні змінні це категоріальні змінні, що складаються зі значень міток. До таких змінних відносяться посада працівника, ступінь посади, рівень освіти, задоволеність заробітною платою, оцінка фінансового стану та стать. Ці змінні нам потрібно перетворити за допомогою методу one hot encoding. Сутність цього методу полягає в тому, що ми перетворюємо кожне категоріальне значення в новий стовпець і призначаємо стовпцям двійкові значення 1 або 0. Це потрібно для того, щоб наша модель змогла прочитати факторні змінні. Ще одним кроком перед початком моделювання є поділ даних на тестову та навчальну вибірку. В даному випадку ми поділили дані в пропорції 70% навчальної вибірки та 30% тестової вибірки.

Отже, для моделювання задачі плинності кадрів на підприємстві пропоную використати 3 методи а саме: метод наївного Байеса, метод к-найближчих сусідів та метод випадкового лісу. Почнемо з побудови моделі методом наївного Байеса.

Цей метод базується на розрахунку умовних ймовірностей та присвоєнні відповідно класу в залежності від даної ймовірності. Пропоную побудувати декілька моделей в кожную з яких будуть поступово додаватися фактори та оцінити кожную з моделей за показниками точності і обрати найкращу модель серед побудованих. Для оцінки побудованих моделей будемо використовувати три ключові показники оцінки точності, а саме: точність моделі, площа під кривою AUC та показник F1. Ці показники використовуються для оцінки точності класифікатора та відхилень прогнозу від факту. В результаті було побудовано 4 моделі методом наївного Байєса. Перша модель складається з цільової змінної та факторів задоволеність заробітною платою, оцінка фінансового стану. Модель має високу якість класифікатора за показником ROC AUC, також точність моделі є доволі високою за показниками точності та F1 score. В другу модель додатково до першої модифікації були додані фактори досвід роботи та рівень освіти. Як бачимо, дана модель має трохи вищу точність, проте класифікатор є трохи гіршим за ROC AUC та F1 score. В наступну модель були додані додатково змінні вік та посада, що займає працівник. В даному випадку модель має ще нижчу точність ніж перші дві. Як бачимо зі зростанням кількості факторів модель має нижчу точність. Остання модель включає в себе усі фактори, які були визначені на етапі збору даних. В даному випадку модель має найнижчу точність прогнозу за показниками accuracy та F1 score. Я би рекомендував застосувати на практиці модель під номером два так як вона є найбільш оптимальною. Тому давайте додатково розглянемо прогноз даної моделі.

Таблиця 3.4. Результати та модифікації побудованих моделей методом Байєса

№	Модель	Точність	ROC AUC	F1 score
1	Attrition ~ salary_satisfaction + fin_status	0,724	0,768	0,615

Продовження таблиці 3.4

2	Attrition ~ salary_satisfaction + fin_status + work_experience + education	0,736	0,707	0,538
3	Attrition ~ salary_satisfaction + fin_status + work_experience + education + age + job_title	0,706	0,627	0,478
4	Attrition ~ salary_satisfaction + fin_status + work_experience + education + age + gender + job_title + job_pos	0,350	0,650	0,448

Джерело: розроблено автором

Додатково для аналізу будують матрицю відповідностей, яка показує кількість спостережень по класам в розрізі прогноз та факт. Як бачимо з рисунку 3.20 дана модель віднесла до класу 1 вірно 82 спостереження зі 116 спостережень 1 класу. Майже половина спостережень класу 1 була віднесена до неправильного класу. Це говорить про те, що дана модель не зовсім точна. В нашому випадку нам дуже важливо, щоб модель правильно прогнозувала саме клас 1, тобто працівників, що звільняються. Додатково можна побудувати криву AUC, яка показує співвідношення частоти правильного позитивного результату до частоти помилково позитивного результату. В даному випадку, правильний позитивний результат достатньо сильно перевищує помилково позитивний (рис 3.21).

	Truth	
Prediction	0	1
0	36	34
1	11	82

Рисунок 3.20 – Матриця відповідностей моделі №2

Джерело: розроблено автором за даними бази практики

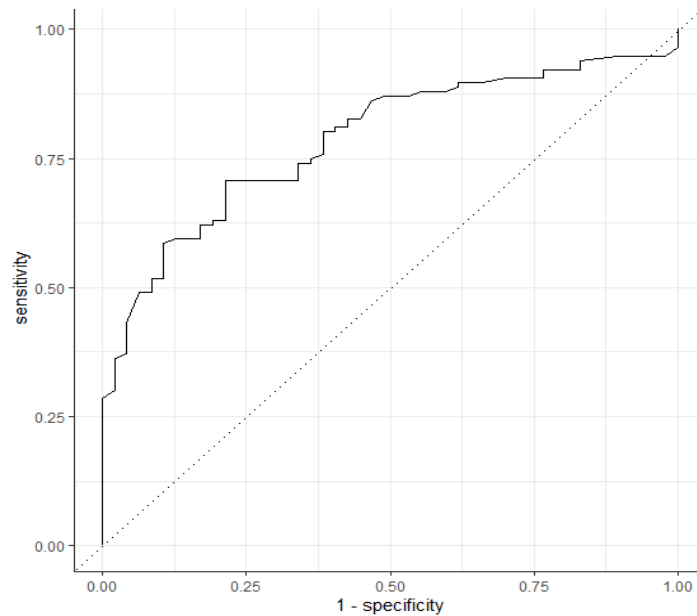


Рисунок 3.21 – Крива AUC моделі №2

Джерело: розроблено автором за даними бази практики

Результат методу найвісного Баєсу є достатньо непоганий, проте дослідимо інші методи моделювання та порівняємо потім результат побудованих моделей. Одним з методів класифікації, що підходить для нашої задачі є метод к-найближчих сусідів. Цей метод використовує значення класу найближчих сусідів для вибору класу спостереження. Для цього обирається кількість найближчих сусідів, які будуть враховуватися при побудові моделі, а потім вираховується евклідова відстань для визначення найближчих сусідів. Тому, спочатку стандартизуємо наші числові змінні методом MinMax. Отже, побудуємо 4 моделі з кількістю сусідів 5, 10, 15 та 20 та оберемо найточнішу модель. У всіх моделях будемо використовувати усіх незалежні фактори, тобто це: задоволеність заробітною платою, оцінка фінансового стану, рівень освіти, вік, стать, посада, ступінь посади, досвід роботи. За показниками точність, ROC AUC, F1 score модель з кількістю найближчих сусідів 20 має найвищу точність та якість класифікатору. В цілому моделі з меншою кількістю найближчих сусідів теж мають високу точність та якість класифікатору, проте модель №4 все ж має незначну перевагу. Тому дослідимо цю модель

додатково на точність та якість прогнозу та порівняємо з найкращою моделлю за методом Байеса.

Таблиця 3.5. Результати побудованих моделей методом k-найближчих сусідів

Модель	Кількість найближчих сусідів	Точність	ROC AUC	F1 score
Модель №1	5	0,728	0,740	0,511
Модель №2	10	0,765	0,796	0,548
Модель №3	15	0,735	0,721	0,442
Модель №4	20	0,778	0,825	0,526

Джерело: розроблено автором

Як бачимо з рисунку 3.22 модель побудована методом k-найближчих сусідів з 20 сусідами вірно віднесла 102 спостереження до 1 класу та помилково віднесла 13 спостережень до класу 0. Тобто модель має доволі високу точність прогнозування класу 1. Це якраз результат який ми хочемо отримати. Як бачимо клас 0 модель прогнозує гірше, більша половина спостережень помилково віднесені до 1 класу. На кривій AUC правильний позитивний результат достатньо сильно перевищує помилково позитивний (рис 3.23). Порівнюючи дану модель з найточнішою моделлю побудованою за методом Байеса, можна дійти висновку, що модель за методом k-найближчих сусідів має вищу точність прогнозу та якість класифікатора. Далі пропоную розглянути метод випадковий ліс та порівняти точність з іншими методами.

	Truth	
Prediction	0	1
0	22	13
1	25	102

Рисунок 3.22 – Матриця відповідностей моделі №4

Джерело: розроблено автором за даними бази практики

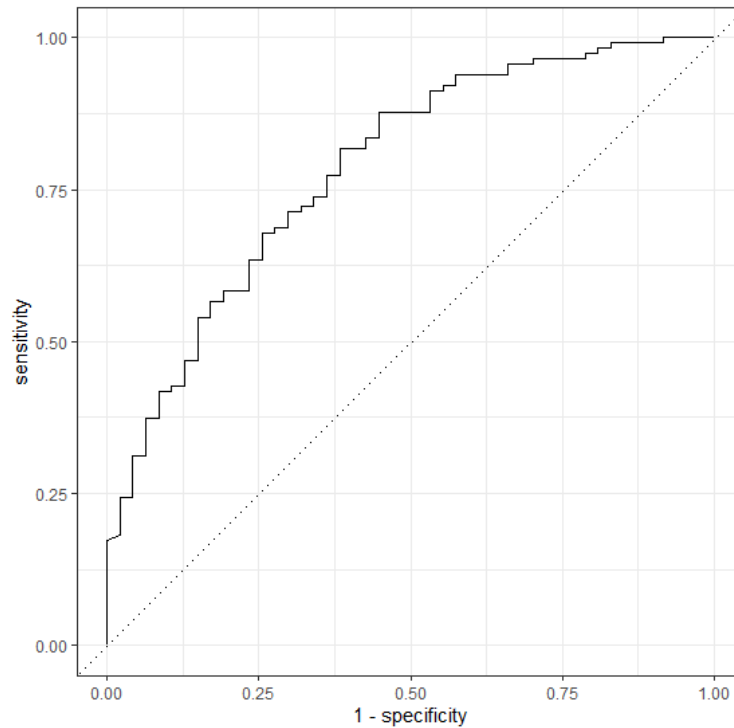


Рисунок 3.23 – Крива AUC моделі №4

Джерело: розроблено автором за даними бази практики

Випадковий ліс є одним з методів побудови дерев рішень. Алгоритм будує цілий набір дерев рішень та має три основні гіперпараметри такі як: кількість дерев, кількість предикторів для вибірки при кожному розділенні та кількість спостережень необхідних для продовження розділення вузлів. Ми будемо використовувати 1000 дерев для побудови моделі. Параметри кількість предикторів та кількість спостережень ми будемо підбирати за допомогою крос-валідації, тобто будемо будувати велику кількість моделей, та перевіряти точність в залежності від обраних параметрів. Після побудови моделей, був побудований графік, який показує зміну показника ROC AUC в залежності від кількості спостережень,

необхідні для продовження розділення вузлів (`min_n`) та кількість предикторів для вибірки при кожному розділенні дерева (`mtry`) (рис 3.24). Як бачимо, в основному при кількості спостережень в діапазоні від 20 до 40 модель має найвищий ROC AUC, а оптимальна кількість предикторів знаходиться в діапазоні від 10 до 20. Тому задамо діапазони гіперпараметрів згідно висновків вище, та побудуємо моделі з такими параметрами та визначимо найкращу модель.

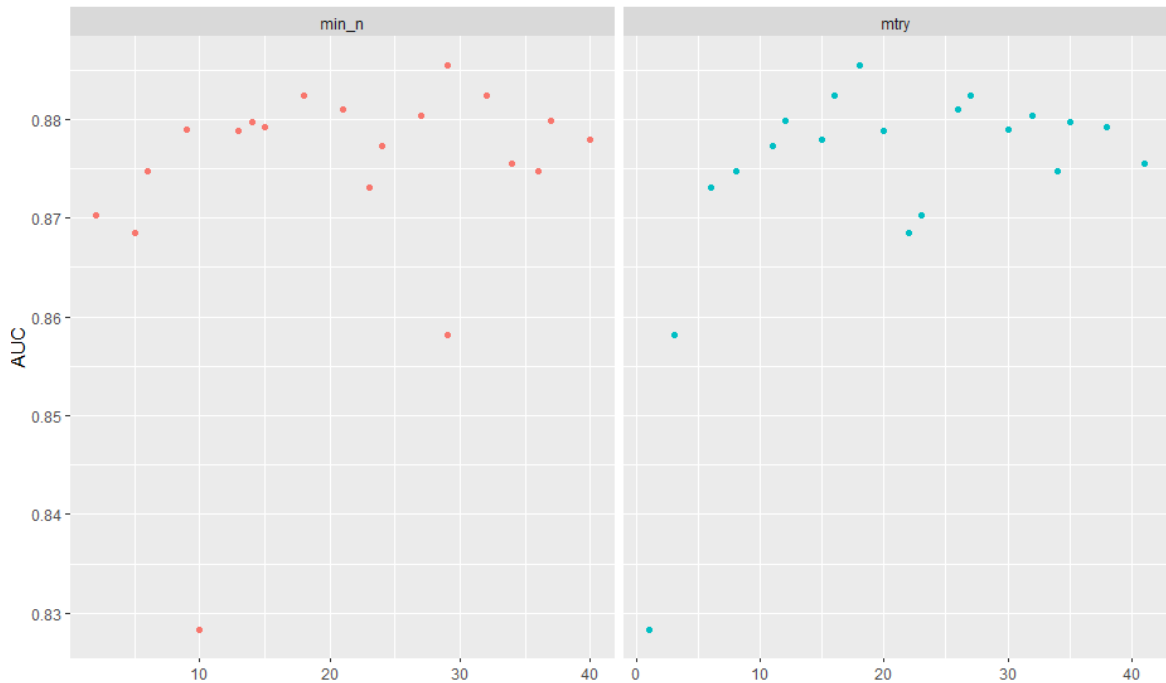


Рисунок 3.24 – ROC AUC в розрізі гіперпараметрів моделі

Джерело: розроблено автором за даними бази практики

В результаті побудови моделей методом випадкового лісу модель з гіперпараметрами 1000 дерев, кількістю предикторів для вибірки при розділенні дерева 20 та кількістю спостережень для продовження розділення вузлів 30 є оптимальною. Як бачимо з рисунку 3.25 модель побудована методом випадкового лісу вірно віднесла 108 спостереження до 1 класу та помилково віднесла 7 спостережень до класу 0. Тобто модель має високу точність прогнозування класу 1. Також дана модель доволі точно прогнозує клас 0, більша половина спостережень віднесені правильно. Дана модель має точність на рівні 0,864 та показник ROC AUC 0,869 і F1 score на рівні 0,744. На кривій AUC правильний позитивний результат

достатньо сильно переважає помилково позитивний (рис 3.26). Можна дійти висновку, що даний метод побудував найкращу та найточнішу модель. Дану модель можна використовувати для подальшого розуміння стану плинності кадрів на підприємстві та планування процесів, що стосуються персоналу.

		Truth	
Prediction		0	1
	0	32	7
	1	15	108

Рисунок 3.25 – Матриця відповідностей моделі випадкового лісу

Джерело: розроблено автором за даними бази практики

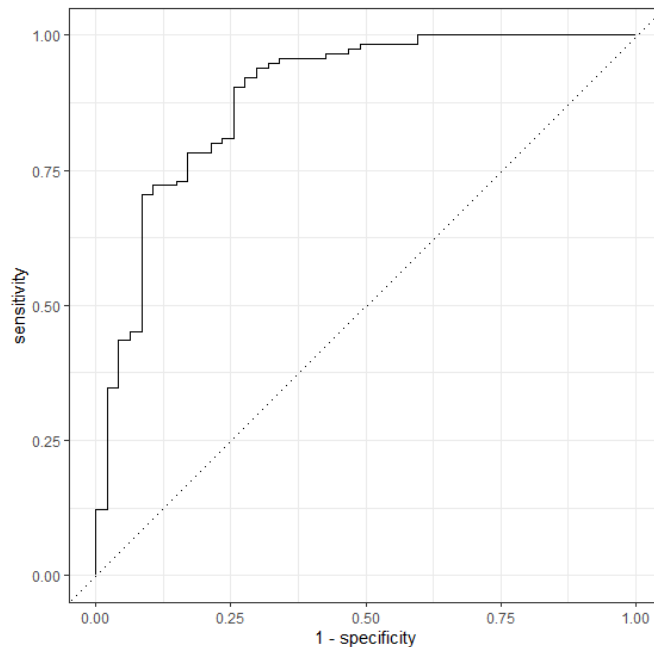


Рисунок 3.26 – Крива AUC моделі випадкового лісу

Джерело: розроблено автором за даними бази практики

ВИСНОВКИ

У кваліфікаційній роботі досліджено сутність ринку праці та особливості його функціонування. Ринок праці є складним економічним процесом та дуже важливим для функціонування держави. Також можна визначити, що ринок праці є системою правових, соціально-трудова, економічних та організаційних відносин, що виникають між особами, які шукають роботу та роботодавцями. Зайнятість та безробіття є важливими компонентами ринку праці. Зайнятість характеризує частину економічно активного населення, яка має місце роботи, а безробіття відповідно характеризує частину населення, що не мають постійного місця праці.

Проаналізовано проблему зі зайнятістю на підприємстві, як одну з актуальних проблем, яка виникає внаслідок негативних тенденцій на ринку праці та в економіці країни. Плинність кадрів характеризує кількість персоналу, які покинули компанію з різних зовнішніх та внутрішніх причин. Дана проблема характерна будь-якому підприємству, проте високий рівень плинності кадрів загрожує функціонуванню бізнесу. На жаль після початку повномасштабного вторгнення у 2022р. дана проблема загострилась. У 76% медіа агенцій через початок війни, внаслідок чого погіршився стан економіки, а також відбулася міграція населення, відбулося скорочення персоналу підприємства. Такі зовнішні чинники вплинули на зростання рівня плинності кадрів. Одним з шляхів вирішення високого рівня плинності кадрів є побудова математичних моделей для прогнозування звільнення працівників, що допоможе прогнозувати рівень плинності кадрів.

В кваліфікаційній роботі проаналізовано стан ринку праці України та зарубіжних країн. Для оцінювання стану ринку праці важливо додатково дослідити демографічну ситуацію в країні. В Україні за останні 20 років спостерігається зниження рівня постійного населення. У 2022 р. після початку повномасштабного

вторгнення за даними ООН рівень постійного населення знизився на 6,2 млн. осіб. Однією з причин зниження рівня населення є рівень смертності, який значно перевищує рівень народжуваності. В Україні станом на 2021 р. був найнижчим показником серед таких країн як: США, Велика Британія, Франція, Німеччина, Італія, Польща, Угорщина, Румунія, Словаччина. Такий низький рівень народжуваності негативно впливає на рівень населення держави. Також це впливає на зростання частки населення пенсійного віку (65 років і старше). За останні 20 років частка населення віком 65 років і старше значно зросла та станом на 2020 р. складає 17,1%. Такий високий показник населення пенсійного віку створює додаткове навантаження на економіку країни. Демографічна ситуація дуже сильно впливає на стан ринку праці. На даний момент демографічну ситуацію в Україні можна охарактеризувати як негативна. Рівень населення знижується з кожним роком, а смертність значно перевищує народжуваність і негативно впливає на рівень населення.

За останні 20 років спостерігається зниження рівня зайнятості та зростання рівня безробіття в Україні. Станом на 2022 р. Україна має найбільший рівень безробіття та один з найнижчих рівнів зайнятості серед країн ЄС, США, та європейських країн сусідів. Найбільшою причиною таких негативних тенденцій став початок повномасштабного вторгнення у 2022р. Велика кількість роботодавців вимушені були скорочувати персонал, а багато підприємств було просто зруйновано. Це вплинуло на значне зниження попиту та пропозиції праці, які можна відслідкувати за індексом приросту нових вакансій та резюме на сайтах пошуку роботи. Проте починаючи з червня 2022р. попит та пропозиція почали поступово відновлюватися, проте попит значно перевищував пропозицію, що створювало дисбаланс на ринку праці. Ринок праці на кінець 2022 р. можна охарактеризувати станом занепаду. Безробіття досягнуло високого рівня та створює перевищення попиту праці над пропозицією. Проте поступове відновлення ринку праці вже

триває. Тому будемо очікувати на покращення ситуації на ринку праці в найближчому майбутньому.

В процесі дослідження ринку праці України побудовано комплекс математичних моделей прогнозування рівня зайнятості в Україні та плинності кадрів на підприємстві. Для прогнозування рівня зайнятості побудовані математичні моделі прогнозування методами класичної та векторної авторегресії, експоненціального згладжування, методу Prophet. В результаті модель побудована методом векторної авторегресії має найкращу точність. Ця модель визначає рівень зайнятості в залежності від рівня безробіття, економічно активного населення, середньої заробітної плати, індексу споживчих цін, прямих інвестицій та номінального ВВП. Модель має високий коефіцієнт детермінації та доволі низьке середньоквадратичне відхилення, що доводить високу точність даної моделі. В результаті дану модель можна використовувати для прогнозування рівня зайнятості на найближче майбутнє. Це допоможе планувати державні заходи та стратегії розвитку ринку праці, та розуміти прогнозований стан ринку праці. Модель також допоможе для підприємства оцінювати стан ринку праці на найближче майбутнє.

Для вирішення однієї з важливих проблем на підприємстві, плинності кадрів, яка тісно пов'язана з ринком праці, побудовані моделі прогнозування звільнення працівника на основі даних опитування персоналу підприємства за факторами: задоволення працею, фінансовим станом, рівня освіти, віку, посади, досвіду роботи, рівня посади, статі. Використано такі методи класифікації: метод наївного Байеса, метод k-найближчих сусідів, метод випадкового лісу. В результаті модель побудована методом випадкового лісу виявилася найкращою. Модель має високу точність, показник ROC AUC та F1_score, що говорить про високу якість класифікатора та точність прогнозу обох класів працівників. В результаті підприємство на основі даних опитування може оцінити рівень плинності кадрів, що допоможе краще планувати стратегію найму персоналу на майбутні періоди.

В результаті проведеного дослідження можна сказати, що ринок праці України в даний момент перебуває в критичному стані та характеризується високим попитом на працю але не високою пропозицією, високим рівнем безробіття та низьким рівнем зайнятості. На такий стан в основному вплинув початок повномасштабного вторгнення, внаслідок чого виникли економічні проблеми та міграція населення, що ще більше загостило демографічну ситуацію. Були побудовані моделі прогнозування рівня зайнятості в залежності від економічних чинників, що допоможе покращити процес планування бізнесу. Також для вирішення важливої проблеми, що виникла внаслідок негативного стану ринку праці, була побудована модель, що прогнозує звільнення працівника. За допомогою такого інструменту підприємство зможе краще планувати заходи з найму персоналу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Кваша О. С. Ринок праці в Україні: аналіз сучасного стану та перспективи розвитку. *Глобальні та національні проблеми економіки*. 2018. №22. С. 1074–1079.
2. Колот А. М., Грішнова О.А., Герасименко О.О. Економіка праці та соціально-трудові відносини. Київ: КНЕУ, 2009. 711 с.
3. Калина А. В. Ринок праці (національний та міжнародний аспекти). Київ: Видавничий дім "Персонал", 2010. 213 с.
4. Закон України «Про зайнятість населення»: від 01.03.1991. URL: <https://ips.ligazakon.net/document/T080300?an=8> (дата звернення: 10.10.2023).
5. Чернявська О. В. Ринок праці. Київ: Центр учбової літератури, 2013. 522с.
6. Державна служба статистики України. Методологічні положення щодо класифікації та аналізу економічної активності населення. URL: https://ukrstat.gov.ua/metod_polog/metod_doc/2011/12/metod.htm (дата звернення: 10.10.2023).
7. Оліскевич М. О. Економетричне моделювання динамічних процесів розвитку ринку праці України. Львів: ЛНУ, 2015. 400 с.
8. Peopleforce. Що таке коефіцієнт плинності кадрів та як його розрахувати? URL: <https://peopleforce.io/uk/blog/yak-rozrahuvaty-koeficient-plynnosti-kadriv> (Дата звернення: 11.10.2023).
9. Шишкін В. О., Лозоба Н. В. Проблема плинності кадрів на підприємстві та шляхи її вирішення. *Інвестиції практика та досвід*. 2009. №12. С.47-48.
10. Процик І.С., Кара Н. І. Плинність кадрів на підприємстві та шляхи її зменшення. *«Молодий вчений»*. 2020. №3. С. 246-252.

- 11.Піхняк Т. Аналіз показників плинності персоналу виробничого підприємства. *Modeling the development of the economic systems*. 2022. №2. С. 154-160.
12. Всеукраїнська рекламна коаліція. Об'єм рекламно-комунікаційного ринку України 2022. URL: <https://vrk.org.ua/news-events/2022/ad-volume-2022.html> (дата звернення: 11.10.2023).
- 13.Всеукраїнська рекламна коаліція. Дослідження «Пульс рекламної індустрії». URL: <https://vrk.org.ua/images/projects/Advertising%20industry%20pulse.pdf> (дата звернення: 11.10.2023).
- 14.Економічна правда. Люди звільняються? Як зупинити плинність кадрів. URL: <https://www.epravda.com.ua/columns/2023/05/23/700401/> (дата звернення: 11.10.2023).
- 15.Єріна А. М. Статистичне моделювання та прогнозування. Київ: КНЕУ, 2001. 170 с.
- 16.Яровий А. Т., Страхов Є.М. Багатовимірний статистичний аналіз. Одеса: ОНУ, 2020. 130 с.
- 17.Hyndman R. J. Forecasting: Principles and PracticeForecasting: Principles and Practice (3rd ed). 2021. URL: <https://otexts.com/fpp3/> (дата звернення: 10.10.2023).
- 18.Юнькова О. О., Кучер П. В. Моделювання та прогнозування товарних ринків на основі методу декомпозиції. *Моделювання та інформаційні системи в економіці*. 2021. №101. С. 182-191.
- 19.Analytics Vidhya. Statistical Tests to Check Stationarity in Time Series. URL: <https://www.analyticsvidhya.com/blog/2021/06/statistical-tests-to-check-stationarity-in-time-series-part-1/>. (дата звернення: 10.10.2023).

20. Medium. What is Stationarity in Time Series? How it can be detected? URL: <https://medium.com/codex/what-is-stationarity-in-time-series-how-it-can-be-detected-7e5dfa7b5f6b> (дата звернення: 10.10.2023).
21. Medium. Tidy Forecasting in R: ARIMA Models. URL: <https://joonim.medium.com/tidy-forecasting-in-r-arima-models-864dc13773c1> (дата звернення: 10.10.2023).
22. Influxdata Exponential Smoothing: A Beginner's Guide to Getting Started. URL: <https://www.influxdata.com/blog/exponential-smoothing-beginners-guide/#:~:text=Exponential%20smoothing%20is%20a%20time,changing%20trends%20in%20the%20data>. (дата звернення: 10.11.2023).
23. Medium. MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared — Which Metric is Better? URL: <https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e> (дата звернення: 10.11.2023).
24. Analytics Yogi. AIC & BIC for Selecting Regression Models: Formula, Examples. URL: <https://vitalflux.com/aic-vs-bic-for-regression-models-formula-examples/> (дата звернення: 10.11.2023).
25. Мінфін. Населення України. URL: <https://index.minfin.com.ua/ua/reference/people/> (дата звернення: 12.11.2023).
26. Державна служба статистики України. Статистична інформація. URL: <https://ukrstat.gov.ua/> (дата звернення: 20.11.2023).
27. Слово і Діло аналітичний портал. Скільки дітей щороку народжувалося в Україні за період незалежності URL : <https://www.slovoidilo.ua/2023/08/04/infografika/suspilstvo/skilky-ditej-shhoroku-narodzhuvalosya-ukrayini-period-nezalezhnosti> (дата звернення: 20.11.2023).

- 28.Опендатабот. Аналітика. URL: <https://opendatabot.ua/analytics> (дата звернення: 22.11.2023).
- 29.Укрінформ. Чисельність населення України варіюється від 28 до 34 мільйонів – демограф. URL : <https://www.ukrinform.ua/rubric-society/3692466-ciselnist-naselenna-ukraini-variuetsa-vid-28-do-34-miljoniv-demografi.html> (дата звернення: 22.11.2023).
- 30.Life-House. Що таке старіння нації? Світова статистика. URL : <https://life-house.ua/blog/ctarinnya-nacziyi-v-ukrayini/> (дата звернення: 22.11.2023).
- 31.Світ КПІ. Населення України та економіка: якщо в Україні будуть створюватися умови для відновлення економіки та виникати нові робочі місця, то українців стане більше. URL: <https://svit.kpi.ua/2023/07/12/%D0%BD%D0%B0%D1%81%D0%B5%D0%BB%D0%B5%D0%BD%D0%BD%D1%8F-%D1%83%D0%BA%D1%80%D0%B0%D1%97%D0%BD%D0%B8-%D1%82%D0%B0-%D0%B5%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D1%96%D0%BA%D0%B0-%D1%8F%D0%BA%D1%89%D0%BE-%D0%B2/> (дата звернення: 22.11.2023).
- 32.Андрейцева І. А. Ринок праці України: сучасні тенденції. *Економіка та суспільство*. 2021. №28.
- 33.Центр економічної стратегії. Трекер економіки України під час війни. URL: <https://ces.org.ua/tracker-economy-during-the-war/#anchor-1> (дата звернення: 22.11.2023).
- 34.Михайлова Ю. Г., Самотоєнкова О.В. Аналіз динаміки середньої заробітної плати в Україні. *Статистика – інструмент соціально-економічних досліджень : збірник наукових студентських праць*. 2017. №3. С. 1–8.

35. Пенсійний фонд України. Дані про розмір середньої заробітної плати для обчислення пенсії URL: <https://www.pfu.gov.ua/statystyka/pokazniki-serednoyi-zarobitnoyi-plat/> (дата звернення: 25.11.2023).
36. World Bank. Open Data. URL : <https://data.worldbank.org/> (дата звернення: 25.11.2023).
37. Zippia. Job satisfaction by country [2023]. URL: <https://www.zippia.com/advice/job-satisfaction-statistics-by-country/> (дата звернення: 25.11.2023).
38. ЕВА. Українці задоволені роботою, але не зарплатою URL: <https://eba.com.ua/ukrayintsi-zadovoleni-robotoyu-ale-ne-zarplatoyu/> (дата звернення: 25.11.2023).
39. OECD. Open data. URL: <https://data.oecd.org/> (дата звернення: 25.11.2023).
40. Махортов Ю. О., Секрет О.В. Продуктивність праці в українському та світовому вимірах. *Економічний простір*. 2020. №163. С. 7–13.
41. ILOSTAT. Statistics on labour productivity. URL: <https://ilostat ilo.org/topics/labour-productivity/#:~:text=Labour%20productivity%20represents%20the%20total,a%20given%20time%20reference%20period>. (дата звернення: 25.11.2023).
42. Wikipedia. Global Competitiveness Report. URL: https://en.wikipedia.org/wiki/Global_Competitiveness_Report#Description (дата звернення: 25.11.2023).
43. Weforum. The Global Competitiveness Report 2019.; URL: https://www3.weforum.org/docs/WEF_TheGlobalCompetitivenessReport2019.pdf (дата звернення: 25.11.2023).
44. WIPO. Global Innovation Index 2021. URL: https://www.wipo.int/global_innovation_index/en/2021/index.html (дата звернення: 25.11.2023).

45. Edureka! Introduction to Classification Algorithms. URL: <https://www.edureka.co/blog/classification-algorithms/#classificationalgorithmsvsclusteringalgorithms> (дата звернення: 10.11.2023).
46. KDnuggets. Naïve Bayes Algorithm: Everything You Need to Know. URL: <https://www.kdnuggets.com/2020/06/naive-bayes-algorithm-everything.html> (дата звернення: 10.11.2023).
47. IBM. What is the k-nearest neighbors algorithm? URL: <https://www.ibm.com/topics/knn#Compute+KNN%3A+defining+k%09%09%09%09%09%09%09%09%09%09%09> (дата звернення: 10.11.2023).
48. IBM. What is a Decision Tree? URL: <https://www.ibm.com/topics/decision-trees#How+to+choose+the+best+attribute+at+each+node> (дата звернення: 10.11.2023).
49. DataCamp. Classification in Machine Learning: An Introduction. URL: <https://www.datacamp.com/blog/classification-machine-learning> (дата звернення: 12.11.2023).
50. Analytics Vidhya. Metrics to Evaluate your Classification Model to take the right decisions. URL: <https://www.analyticsvidhya.com/blog/2021/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/#h-accuracy> (дата звернення: 12.11.2023).
51. Medium. Metrics to Evaluate your Machine Learning Algorithm. URL: <https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234> (дата звернення: 12.11.2023).
52. Національний банк України. Макроекономічні показники URL: <https://bank.gov.ua/ua/statistic/macro-indicators> (дата звернення: 25.11.2023)

ДОДАТКИ

Додаток А

```

library(fpp3)
library(tidyverse)
library(feasts)
library(GGally)
library(tseries)
library(prophet)
library(tidymodels)
library(infer)
library(car)
library(readxl)
data <- read_excel("employment_Ukraine_quartarly_data.xlsx")
na.omit(data)
data$Період <- yearquarter(data$Період)
data <- data %>%
  as_tsibble(index = Період)
data
# Time-series analysis
data %>%
  autoplot(`Рівень зайнятості %`) +
  labs(title = "Рівень зайнятості в розрізі періоду")
data %>%
  gg_season(`Рівень зайнятості %`, labels = "both") +
  labs(title = "Рівень зайнятості в розрізі періоду")
data %>%
  gg_subseries(`Рівень зайнятості %`) +

```

```

labs(title = "Рівень зайнятості в розрізі періоду")

data %>%
  model(STL(`Рівень зайнятості %` ~ trend() +
    season(),
    robust = TRUE)) %>%
  components() %>%
  autoplot()
# Autocorrelation analysis
data %>%
  ACF(`Рівень зайнятості %`, lag_max = 16) %>%
  autoplot()

data %>%
  gg_lag(`Рівень зайнятості %`, geom = "point", lag = 1:12)
data %>%
  features(`Рівень зайнятості %`, guerrero)
adf.test(data$`Рівень зайнятості %`)
data %>%
  features(`Рівень зайнятості %`, unitroot_kpss)
data %>%
  features(`Рівень зайнятості %`, unitroot_ndiffs)
data_difference <- data %>%
  dplyr::select(`Період`, `Рівень зайнятості %`) %>%
  mutate(diff1 = difference(`Рівень зайнятості %`, lag = 1))
data_difference %>%
  features(diff1, unitroot_kpss)
# differencing data

```

```

data %>%
  autoplot(difference(`Рівень зайнятості %`, lag = 1)) +
  labs(title = "Log 1 difference (Рівень зайнятості)")
data %>%
  autoplot(difference(`Рівень зайнятості %`, lag = 2)) +
  labs(title = "Log 2 difference (Рівень зайнятості)")
data %>%
  autoplot(difference(`Рівень зайнятості %`, lag = 4)) +
  labs(title = "Log 4 difference (Рівень зайнятості)")
train <- data_difference %>%
  filter(year(Період) <= 2018)
test <- data_difference %>%
  filter(year(Період) > 2018)
#ARIMA, SARIMA
data_difference %>%
  gg_tsdisplay(difference(`Рівень зайнятості %`), plot_type = 'partial')
fit <- data_difference %>%
  model(Arima = ARIMA(`Рівень зайнятості %`),
        auto_season = ARIMA(`Рівень зайнятості %`, stepwise = FALSE, approx
= FALSE))
fit
report(fit)
fit %>%
  dplyr::select(Arima) %>%
  augment() %>%
  features(.innov, ljung_box)
fit %>%
  dplyr::select(Arima) %>%

```

```

  gg_tsresiduals()
fitted <- fit %>%
  dplyr::select(Arima) %>%
  fitted()
ARIMA <- fit %>%
  dplyr::select(Arima)

fit %>%
  dplyr::select(Arima) %>%
  forecast(h = "2 years") %>%
  autoplot(data_difference, series="train data") + autolayer(fitted, series="fitted
value", col = "#D55E00")

#ARIMA, SARIMA, ETS, prophet
library(fable.prophet)
fit <- train %>%
  model(auto_season = ARIMA(`Рівень зайнятості %`, stepwise = FALSE,
approx = FALSE),
  ETS = ETS(`Рівень зайнятості %`),
  prophet = prophet(`Рівень зайнятості %`) ~ season(period = 4, type =
"multiplicative", order = 10))) %>%
  report(fit)

fc <- fit %>%
  forecast(h = "4 years")

fit %>%
  dplyr::select(prophet) %>%

```

```

forecast(h = "4 years") %>%
autoplot(data_difference)

fc %>%
  fabletools::accuracy(data_difference)
fit %>%
  dplyr::select(ETS) %>%
  augment() %>%
  features(.innov, ljung_box)
fit %>%
  dplyr::select(ETS) %>%
  gg_tsresiduals()
fit_ARIMA <- fit %>%
  dplyr::select(auto)
fit %>%
  dplyr::select(prophet) %>%
  augment() %>%
  features(.innov, ljung_box)
fit %>%
  dplyr::select(prophet) %>%
  gg_tsresiduals()

fit %>%
  dplyr::select(prophet) %>%
  forecast(h = "4 years") %>%
  autoplot(data_difference)

adf.test(data$`Середня заробітна плата грн.`)
```

```

adf.test(data$`Економічно активне населення тис. ос.`)
adf.test(data$`Індекс споживчих цін`)
adf.test(data$`Номінальний ВВП млн.грн`)
adf.test(data$`Прямі інвестиції млн. дол.`)
data_difference
data %>%
  features(`Рівень безробіття %`, unitroot_kpss)
data %>%
  features(`Рівень безробіття %`, unitroot_ndiffs)
data %>%
  features(`Середня заробітна плата грн.`, unitroot_kpss)
data %>%
  features(`Середня заробітна плата грн.`, unitroot_ndiffs)

# >=0,05 ряд стаціонарний
data %>%
  features(`Економічно активне населення тис. ос.`, unitroot_kpss)
data %>%
  features(`Економічно активне населення тис. ос.`, unitroot_ndiffs)

data %>%
  features(`Індекс споживчих цін`, unitroot_kpss)

data %>%
  features(`Номінальний ВВП млн.грн`, unitroot_kpss)

data %>%
  features(`Номінальний ВВП млн.грн`, unitroot_ndiffs)

```

```

data %>%
  features(`Прямі інвестиції млн. дол.`, unitroot_kpss)

library(vars)
data %>%
  ggpairs(columns = 2:8)

train <- data %>%
  filter(year(Період) <= 2018)
test <- data %>%
  filter(year(Період) > 2018)

VAR_data <- train[,c("Рівень зайнятості %", "Рівень безробіття %")]
lag <-
  VARselect      (VAR_data, lag.max = 10, type = "const")
lag

VAR_est <- VAR(VAR_data, p = 2, type = "const", season = NULL,
               exog = NULL)
summary(VAR_est$varresult$Рівень.зайнятості.)

forecasts <- predict(VAR_est, n.ahead = 16)
forecasts

test <- test %>%
  mutate(fcst = forecasts$fcst$Рівень.зайнятості..[1:16],
         lower = forecasts$fcst$Рівень.зайнятості..[17:32],

```

```

upper = forecasts$fcst$Рівень.зайнятості..[33:48])
test

MSE <- mean((test$`Рівень зайнятості %` - test$fcst)^2)
RMSE <- sqrt(MSE)
MSE
RMSE

serial <- serial.test(VAR_est, lags.pt = 10, type = "PT.asymptotic")
serial

plot(serial, names = "Рівень.зайнятості..")

grangertest(`Рівень зайнятості %` ~ `Рівень безробіття %`, order = 1, data =
train)
grangertest(`Рівень безробіття %` ~ `Рівень зайнятості %`, order = 1, data =
train)

autoplot(data, .vars = `Рівень зайнятості %`) +
  autolayer(test, .vars = `fcst`, colour = "red") +
  autolayer(test, .vars = `lower`, colour = "blue", linetype = "dashed") +
  autolayer(test, .vars = `upper`, colour = "blue", linetype = "dashed") +
  scale_color_manual(name='iivi',
    breaks=c('Прогноз', 'Нижня межа', 'Верхня межа'),
    values=c('Прогноз' ='red', 'Нижня межа' ='blue', 'Верхня межа
'='blue'))

```

```
VAR_data <- train[,c("Рівень зайнятості %", "Рівень безробіття
%", "Економічно активне населення тис. ос.")]
```

```
lag <-
```

```
VARselect (VAR_data, lag.max = 10, type = "const")
```

```
lag
```

```
VAR_est <- VAR(VAR_data, p = 5, type = "const", season = NULL,
               exog = NULL)
```

```
summary(VAR_est$varresult$Рівень.зайнятості.)
```

```
forecasts <- predict(VAR_est, n.ahead = 16)
```

```
forecasts
```

```
test <- test %>%
```

```
  mutate(fcst = forecasts$fcst$Рівень.зайнятості..[1:16],
```

```
         lower = forecasts$fcst$Рівень.зайнятості..[17:32],
```

```
         upper = forecasts$fcst$Рівень.зайнятості..[33:48])
```

```
test
```

```
MSE <- mean((test$`Рівень зайнятості %` - test$fcst)^2)
```

```
RMSE <- sqrt(MSE)
```

```
MSE
```

```
RMSE
```

```
serial <- serial.test(VAR_est, lags.pt = 12, type = "PT.asymptotic")
```

```
serial
```

```
plot(serial, names = "Рівень.зайнятості..")
```

```
grangertest(`Рівень зайнятості %` ~ `Економічно активне населення тис. ос.`,  
order = 5, data = train)
```

```
grangertest(`Економічно активне населення тис. ос.` ~ `Рівень зайнятості %`  
, order = 5, data = train)
```

```
autoplot(data, .vars = `Рівень зайнятості %`) +  
  autolayer(test, .vars = `fcst`, colour = "red") +  
  autolayer(test, .vars = `lower`, colour = "blue", linetype = "dashed") +  
  autolayer(test, .vars = `upper`, colour = "blue", linetype = "dashed") +  
  scale_color_manual(name='iivi',  
                      breaks=c('Прогноз', 'Нижня межа', 'Верхня межа'),  
                      values=c('Прогноз' ='red', 'Нижня межа' ='blue', 'Верхня межа'  
                                ='blue'))
```

```
VAR_data <- train[,c("Рівень зайнятості %", "Рівень безробіття  
%", "Економічно активне населення тис. ос.", "Середня заробітна плата грн.")]
```

```
lag <-
```

```
VARselect (VAR_data, lag.max = 10, type = "const")
```

```
lag
```

```
VAR_est <- VAR(VAR_data, p = 5, type = "const", season = NULL,  
               exog = NULL)
```

```
summary(VAR_est$varresult$Рівень.зайнятості.)
```

```
forecasts <- predict(VAR_est, n.ahead = 16)
```

```
forecasts
```

```
test <- test %>%
```

```
  mutate(fcst = forecasts$fcst$Рівень.зайнятості..[1:16],
         lower = forecasts$fcst$Рівень.зайнятості..[17:32],
         upper = forecasts$fcst$Рівень.зайнятості..[33:48])
```

```
test
```

```
MSE <- mean((test$`Рівень зайнятості %` - test$fcst)^2)
```

```
RMSE <- sqrt(MSE)
```

```
MSE
```

```
RMSE
```

```
serial <- serial.test(VAR_est, lags.pt = 12, type = "PT.asymptotic")
```

```
serial
```

```
plot(serial, names = "Рівень.зайнятості..")
```

```
grangertest(`Рівень зайнятості %` ~ `Середня заробітна плата грн.`, order = 5,
data = train)
```

```
grangertest(`Середня заробітна плата грн.` ~ `Рівень зайнятості %`, order = 5,
data = train)
```

```
autoplot(data, .vars = `Рівень зайнятості %`) +
```

```
  autolayer(test, .vars = `fcst`, colour = "red") +
```

```
  autolayer(test, .vars = `lower`, colour = "blue", linetype = "dashed") +
```

```
  autolayer(test, .vars = `upper`, colour = "blue", linetype = "dashed") +
```

```
  scale_color_manual(name='iivi',
```

```
breaks=c('Прогноз', 'Нижня межа', 'Верхня межа'),
values=c('Прогноз' ='red', 'Нижня межа' ='blue', 'Верхня межа'
='blue'))
```

```
VAR_data <- train[,c("Рівень зайнятості %", "Рівень безробіття
%", "Економічно активне населення тис. ос.", "Середня заробітна плата грн.", "Індекс
споживчих цін")]
```

```
lag <-
```

```
VARselect (VAR_data, lag.max = 10, type = "const")
```

```
lag
```

```
VAR_est <- VAR(VAR_data, p = 5, type = "const",
exog = NULL)
```

```
summary(VAR_est$varresult$Рівень.зайнятості.)
```

```
forecasts <- predict(VAR_est, n.ahead = 16)
```

```
forecasts
```

```
test <- test %>%
```

```
mutate(fcst = forecasts$fcst$Рівень.зайнятості..[1:16],
```

```
lower = forecasts$fcst$Рівень.зайнятості..[17:32],
```

```
upper = forecasts$fcst$Рівень.зайнятості..[33:48])
```

```
test
```

```
MSE <- mean((test$`Рівень зайнятості %` - test$fcst)^2)
```

```
RMSE <- sqrt(MSE)
```

```
MSE
```

```
RMSE
```

```

serial <- serial.test(VAR_est, lags.pt = 12, type = "PT.asymptotic")
serial

plot(serial, names = "Рівень.зайнятості..")

grangertest(`Рівень зайнятості %` ~ `Індекс споживчих цін`, order = 5, data =
train)
grangertest(`Індекс споживчих цін` ~ `Рівень зайнятості %`, order = 5, data =
train)

autoplot(data, .vars = `Рівень зайнятості %`) +
  autolayer(test, .vars = `fcst`, colour = "red") +
  autolayer(test, .vars = `lower`, colour = "blue", linetype = "dashed") +
  autolayer(test, .vars = `upper`, colour = "blue", linetype = "dashed") +
  scale_color_manual(name='iivi',
    breaks=c('Прогноз', 'Нижня межа', 'Верхня межа'),
    values=c('Прогноз' ='red', 'Нижня межа' ='blue', 'Верхня межа
'='blue'))

VAR_data <- train[,c("Рівень зайнятості %", "Рівень безробіття
%", "Економічно активне населення тис. ос.", "Середня заробітна плата грн.", "Індекс
споживчих цін", "Прямі інвестиції млн. дол.")]

lag <-
  VARselect (VAR_data, lag.max = 10, type = "const")
lag

```

```

VAR_est <- VAR(VAR_data, p = 5, type = "const", season = NULL,
               exog = NULL)
summary(VAR_est$varresult$Рівень.зайнятості.)

forecasts <- predict(VAR_est, n.ahead = 16)
forecasts

test <- test %>%
  mutate(fcst = forecasts$fcst$Рівень.зайнятості..[1:16],
         lower = forecasts$fcst$Рівень.зайнятості..[17:32],
         upper = forecasts$fcst$Рівень.зайнятості..[33:48])
test

MSE <- mean((test$`Рівень зайнятості %` - test$fcst)^2)
RMSE <- sqrt(MSE)
MSE
RMSE

serial <- serial.test(VAR_est, lags.pt = 12, type = "PT.asymptotic")
serial

plot(serial, names = "Рівень.зайнятості..")

grangertest(`Рівень зайнятості %` ~ `Прямі інвестиції млн. дол.`, order = 5,
data = train)
grangertest(`Прямі інвестиції млн. дол.` ~ `Рівень зайнятості %`, order = 5,
data = train)

```

```

autoplot(data, .vars = `Рівень зайнятості %`) +
  autolayer(test, .vars = `fcst`, colour = "red") +
  autolayer(test, .vars = `lower`, colour = "blue", linetype = "dashed") +
  autolayer(test, .vars = `upper`, colour = "blue", linetype = "dashed") +
  scale_color_manual(name='iivi',
    breaks=c('Прогноз', 'Нижня межа', 'Верхня межа'),
    values=c('Прогноз' ='red', 'Нижня межа' ='blue', 'Верхня межа'
    ='blue'))

```

```

VAR_data <- train[,c("Рівень зайнятості %", "Рівень безробіття
%", "Економічно активне населення тис. ос.", "Середня заробітна плата грн.", "Індекс
споживчих цін", "Прямі інвестиції млн. дол.", "Номинальний ВВП млн.грн" )]

```

```
lag <-
```

```
VARselect (VAR_data, lag.max = 10, type = "const")
```

```
lag
```

```
VAR_est <- VAR(VAR_data, p = 6, type = "const",
```

```
exog = NULL)
```

```
summary(VAR_est$varresult$Рівень.зайнятості.)
```

```
forecasts <- predict(VAR_est, n.ahead = 16)
```

```
forecasts
```

```
test <- test %>%
```

```
mutate(fcst = forecasts$fcst$Рівень.зайнятості..[1:16],
```

```
lower = forecasts$fcst$Рівень.зайнятості..[17:32],
```

```
upper = forecasts$fcst$Рівень.зайнятості..[33:48])
```

```
test
```

```
MSE <- mean((test$`Рівень зайнятості %` - test$fcst)^2)
```

```
RMSE <- sqrt(MSE)
```

```
MSE
```

```
RMSE
```

```
serial <- serial.test(VAR_est, lags.pt = 12, type = "PT.asymptotic")
```

```
serial
```

```
plot(serial, names = "Рівень зайнятості..")
```

```
grangertest(`Рівень зайнятості %` ~ `Номінальний ВВП млн.грн`, order = 6,  
data = train)
```

```
grangertest(`Номінальний ВВП млн.грн` ~ `Рівень зайнятості %`, order = 6,  
data = train)
```

```
autoplot(data, .vars = `Рівень зайнятості %`) +  
  autolayer(test, .vars = `fcst`, colour = "red") +  
  autolayer(test, .vars = `lower`, colour = "blue", linetype = "dashed") +  
  autolayer(test, .vars = `upper`, colour = "blue", linetype = "dashed") +  
  scale_color_manual(name='iivi',  
                      breaks=c('Прогноз', 'Нижня межа', 'Верхня межа'),  
                      values=c('Прогноз' ='red', 'Нижня межа' ='blue', 'Верхня межа  
'='blue'))
```

```
forecasts <- predict(VAR_est, n.ahead = 20)
```

```
forecasts
```

```
df <- data.frame(Період = yearquarter(c("2023 Q1", "2023 Q2", "2023 Q3",  
"2023 Q4")),  
                var2= c(NA),
```

```

var3= c(NA),
var4= c(NA),
var5= c(NA),
var6= c(NA),
var7= c(NA),
var8= c(NA))

```

```

fcst_2023 <- bind_rows(test, df)
fcst_2023 <- fcst_2023 %>%
  dplyr::select(1:8) %>%
  mutate(fcst = forecasts$fcst$Рівень.зайнятості..[1:20],
         lower = forecasts$fcst$Рівень.зайнятості..[21:40],
         upper = forecasts$fcst$Рівень.зайнятості..[41:60])
autoplot(data, .vars = `Рівень зайнятості %`) +
  autolayer(fcst_2023, .vars = `fcst`, colour = "red") +
  autolayer(fcst_2023, .vars = `lower`, colour = "blue", linetype = "dashed") +
  autolayer(fcst_2023, .vars = `upper`, colour = "blue", linetype = "dashed") +
  scale_color_manual(name='iivі',
                     breaks=c('Прогноз', 'Нижня межа', 'Верхня межа'),
                     values=c('Прогноз' ='red', 'Нижня межа' ='blue', 'Верхня межа
='blue'))

```

```

data_HR <- readr::read_csv('data_attrition.csv')

glimpse(data_HR)

data_HR <- data_HR %>%

  dplyr::select(-c(Timestamp,    project_domen,company_type,    company_size,
overtime, curr_location))

glimpse(data_HR)


data_HR <- data_HR%>%

  drop_na() %>%

  unique()


data_HR %>%

  ggplot(aes(x = gender, color = gender, fill = gender )) +

  geom_bar()


library(forcats)


data_HR %>%

  ggplot(aes(x = fct_rev(fct_infreq(education)), color = education, fill = education
)) +

  geom_bar() + coord_flip()

```

```

data_HR %>%

  ggplot(aes(age)) +

  geom_histogram(colour = 4, fill = "grey", bins = 30) +

  geom_vline(xintercept = median(data_HR$age), color="red")

set.seed(2000)

split <- initial_split(data_HR, prop = 0.70, strata = Attrition)

split

split_test <- testing(split)

split_test

split_train <- training(split)

split_train

library(discrim)

library(caret)

library(randomForest)

library(fastDummies)

# Naive Bayes

nb_spec <- naive_Bayes() %>%

  set_mode('classification') %>%

  set_engine('naivebayes') %>%

```

```

set_args(usekernel = FALSE)

nb_2var <- recipe(Attrition ~ salary_satisfaction + fin_status,
                  data = split_train) %>%

  step_dummy(salary_satisfaction, fin_status) %>%

  prep(training = split_train)

nb_model <- workflow() %>%

  add_model(nb_spec) %>%

  add_recipe(nb_2var)

nb_model

nb_fit <- nb_model %>%

  fit(split_train)

nb_fit

augment(nb_fit, new_data = split_test) %>%

  conf_mat(truth = Attrition, estimate = .pred_class)

result_test <- split_test %>%

  select(Attrition) %>%

  bind_cols(

    predict(nb_fit, new_data = split_test),

    predict(nb_fit, new_data = split_test, type = "prob")

  )

```

```

result_test

result_test %>%

  roc_curve(Attrition, .pred_1, event_level = "second") %>%

  autoplot()

result_test <-

  as.data.frame(result_test)

multi_metric <-

  bind_rows(yardstick::accuracy(result_test, truth = Attrition, estimate =
.pred_class),

    yardstick::recall(result_test, truth = Attrition, estimate = .pred_class),

    roc_auc(result_test, Attrition, .pred_1, event_level = "second"),

    yardstick::precision(result_test, Attrition, .pred_class),

    f_meas(result_test, Attrition, .pred_class))

multi_metric

nb_4var <- recipe(Attrition ~ salary_satisfaction + fin_status + work_experience
,

  data = split_train) %>%

  step_dummy(salary_satisfaction, fin_status) %>%

  prep(training = split_train)

nb_model <- workflow() %>%

  add_model(nb_spec) %>%

```

```

  add_recipe(nb_4var)

nb_model

nb_fit <- nb_model %>%
  fit(split_train)

nb_fit

augment(nb_fit, new_data = split_test) %>%
  conf_mat(truth = Attrition, estimate = .pred_class)

result_test <- split_test %>%
  select(Attrition) %>%
  bind_cols(
    predict(nb_fit, new_data = split_test),
    predict(nb_fit, new_data = split_test, type = "prob")
  )

result_test

result_test %>%
  roc_curve(Attrition, .pred_1, event_level = "second") %>%
  autoplot()

result_test <-
  as.data.frame(result_test)

multi_metric <-

```

```

    bind_rows(yardstick::accuracy(result_test, truth = Attrition, estimate =
.pred_class),

    yardstick::recall(result_test, truth = Attrition, estimate = .pred_class),

    roc_auc(result_test, Attrition, .pred_1, event_level = "second"),

    yardstick::precision(result_test, Attrition, .pred_class),

    f_meas(result_test, Attrition, .pred_class))

multi_metric

nb_6var <- recipe(Attrition ~ salary_satisfaction + fin_status + work_experience
+ education + age + job_title,

    data = split_train) %>%

    step_dummy(salary_satisfaction, fin_status, education, job_title) %>%

    prep(training = split_train)

nb_model <- workflow() %>%

    add_model(nb_spec) %>%

    add_recipe(nb_6var)

nb_model

nb_fit <- nb_model %>%

    fit(split_train)

nb_fit

augment(nb_fit, new_data = split_test) %>%

    conf_mat(truth = Attrition, estimate = .pred_class)

result_test <- split_test %>%

```

```

select(Attrition) %>%

bind_cols(

  predict(nb_fit, new_data = split_test),

  predict(nb_fit, new_data = split_test, type = "prob")

)

result_test

result_test %>%

  roc_curve(Attrition, .pred_1, event_level = "second") %>%

  autoplot()

result_test <-

  as.data.frame(result_test)

multi_metric <-

  bind_rows(yardstick::accuracy(result_test, truth = Attrition, estimate =
.pred_class),

    yardstick::recall(result_test, truth = Attrition, estimate = .pred_class),

    roc_auc(result_test, Attrition, .pred_1, event_level = "second"),

    yardstick::precision(result_test, Attrition, .pred_class),

    f_meas(result_test, Attrition, .pred_class))

multi_metric

nb_8var <- recipe(Attrition ~ salary_satisfaction + fin_status + work_experience
+ education + age + job_title + gender,

  data = split_train) %>%

```

```

step_dummy(salary_satisfaction, fin_status, education, job_title, gender) %>%
prep(training = split_train)

nb_model <- workflow() %>%
  add_model(nb_spec) %>%
  add_recipe(nb_8var)
nb_model
nb_fit <- nb_model %>%
  fit(split_train)
nb_fit
augment(nb_fit, new_data = split_test) %>%
  conf_mat(truth = Attrition, estimate = .pred_class)
result_test <- split_test %>%
  select(Attrition) %>%
  bind_cols(
    predict(nb_fit, new_data = split_test),
    predict(nb_fit, new_data = split_test, type = "prob")
  )
result_test
result_test %>%
  roc_curve(Attrition, .pred_1, event_level = "second") %>%

```

```

    autoplot()

result_test <-

  as.data.frame(result_test)

multi_metric <-

  bind_rows(yardstick::accuracy(result_test, truth = Attrition, estimate =
.pred_class),

            yardstick::recall(result_test, truth = Attrition, estimate = .pred_class),

            roc_auc(result_test, Attrition, .pred_1, event_level = "second"),

            yardstick::precision(result_test, Attrition, .pred_class),

            f_meas(result_test, Attrition, .pred_class))

multi_metric

# KNN

process <-

  preProcess(data_HR, method=c("range"))

data_scaled <-

  predict(process, data_HR)

split <- initial_split(data_scaled, prop = 0.70, strata = Attrition)

split

train_scaled <- training(split)

```

```

test_scaled <- testing(split)

knn_spec <- nearest_neighbor (neighbors = 20) %>%

  set_mode('classification') %>%

  set_engine('kknn') %>%

  set_args(usekernel = FALSE)


knn <- recipe(Attrition ~ salary_satisfaction + fin_status + work_experience +
education + age + job_title + job_pos + gender,

              data = train_scaled) %>%

  step_dummy(salary_satisfaction, gender, fin_status, education, job_title,
job_pos) %>%

  prep(training = train_scaled)

knn_model <- workflow() %>%

  add_model(knn_spec) %>%

  add_recipe(knn)

knn_model

knn_fit <- knn_model %>%

  fit(train_scaled)

knn_fit

augment(knn_fit, new_data = test_scaled) %>%

  conf_mat(truth = Attrition, estimate = .pred_class)

result_test <- test_scaled %>%

```

```

select(Attrition) %>%

bind_cols(

  predict(knn_fit, new_data = test_scaled),

  predict(knn_fit, new_data = test_scaled, type = "prob")

)

result_test

result_test %>%

  roc_curve(Attrition, .pred_1, event_level = "second") %>%

  autoplot()

result_test <-

  as.data.frame(result_test)

multi_metric <-

  bind_rows(yardstick::accuracy(result_test, truth = Attrition, estimate =
.pred_class),

    yardstick::recall(result_test, truth = Attrition, estimate = .pred_class),

    roc_auc(result_test, Attrition, .pred_1, event_level = "second"),

    yardstick::precision(result_test, Attrition, .pred_class),

    f_meas(result_test, Attrition, .pred_class))

multi_metric

# Random forest

```

```

split <- initial_split(data_HR, prop = 0.70, strata = Attrition)

split

split_test <- testing(split)

split_test

split_train <- training(split)

split_train

tree_spec <- rand_forest(

  mtry = tune(),

  trees = 1000,

  min_n = tune()

) %>%

  set_mode("classification") %>%

  set_engine("ranger")

tree <- recipe(Attrition ~ salary_satisfaction + fin_status + work_experience +
education + age + job_title + gender + job_pos ,

  data = split_train) %>%

  step_dummy(salary_satisfaction, fin_status, education, job_title, gender,
job_pos) %>%

  prep(training = split_train)

set.seed(123)

trees_folds <- vfold_cv(split_train)

tune_wf <- workflow() %>%

```

```

add_recipe(tree) %>%

add_model(tree_spec)

tune_tree <- tune_grid(

  tune_wf,

  resamples = trees_folds,

  grid = 20

)

tune_tree

tree_prep <- prep(tree)

juiced <- juice(tree_prep)

tune_tree %>%

  collect_metrics() %>%

  filter(.metric == "roc_auc") %>%

  select(mean, min_n, mtry) %>%

  pivot_longer(min_n:mtry,

               values_to = "value",

               names_to = "parameter"

  ) %>%

  ggplot(aes(value, mean, color = parameter)) +

  geom_point(show.legend = FALSE) +

  facet_wrap(~parameter, scales = "free_x") +

```

```

  labs(x = NULL, y = "AUC")

tree_grid <- grid_regular(
  mtry(range = c(10, 20)),
  min_n(range = c(20, 40)),
  levels = 5
)

tree_grid

tree_reg <- tune_grid(
  tune_wf,
  resamples = trees_folds,
  grid = tree_grid
)

tree_reg

tree_reg %>%

  collect_metrics() %>%

  filter(.metric == "roc_auc") %>%

  mutate(min_n = factor(min_n)) %>%

  ggplot(aes(mtry, mean, color = min_n)) +
  geom_line(alpha = 0.5, size = 1.5) +
  geom_point() +
  labs(y = "AUC")

```

```

best_auc <- select_best(tree_reg, "roc_auc")

final_rf <- finalize_model(tree_spec, best_auc)

final_rf

final_wf <- workflow() %>%

  add_recipe(tree) %>%

  add_model(final_rf)

final_res <- final_wf %>%

  last_fit(split)

final_res$.predictions

final_res %>%

  collect_predictions() %>%

  conf_mat(truth = Attrition, estimate = .pred_class)

final_res %>%

  collect_metrics()

result <- final_res %>%

  collect_predictions()

final_res %>%

  collect_predictions() %>%

  roc_curve(Attrition, .pred_1, event_level = "second") %>%

  autoplot()

multi_metric <-

```

```

      bind_rows(yardstick::accuracy(result, truth = Attrition, estimate =
.pred_class),

      yardstick::recall(result, truth = Attrition, estimate = .pred_class),

      roc_auc(result, Attrition, .pred_1,event_level = "second"),

      yardstick::precision(result, Attrition, .pred_class),

      f_meas(result, Attrition, .pred_class))

multi_metric

library(vip)

final_rf %>%

  set_engine("ranger", importance = "permutation") %>%

  fit(Attrition ~ .,

      data = juiced

  ) %>%

  vip(geom = "point")

```