Вінницький національний технічний університет
Факультет інформаційних технологій та комп'ютерної інженерії
Кафедра обчислювальної техніки

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему:

МЕТОДИ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ У ДАТА-МАЙНІНГУ

Виконав студент 2 курсу, групи 1КІ-22м
спеціальності123 — Комп'ютерна інженерія
Мазуренко В.В.
Керівник к.пед.н., доц. каф. ОТ
Добровольська Н.В.
"11" 12 2023 р.
Опонент д.т.н., проф. каф. ПЗ
Ліщинська Л.Б.
"12" 12 2023 р.

Допущено до захисту
Завідувач кафедри ОТ
д.т.н., проф. Азаров О.Д.

"15" 12 2023 р.

ВНТУ 2023

ВІННИЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

Факультет інформаційних технологій та комп'ютерної інженерії

Кафедра обчислювальної техніки

Галузь знань — Інформаційні технології

Освітній рівень — магістр

Спеціальність — 123 Комп'ютерна інженерія

Освітньо-професійна програма — Комп'ютерна інженерія

ЗАТВЕРДЖУЮ

Завідувач кафедри
обчислювальної техніки
_____ проф., д.т.н. О.Д. Азаров
"26" вересня 2023 р.

З А В Д А Н Н Я

НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ

студенту **Мазуренку Владиславу Володимировичу**

1 Тема роботи «Методи машинного навчання для прогнозування у дата-майнінгу» керівник роботи Добровольська Наталя Вікторівна к.пед.н., доцент, затверджено наказом вищого навчального закладу від 18.09.2023р. року № 247.

2 Строк подання студентом роботи 13.12.2023р.

3 Вихідні дані до роботи: структура системи прогнозування в дата-майнінгу, сервер бази даних Microsoft SQL Server, основна методика створення інформаційної технології є Model-view-controller.

4 Зміст розрахунково-пояснювальної записки: вступ, аналіз методів прогнозування в дата-майнінгу та здійснення обгрунтування вибору методу аналізу для прогнозування в дата-майнінгу, розробка архітектури веб-додатку та алгоритму прогнозування в дата-майнінгу, експериментальне дослідження ефективності розробленого веб-додатку, економічна частина.

5 Перелік графічного матеріалу: технічне завдання, схема Model-view-

controller, схема реалізації запропонованої моделі прогнозування часового ряду, діаграма формування часового ряду та представлення загальної моделі ряду.

6 Консультанти розділів роботи приведені в таблиці 1.

Таблиця 1— Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1-4	Добровольська Н.В., к.пед.н., доц.каф. ОТ		
5	Небава М.І., к.е.н., проф.каф. ЕПВМ		

7 Дата видачі завдання 19.09.2023р.

8 Календарний план виконання МКР приведений в таблиці 2.

Таблиця 2 — Календарний план

№ з/п	Назва етапів МКР	Строк виконання	Підпис
1	Постановка задачі	21.09.2023	
2	Огляд та аналіз підходів до прогнозування в дата-майнінгу	26.09.2023	
3	Дослідження математичних моделей	03.10.2023	
4	Моделювання оптимізаційного рішення	10.10.2023	
5	Постановка задачі на проектування веб-додатку	17.10.2023	
6	Розробка структури веб-додатку	31.10.2023	
7	Тестування веб-додатку	05.11.2023	
8	Розрахунок економічної частини	18.11.2023	
9	Оформлення пояснювальної записки та ілюстративного матеріалу	24.11.2023	
10	Перевірка якості виконання магістерської кваліфікаційної роботи та усунення недоліків	04.12.2023	

Студент Мазуренко Владислав Володимирович

Керівник к.пед.н., доц. Добровольська Наталя Вікторівна

АНОТАЦІЯ

УДК 004.056

Мазуренко В. В. Методи машинного навчання для прогнозування у дата-майнінгу. Магістерська кваліфікаційна робота зі спеціальності 123 — Комп’ютерна Інженерія, Вінниця: ВНТУ, 2023 — 129 с. На укр. мові. Бібліогр.: 59 назв; рис.: 42; табл. 8.

У роботі здійснено огляд теорії в дата-майнінгу; проведено аналіз методів прогнозування в дата-майнінгу та здійснено обґрунтування вибору методу аналізу для прогнозування в дата-майнінгу. Також розкрито методику розробки веб-додатків для прогнозування в дата-майнінгу та розроблено архітектуру веб-додатків та алгоритми прогнозування в дата-майнінгу. У роботі описано додаток для прогнозування в дата-майнінгу та проведено експериментальне дослідження ефективності розробленого веб-додатку.

Ключові слова: дата-майнінг, управління, алгоритм, прогнозування.

ABSTRACT

Surname I. B. Machine learning methods for forecasting in data mining. Master's thesis on specialty 123 — Computer Engineering, Vinnytsia: VNTU, 2023 — 129 p.

In Ukrainian speech Bibliography: 59 titles; Fig.: 42; table 8.

The paper reviews the theory in data mining; the analysis of forecasting methods in data mining was carried out and the justification of the choice of the analysis method for forecasting in data mining was made. The methodology for developing web applications for data mining forecasting is also disclosed, and the architecture of web applications and data mining forecasting algorithms are developed. The paper describes an application for forecasting in data mining and conducts an experimental study of the effectiveness of the developed web application.

Keywords: data mining, management, algorithm, forecasting.

ЗМІСТ

ВСТУП 8

1 СУТНІСТЬ ТА ОСОБЛИВОСТІ ПРОГНОЗУВАННЯ В ДАТА-МАЙНІНГУ 11

1.1 Поняття дата-майнінгу 11

1.2 Механізми прогнозування 15

1.3 Огляд та аналіз аналогів 33

2 МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ЯК ІНСТРУМЕНТ ДЛЯ РОЗРОБКИ ВЕБ-ДОДАТКУ ДЛЯ ПРОГНОЗУВАННЯ В ДАТА-МАЙНІНГУ 41

2.1 Аналіз підходів до прогнозування в дата-майнінгу 41

2.2 Дослідження математичних моделей для прогнозування в дата-майнінгу 46

2.3 Моделювання оптимізаційного рішення 53

3 ПРОГРАМНА РЕАЛІЗАЦІЯ ВЕБ-ДОДАТКУ ДЛЯ ПРОГНОЗУВАННЯ В ДАТА-МАЙНІНГУ 62

3.1 Постановка задачі на проектування веб-додатку для прогнозування в дата-майнінгу 62

3.2 Інформаційне забезпечення системи 63

3.3 Розробка структури веб-додатку для прогнозування в дата-майнінгу 66

4 ПЕРЕВІРКА ПРАЦЕЗДАТНОСТІ ТА ТЕСТУВАННЯ 70

4.1 Тестування веб-додатку для прогнозування в дата-майнінгу 70

4.2 Верифікація результатів 73

5 ЕКОНОМІЧНА ЧАСТИНА 94
5.1 Проведення комерційного та технологічного аудиту науково-технічної розробки 94
5.2 Визначення рівня конкурентоспроможності розробки 98
5.3 Розрахунок витрат на проведення науково-дослідної роботи 101
ВИСНОВКИ 113
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ 115
ДОДАТОК А Технічне завдання 120
ДОДАТОК Б Результати розрахунку 123
ДОДАТОК В Результати розрахунку 126
ДОДАТОК Г Результати розрахунку за методом Вінтерса 129
ДОДАТОК Д Протокол перевірки кваліфікаційної роботи 132

ВСТУП

Роль прогнозування і масштаби його практичного застосування в сучасній економіці значно зросли. Незважаючи на це, питань «що буде» стає все більше і більше, а впевнених відповідей — менше і менше. Довготривалі реформи породжують процеси, закономірності розвитку яких недостатньо вивчені вітчизняною економічною наукою. Відсутність чіткого розуміння внутрішніх механізмів цих процесів, а також відсутність стійкої моделі економічного розвитку країни ускладнюють використання старої бази знань для побудови адекватних моделей та отримання надійних прогнозів.

Розвиток адаптивного прогнозування економічних процесів в даний час здійснюється в двох напрямках. Перший напрям пов'язаний з ускладненням структурних характеристик прогнозних моделей, а другий — з вдосконаленням їх адаптивного механізму. Кожен з напрямків отримав достатній розвиток в моделях прогнозування по одновимірним часовим рядам. У меншій мірі це відноситься до моделей, що використовуються в аналітичних цілях. І зовсім не розроблені питання, пов'язані з проблемою розширення сфери застосування принципів адаптації при вирішенні сучасних завдань перспективного аналізу процесів економіки перехідного періоду.

Саме така мета переслідується у магістерській роботі, що і зумовлює її **актуальність.**

Метою роботи є спрощення прогнозування в дата-майнінгу та пришвидшення отримання результатів.

Досягнення поставленої мети пов'язане з дослідженням динамічних ефектів сучасних економічних процесів, створенням спільних принципів побудови і методичних основ практичного використання в перспективному аналізі.

Для досягнення поставленої мети у роботі необхідно виконати низку **завдань**:

— здійснити огляд теорії в дата-майнінгу;

— провести аналіз методів прогнозування в дата-майнінгу та здійснити обґрунтування вибору методу аналізу для прогнозування в дата-майнінгу;

— розкрити методику розробки веб-додатків для прогнозування в дата-майнінгу;

— розробити архітектуру веб-додатків та алгоритми прогнозування в дата-майнінгу;

— описати додаток для прогнозування в дата-майнінгу;

— провести експериментальне дослідження ефективності розробленого веб-додатку;

— здійснити порівняння результатів з результатами, отриманими іншими методами прогнозування часових рядів;

— провести аналіз отриманих результатів.

Об’єктом роботи є процеси прогнозування в дата-майнінгу.

Предметом дослідження є методи і засоби прогнозування часових рядів в дата-майнінгу.

Проведені дослідження заклали основи регулярного застосування принципів аналізу в математичному апараті перспективного аналізу процесів, що характеризують економічну систему на етапі зміни орієнтирів її розвитку.

Комбінування розробленого апарату з ідеями імітаційного моделювання, дозволяє конкретні моделі наділяти механізмом відтворення властивостей реальних процесів, є новим внеском в методологію застосування математичних засобів для дослідження економічної динаміки в умовах часткової або повної невизначеності.

Наукова новизна отриманих результатів полягає в тому, що вперше були вдосконалені методи машинного навчання для прогнозування в дата-майнінгу, особливістю яких є спрощення реалізації процесу прогнозування та пришвидшення отримання результатів їх роботи. Це стало можливим за рахунок формування єдиної системи прогнозу, яка базується на сучасних алгоритмах прогнозування та механізмах економічного аналізу.

Практичне значення отриманих результатів полягає в тому, що

запропоновано веб-додаток для прогнозування часових рядів, який дозволяє зменшити кількість помилок та спрощує процес в цілому. Використані науково-технічні рішення дозволили реалізувати та адаптувати веб-додаток для прогнозування часових рядів.

Апробація результатів магістерської роботи буде здійснено на міжнародній науково-практичній Інтернет-конференції студентів, аспірантів та молодих науковців «МОЛОДЬ В НАУЦІ: ДОСЛІДЖЕННЯ, ПРОБЛЕМИ, ПЕРСПЕКТИВИ (МН-2024)».

Матеріали роботи опубліковувались:

В. В. Мазуренко Н. В. Добровольська / Методи машинного навчання для прогнозування у дата-майнінгу // Тези доповіді. Міжнародна науково-практична Інтернет-конференція студентів, аспірантів та молодих науковців «МОЛОДЬ В НАУЦІ: ДОСЛІДЖЕННЯ, ПРОБЛЕМИ, ПЕРСПЕКТИВИ (МН-2024)»

1 СУТНІСТЬ ТА ОСОБЛИВОСТІ ПРОГНОЗУВАННЯ В ДАТА-МАЙНІНГУ

1.1 Поняття дата-майнінгу

Термін Data Mining походить від двох понять: пошук цінної інформації у великих базах даних (data) і видобуток даних (mining). Обидва процеси вимагають просіювання величезної кількості сировини або проведення інтелектуальних досліджень, щоб знайти бажану цінність.

Термін Data Mining стосується інтелектуального аналізу даних, вилучення інформації, вилучення даних, інструментів пошуку шаблонів, вилучення знань, аналізу шаблонів, вилучення частинок знань із гори даних, вилучення даних у базу даних, інформації в даних і часто перекладається як аналіз . Очищення даних. Поняття «виявлення знань у базах даних» (KDD) можна вважати синонімом інтелектуального аналізу даних [1].

Концепція інтелектуального аналізу даних, що виникла в 1978 році, набула популярності в сучасній інтерпретації приблизно з першої половини 1990-х років. Вже протягом багатьох років обробка та аналіз даних здійснюються в прикладних рамках статистики, при цьому в основному вирішуються невеликі завдання обробки баз даних.

Інтелектуальний аналіз даних — це міждисциплінарний напрямок, який виник і розвивається на базі таких наук, як прикладна статистика, розпізнавання образів і синтетичний аналіз.

Інтелект, теорія баз даних тощо. Інтелектуальний аналіз даних – це процес виявлення раніше невідомих, нетривіальних, практично корисних і доступних інтерпретацій у необроблених даних знань, необхідних для прийняття рішень у різних сферах людської діяльності.

Суть і призначення технології інтелектуального аналізу даних можна описати наступним чином: техніка, призначена для пошуку неочевидних, об’єктивних і практично корисних закономірностей у великих обсягах даних.

Нечітко означає, що існуючі шаблони не виявлені стандартними методами

обробки інформації або експертно.

Об'єктивних — це означає, що виявлені закономірності повні і будуть відповідати дійсності, на відміну від експертного висновку, який завжди є суб'єктивним.

Практично корисні — це означає, що висновки мають конкретний зміст, який можна використовувати на практиці.

Знання — це сукупність інформації, яка формує послідовний опис, що відповідає певному рівню усвідомлення описуваного питання, теми, проблеми тощо.

Використання знань (поширення знань) означає реальне застосування, що приносять певні переваги (наприклад, у конкурентний ринок).

В основу технології Data Mining покладено концепцію шаблонів (patterns), які є закономірностями, властивими підвиборкам даних, які можуть бути виражені у формі, зрозумілій людині.

Традиційні методи аналізу даних (статистичні методи) і OLAP зосереджуються в основному на перевірці попередньо сформульованих гіпотез (видобуток даних на основі валідації) та «грубий» аналіз розвідки формує основу для обробки даних оперативного аналізу (OnLine Analytical обробка, OLAP), і одним з основних положень інтелектуального аналізу даних є пошук неочевидних закономірностей. Інструменти аналізу даних можуть виявити наступні закономірності самостійно встановлюють гіпотези про взаємозв'язки. Тому що найскладніше — це сформулювати припущення про залежність, переваги аналізу даних порівняно з іншими методами аналізу є очевидними.

Більшість статистичних методів, які використовуються для виявлення зв’язків у даних використовуючи концепцію усереднення вибірки, це призводить до праці з неіснуючими значеннями, тоді як аналіз даних працює з реальними значеннями.

Значення OLAP найкраще підходить для розуміння ретроспективних даних, дата-майнінг покладається на ретроспективні дані, щоб отримати відповіді на запитання про майбутнє.

Потенціал інтелектуального аналізу даних дає «зелене світло» для розширення кордонів застосування технології. Що стосується перспектив інтелектуального аналізу даних, то тут можуть бути наступні напрямки розвитку:

— вибір типів предметної області та відповідну їй предметну область Евристики, формалізація яких допоможе у вирішенні актуальних завдань з даними гірничодобувна промисловість, пов'язана з цими районами;

— створення формальних мов та логічних засобів, за допомогою яких буде формалізовано міркування та автоматизація яких стане інструментом вирішення завдань Data Mining у конкретних предметних галузях;

— створення методів Data Mining, здатних як витягувати з даних закономірності, так і формувати деякі теорії, що спираються на емпіричні дані;

— подолання суттєвого відставання можливостей інструментальних засобів Data Mining від теоретичних досягнень у цій галузі.

Якщо розглядати майбутнє інтелектуального аналізу даних у короткостроковій перспективі, то зрозуміло, що розвиток цієї технології в основному спрямований на сфери, пов’язані з бізнесом.

У короткостроковій перспективі продукти Data Mining можуть стати такими ж звичайними та необхідними, як електронна пошта, і, наприклад, використовуватись користувачами для пошуку найнижчих цін на певний товар або найдешевших квитків.

У довгостроковій перспективі майбутнє інтелектуального аналізу даних справді захоплююче — воно може призвести до пошуку нових методів лікування різних захворювань за допомогою інтелектуальних агентів, а також до нового розуміння універсалій.

Основною характеристикою інтелектуального аналізу даних є поєднання широкого спектру математичних інструментів (від класичного статистичного аналізу до нових математичних інструментів) та новітні досягнення в галузі інформації технології. У технології інтелектуального аналізу даних суворе і гармонійне поєднання формальних методів та неформальні методи аналізу, тобто кількісний та якісний аналіз даних.

Методи та алгоритми інтелектуального аналізу даних включають: штучні нейронні мережі, дерева рішень, символічні правила, метод найближчого сусіда та метод k-найближчого сусіда, метод опорного вектора, байєсовську мережу, лінійну регресію, кореляційний регресійний аналіз; ієрархічні методи кластерного аналізу, кластеризацію не- ієрархічні методи аналізу, включаючи алгоритми k-середніх і k-медіан; методи пошуку правил асоціації, особливо алгоритм Apriori; методи обмеженого вибору, еволюційне програмування та генетику Алгоритмів, різноманітні методи візуалізації даних та багато іншого.

Більшість аналітичних методів, що використовуються в технології дата Майнінг — це відомі математичні алгоритми і методи. Новизна їх застосування полягає в тому, що завдяки можливостям технологій і програмних засобів, що з'явилися, вони можуть бути використані для вирішення окремих конкретних задач. Слід зауважити, що більшість методів інтелектуального аналізу даних розроблені в рамках теоретичних рамок штучного інтелекту.

Останнім часом методи Data Mining [8] застосовуються для аналізу закономірностей часового ряду, вони призначені для виявлення різних шаблонів (паттернів) в часовому ряді [5, 6]. При цьому особливу цінність для виявлення таких шаблонів мають логічні методи. Ці методи дозволяють знаходити логічні if-then правила. Вони придатні для аналізу як числових, так і символьних послідовностей, та їх результати мають прозору інтерпретацію.

Разом з тим, при виборі того чи іншого методу пошуку слід спиратися на критерій, що відображає його здатність виявляти найбільш повні та точні if-then правила для кожного рядка часової матриці за прийнятний час. На жаль, відомі методи слабко здатні задовольняти цьому критерію.

Так, дерева рішень принципово не здатні знаходити «найкращі» комбінації даних. Вони реалізують наївний принцип послідовного перегляду ознак і виявляють фактично «уламки» реальних закономірностей, створюючи лише ілюзію логічного висновку.

Критерії відбору «хромосом» у генетичних алгоритмах і програмах використовуються, евристики та не гарантують знаходження «найкращого»

рішення. Як і в реальному житті, еволюція може «застрягти» у будь-які невиробничі галузі. Натомість ви можете навести приклад двох безперспективних батьків, які були б генетично виключені з еволюції алгоритмом, здатним виробляти ефективне потомство. Це особливо стає помітно під час вирішення високорозмірних завдань зі складними внутрішніми зв'язками.

1.2 Механізми прогнозування

Сучасна економічна теорія, пройшовши складну еволюцію, диференціюючи на безліч шкіл та напрямків, стійко відтворювала істотну відмінність у підходах до прогнозування. На початку 60-их років XX ст. для економічного прогнозування почали використовувати економетричні моделі [1]. Як зазначає О. Ю. Кучанський з цього часу структура економіки розвинених країн і методи економетричного аналізу зазнали кардинальних змін [2].

Сфера економіки відрізняється наявністю явищ, які еволюціонують і флуктують в часі. З плином часу змінюються ціни, економічні умови, режим протікання того чи іншого виробничого процесу. В. О. Андрієнко стверджує, що сукупність вимірів подібного роду показників протягом деякого періоду часу і являє часовий ряд [3].

В рамках сучасних досліджень, економіки та математичної науки, вона присвячена різним аспектам вирішення питання аналізу фінансових часових рядів. До них відносяться генетичні алгоритми, що характеризуються використанням випадкового вибору, комбінування та варіювання бажаних параметрів за допомогою механізмів, подібних до природного відбору в природі [4]; нейронні мережі, побудовані за принципом організації та функціонування біологічних нейронних мереж — живої мережі нервових клітин організму [5]; нечіткі методи, в основі яких лежить нечітка множина як об'єкт функція елемента, що належить множині, яка приймає будь-яке значення в діапазоні [0,1], а не тільки 1 або 0 [6]; та інші. Однак, важливими є наступні міркування. Слід зазначити, що при аналізі динаміки ринку жоден з наведених вище методів не може врахувати таку властивість ринку, як самоорганізація,

тому використовувати ці методи для розв'язання поставленої задачі недоцільно. Певною мірою цю проблему може вирішити теорія фракталів.

Масштабованість фінансових часових рядів з фрактальними властивостями вимагає впровадження єдиного універсального механізму аналізу крихкості динамічних процесів, що відбуваються в економічних системах. Пошук такого механізму є одним з найактуальніших завдань науки на сьогоднішній день.

Н. Л. Дорош наголошує, що часовим рядом називають послідовність вимірювань y_t, $t = [1 \dots N]$, яка, як правило, впорядкована за часом [7]. Дані типу часових рядів у рамках економічної теорії являють собою щоденні ціни на акції, курси валют, щотижневі і місячні обсяги продажів, річні обсяги виробництва, тощо.

Згідно з І. В. Бакланом, основними фрактальними властивостями часового ряду є показник Херста H, показник крихкості μ, розмірність Хаусдорфа D та спектральний показник b [5].

У 1951 році британський гідролог Гарольд Едвін Херст запропонував R/S аналіз або метод нормалізованого спреду як основний метод дослідження фрактальної структури фінансових ринків [8].

Фрактальна розмірність має два різновиди — D та A . Так, фрактальна розмірність D (D - це оцінка ступеня зламу ряду) визначається за формулою:

$$D = 2 - H \tag{1.1}$$

Б. Мандельброт показав, що фрактальна розмірність є оберненою до коефіцієнта Херста (Hurst H). Наприклад, при H = 0.5 фрактальна розмірність дорівнює 2(1/0.5), а при H = 0.8 фрактальна розмірність дорівнює 1.25(1/0.8).

Так, згідно з Б. Мандельбротом, фрактальна розмірність, що позначається літерою A (A — розмірність ймовірнісного простору, тобто оцінка товщини хвостів як функції густини ймовірності), обчислюється за формулою:

$$A = {}^{1}/_{H} \quad (1.2)$$

Таблиця 1.1 — Значення варіацій фрактальної розмірності

Позначення Найменування	Пряма лінія	Випадковий ряд	Нескінчений лінійний тренд
Показник Херста (H)	$H \approx 0$	$H = 0{,}5$	$H = 1$
Фрактальна розмірність D	$D \approx 0$	$D = 1{,}5$	$D = 1$
Фрактальна розмірність A	$A \to 0$	$A = 2$	$A = 1$

Показник Херста визначається на відрізку $[0, 1]$, і розраховується у наступних межах:

якщо $0{,}5 < H \leq 1$ — ціни є фрактальними, підтверджено справедливість гіпотези фрактального ринку (Fractal Market Hypothesis, FMH), існують "важкі хвости" у розподілі змінних та стійкі ряди, тобто зміни цін позитивно корелюють з чорним шумом, що свідчить про наявність трендів на ринку;

якщо $H = 0{,}5$ — ціни є стохастичними, підтверджується справедливість гіпотези ефективного ринку (EMH), рух цін фінансових активів є прикладом стохастичного броунівського руху (вінерівського процесу), часові ряди загалом є нормально розподіленими, відсутня кореляція у зміні вартості активів (пам'ять) та підтверджено існування білого шуму;

якщо $0 \leq H < 0{,}5$ — ціни фрактальні, підтверджено ПІВ, є "важкі хвости" у розподілі змінних, антиперсистентні ряди, тобто від'ємна кореляція у зміні цін, рожевий шум з частою зміною напрямку руху цін.

Теорія броунівського руху була розроблена Альбертом Ейнштейном у 1905 році [9]. Херст взяв формулу з роботи Ейнштейна за відправну точку:

$$R = \sqrt{T} \quad (1.3)$$

де R — відстань, пройдена броунівською часткою за час T,

T — показник часу.

Згідно з (1.3), броунівська частинка пройшла відстань, що дорівнює квадратному кореню з часу, витраченого на цей рух.

Розрахунок показника Херста можна провести за такою формулою:

$$R/S = (aN)^H \tag{1.4}$$

Звідси

$$H = \frac{log(R/S)}{log(aN)} \tag{1.5}$$

де H — показник Херста;

N — число періодів спостережень;

a — задана константа, позитивне число, є константою. Херст емпірично розрахував цю константу для порівняно короткострокових часових рядів природних явищ як 0.5;

S — середньоквадратичне відхилення ряду спостережень x:

$$S = \sqrt{1/N \sum_{i=1}^{N} (x_i - x_{cp})^2} \tag{1.6}$$

де x_{cp} — Середнє арифметичне число спостережень за N періодів:

$$x_{cp} = 1/N \sum_{i=1}^{N} x_i \tag{1.7}$$

R — розмах накопиченого відхилення Zu, даний показник є найбільш

важливим елементом показника Херста, його обчислюють за формулою:

$$R = \max_{l \le u \le N}(Z_u) - \min_{l \le u \le N}(Z_u) \quad (1.8)$$

де Z_u — накопичене відхилення ряду x від середнього x_{cp}:

$$Z_u = \sum_{i=1}^{u}(x_i - x_{cp}) \quad (1.9)$$

Однак, якщо ми використовуємо константу a 0,5, то коли кількість спостережень N невелика, індекс Херста схильний розглядати навіть випадкові ряди як персистентні, тобто як ряди з трендами, і завищує значення H. Тому для більш детального аналізу ринкових рядів слід використовувати константу $a = p / 2$.

Для часових рядів важливим є показник Херста. Значення H та D вказують на ступінь "розриву" або "шорсткості" часового ряду.

Однак, хоча він широко використовується в різних сферах аналізу даних, для розрахунку показника Херста потрібна велика вибірка даних.

Поняття фрактальної характеристики або індексу фрактальності описує Т. В. Кравець [10]. Ця функція дозволяє оцінити фрактальну розмірність часового ряду за кількома десятками вибірок

Для функції $f(t)$, що визначена на відрізку $[a, b]$, та яка має деяке кінцеве число точок розриву першого роду. Задамо рівномірний розподіл відрізка

$$\omega_n = [a = t_0 < t_1 < \cdots < t_n = b] \quad (1.10)$$

де $t_i - t_{i-1} = \delta = {(b-a)}/{n}, (i = 1,2 \dots n)$

Введемо позначення:

$$V_f(\delta) \equiv \sum_{i=1}^{n} A_i(\delta) \tag{1.11}$$

Величина $V_f(\delta)$ — це амплітудна варіація функції $f\,(t)$, яка відповідає масштабу розбиття δ на відрізку $[a, b]$.

Загальна площа мінімального покриття визначається, як

$$S_\mu(\delta) = V_f(\delta) \cdot \delta \tag{1.12}$$

Оскільки $S_\mu(\delta) \sim \delta^{2-D_\mu}$, то $V_f(\delta) = \delta^{-\mu}$

При помноженні (2.21) на δ^2 спостерігається наступна залежність:

$$V_f(\delta) \equiv \sum_{i=1}^{n} A_i(\delta) \sim \delta^{-\mu} \text{ при } \delta \to 0 \tag{1.13}$$

це говорить про наявність фрактальної розмірності D для $f(t)$, причому $D = \mu + 1$, тобто

$$\mu = D - D_T \tag{1.14}$$

де μ — це індекс фрактальності.

Виходячи з того, що амплітудна варіація функції $f\,(t)$, яка відповідає масштабу розбиття δ на відрізку $[a, b]$, швидко виходить на статечний асимптотичний режим, індекс μ потребує вдвічі менше даних порівняно з іншими фрактальними параметрами. Теоретично, при однаковій кількості даних, індекс фрактальності має обчислюватися точніше, ніж індекс Херста, оскільки для його визначення використовується мінімальне, а отже, оптимальне охоплення часового ряду.

Моделі та методи аналізу складають сукупність послідовних дій необхідних для функціонального уявлення результату. Їх загальна класифікація

наведена на рис. 1.1.

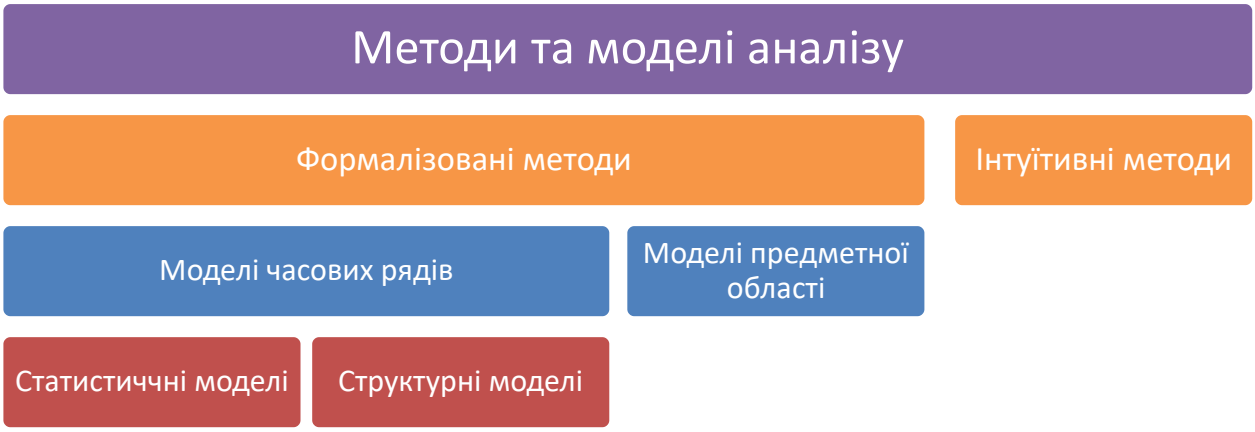


Рисунок 1.1 — Загальна класифікація моделей та методів аналізу

Метод прогнозування є поняттям ширшим за модель аналізу. На першому етапі методи прогнозування умовно діляться на дві групи: формалізовані, які за своєю суттю включають математичні моделі, тобто визначають таку математичну залежність, яка дозволяє обчислити майбутнє значення процесу, тобто зробити прогноз та інтуїтивні, які працюють з судженнями і оцінками експертів. Другі — методи розв'язання складних неформалізованих задач шляхом отримання оціночних прогнозів стану розвитку об'єкта в майбутньому, незалежно від інформаційного забезпечення методом експертних оцінок, застосовуються переважно в маркетингу, економіці та політиці [11], оскільки система, поведінку якої необхідно спрогнозувати, або занадто складна і не може бути описана математично, або занадто проста і не потребує такого опису.

Формалізовані методи прогнозування поділяються на моделі предметної області та моделі часових рядів, які є математичними моделями прогнозування.

Моделі предметної області — це математичні моделі прогнозування, побудовані за законами певної предметної області. Такі моделі використовують залежності, характерні для конкретної предметної області. Такі моделі характеризуються індивідуальним підходом до розробки.

Моделі часових рядів — це математичні моделі прогнозування, які намагаються знайти залежність майбутнього значення від минулого в самому процесі і розрахувати прогноз на основі цієї залежності. Ці моделі є

універсальними для різних предметних областей, тобто їх загальна форма не змінюється залежно від природи часового ряду.

Останні поділяють на статистичні та структурні. Статистичні методи це різновид математичних методів прогнозування, що дозволяють побудувати динамічні ряди на перспективу. Статистичні методи прогнозування охоплюють:

— розробка, вивчення та застосування сучасних математико-статистичних методів прогнозування на основі об'єктивних даних (включаючи непараметричні методи найменших квадратів, адаптивні методи, методи авторегресії та інші, з оцінкою точності прогнозування);

— розвиток теорії та практики ймовірнісно-статистичного моделювання методів експертного прогнозування, включаючи методи аналізу суб'єктивних експертних оцінок на основі статистики нечислових даних;

— розробка, дослідження та застосування методів прогнозування в умовах ризику та загальноекономічних, математичних та економетричних (як математико-статистичних, так і експертних) моделей.

Наукова база статистичних методів прогнозування — прикладна статистика і теорія прийняття рішень розкрита в працях Т. О. Семененко та М. Я. Кушніра [12, 13].

Статистичні моделі (статистичні методи), в яких функціональний зв'язок між майбутніми та фактичними значеннями часового ряду оцінюється на основі часового ряду та зовнішніх факторів, якщо такі враховуються, задана аналітично, тобто формулою варто віднести:

— множинну лінійну регресію;

— стохастичні часові ряди;

— загальне експоненціальне згладжування;

— простір станів;

— опорні вектори регресії.

Структурні моделі (методи штучного інтелекту) — це моделі, в яких функціональний зв'язок між майбутніми і фактичними значеннями часового ряду і зовнішніми факторами представлений структурно, наприклад, графічно:

— нейромережеві моделі;

— моделі на базі ланцюгів Маркова;

— моделі на базі класифікаційно-регресійних дерев.

Отже, на основі проведеного дослідження варто підкреслити, що послідовність, яка описує поведінку певного ринкового процесу, є фінансовим часовим рядом. Рух цін більшості фінансових інструментів виглядає схожим на різних часових та цінових масштабах. За зовнішнім виглядом графіка неможливо визначити, чи належать дані до тижневих, денних або годинних змін. Це основний аспект фрактальності. Фінансові часові ряди з фрактальною структурою характеризуються нелінійною динамікою, хаосом, нестаціонарністю, невизначеністю і значним шумом.

На сьогодні, у науковій літературі, для прогнозування часових рядів використовується ряд методів. Так А.О. Лузіна та С.П. Кобець описують принцип регресійного аналізу. Автори зазначають, що регресійний аналіз є математичним методом, який базується на безлічі інших методів для моделювання та аналізу декількох змінних, коли основна увага приділяється взаємозв'язку між залежною змінною і однією або декількома незалежними [14]. У сфері статистичного моделювання окреслене поняття являє собою дослідження, що застосовуються з метою оцінки взаємозв'язку між змінними. Однією з головних ознак регресійного аналізу виступає функція регресії, що яка може бути описана за допомогою розподілу ймовірностей.

У 1805 році Лежандром і Гаусом у 1809 році був опублікований метод найменших квадратів, який є однією з перших форм регресії [15]. На початку вчені застосували метод до задачі визначення з астрономічних спостережень орбіти тіл навколо Сонця. Подальший розвиток методу найменших квадратів датується 1821 роком, саме тоді Гаус довів теорему Гауса-Маркова [15].

Метою регресійного аналізу є визначення залежності між вихідною змінною і безліччю регресорів.

Специфіка побудови регресійних моделей для часових рядів полягає у можливій присутності лінійної залежності (високим ступенем кореляції) між пояснюючими змінними.

В. В. Веровчук та О. Б. Вовк зазначають, що якщо залежність є, то обчислювальні процедури будуть володіти мультиколінеарністю і поганою обумовленістю, а оцінки коефіцієнтів моделі — поганими статистичними властивостями [16].

У даній системі для оцінки параметрів регресії використовується метод головних компонент, який може бути віднесений до числа робастних (стійких) методів. У ньому в якості нових змінних використовуються лінійні комбінації пояснюючих змінних, які вибрані так, щоб кореляції між новими змінними були малі або взагалі були відсутні, що дозволяє обійти зазначені труднощі.

Однією з характеристик якості регресійної моделі може служити коефіцієнт детермінації, або його модифікація — скоригований коефіцієнт детермінації, який є несуміщеною оцінкою істинного коефіцієнта детермінації.

Оцінки коефіцієнтів моделі мають (в асимптотиці) розподіл Стьюдента [17]. На цьому заснована перевірка гіпотези про рівність коефіцієнтів нулю, і побудова довірчих меж для коефіцієнтів.

Для перевірки наявності автокореляції залишків зазвичай застосовують коефіцієнт Дарбіна-Уотсона [18] (наявність або відсутність кореляції за часом в помилках системи).

В. С. Степашко розкрив множинну регресійну модель [19]. У загальному випадку модель має вигляд:

$$y_t = \mathrm{C} + \sum_{k=1}^{n} \beta_k \cdot x_{t,k} + e_t \qquad (1.15)$$
$$\mathrm{t} = 1 \ldots \mathrm{T}$$

де T — число спостережень;

$\beta_k, k = 1 \ldots n$ — оцінювані коефіцієнти при пояснюючих змінних;

$x_k, \ k = 1 \ \ldots \ n$ — пояснюючі змінні;

e_t — залишок;

Y_t — змінна яка пояснюється.

Для моделювання часових рядів, ряд науковців [20-22] пропонують використовувати моделі типу AutoRegressive Fractional Integrated Moving Average (ARFIMA) та Fractional Integrated Generalized AutoRegressive Conditional Heteroscedasticity (FIGARCH). Перші – дрібно-інтегровані моделі авторегресії-ковзного середнього є розширенням моделі ARIMA, яка, у свою чергу, є інтегрованою моделлю авторегресії — змінного середнього — модель і методологія аналізу часових рядів. ARIMA є розширенням моделі ARMA для нестаціонарних часових рядів і може бути стаціонаризована шляхом взяття різниць певного порядку від вихідного часового ряду.

Другі — дрібно-інтегровані узагальнені авторегресійні моделі умовної гетероскедастичності, у рамках яких, умовна дисперсія залежить також від минулих значень самої умовної дисперсії [23].

О. Михайловська наводить модель $ARIMA\ (p, d, q)$ означає, що різниці часового ряду порядку підкоряються моделі $ARMA\ (p, q)$ [24].

$$\Delta^d X_t = c + \sum_{i=1}^{p} a_i \Delta^d X_{t-i} + \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t \qquad (1.16)$$

де ε_t — стаціонарний часовий ряд;

c, a_i, b_j — параметри моделі;

Δ^d — оператор різниці часового ряду порядку d (послідовне взяття d раз різниць першого порядку – спочатку від тимчасового ряду, потім від отриманих різниць першого порядку, потім від другого порядку, тощо.)

ARIMA-моделі можна привести до стаціонарності за допомогою послідовного диференціювання спостережень. Аналіз часового ряду починається, як правило, з визначення параметра d процесу $ARIMA\ (p, d, q)$ —

порядку інтегрованості часового ряду [25]. Зазвичай розглядають два окремих випадки: 0 і 1. У випадку коли, $d = 0$ часовий ряд має коротку пам'ять, при $d = 1$ кожний окремий вплив на часовий ряд триває нескінченно довго, що не враховує проміжні випадки довгої пам'яті та є обмеженим і приватним. У зв’язку з цим у 1981 році Хоскинг [26] узагальнив модель ARIMA для дрібного диференціювання, у ввів нову модель ARFIMA, яка дозволяє моделювати ефекти довгої і короткої пам'яті.

Моделі ARFIMA передбачають використання дробових порядків різниць, оскільки теоретично порядок інтегрованості d тимчасового ряду може бути не цілою величиною, а дробовою. Для розуміння сутності дрібного інтегрування необхідно розглянути розкладання оператора взяття d-ої різниці у статечний ряд за ступенями лагового оператора для дрібних d (розкладання в ряд Тейлора [27]):

$$\Delta^d = (1-L)^d = \sum_{k=0}^{\infty}(-1)^k \prod_{j=0}^{k-1}(d-j) \Big/ k! \, L^k \qquad (1.17)$$

де L — лагів оператор;

k — порядок інтегрування.

Для моделювання мінливості волатильності часових рядів в часі зазвичай вдаються до використання моделей сімейства GARCH — аналогічно моделі ARFIMA для ARCH, узагальнюючи її на дробовий випадок, вводиться модель із частковим порядком інтегрування FIGARCH для GARCH.

Авторегресійні інтегровані моделі середнього ковзаючого (ARIMAX) розширюють моделі ARIMA шляхом включення екзогенних змінних X. Модель ARIMAX (p, d, q) є моделлю для деяких даних часових рядів Y_t та екзогенних даних X_t, де p — це число авторегресійних лагів, d — ступінь диференціювання та q — кількість переміщень середніх лагів:

$$\Delta^D Y_t = \sum_{i=1}^{p} \phi_i \Delta^D Y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{m=1}^{M} \beta_m Xm, t + \epsilon_t \qquad (1.18)$$

Моделі з розподіленими лагами розкривають В. М. Горбачук. Моделі з розподіленими лагами це моделі, що містять в якості лагових змін тільки незалежні (пояснюючі) змінні [28,29]. А. І. Яковлев та О. П. Косенко, зазначають, що дані моделі широко використовуються в економетричному аналізі [30], так як в багатьох випадках вплив деяких економічних факторів на інші здійснюється не миттєво, а з деяким тимчасовим запізненням — лагом. Метод розподілених лагів дозволяє досліджувати влив такого роду.

Нехай досліджується показник Y. Його значення в поточний момент часу — y_t; значення Y в наступні моменти часу — $y_{t+1}, y_{t+2}, \ldots, y_{t+q}$; значення Y в попередні моменти часу — $y_{t-1}, y_{t-2}, \ldots, y_{t-q}$.

У разі відсутності авторегресійних членів модель розподілених лагів – це модель виду:

$$y_t = a_0 + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_q x_{t-q} + \varepsilon_t \qquad (1.19)$$

де x_t — лаговані значення екзогенних змінних;

β_q — короткостроковий мультиплікатор. Характеризує зміну середнього значення Y в момент часу t під впливом одиничної зміни змінної X в момент часу $t-q$.

Сума всіх коефіцієнтів при екзогенних змінних — довгостроковий мультиплікатор. Він характеризує зміну Y під впливом одиничної зміни змінної X в кожному з розглянутих часових періодів.

Будь-яку суму коефіцієнтів $\sum_{j=0}^{p} \beta_j \, (p < q)$ називають проміжним мультиплікатором.

Для вимірювання швидкості реакції Y на зміну X розглядається величина середнього лага:

$$\sum_{j=0}^{q} j \cdot w_j \qquad (1.20)$$

де $w_j = {\beta_j}/{\sum_{j=0}^{q} \beta_j}$ — це внесок окремого лага або розподіл лага.

Малі значення середнього лага відповідають швидкій реакції Y на зміну X, великим значенням середнього лага відповідає уповільнена реакція.

В. М. Горбачук описує моделі експоненціального згладжування. Науковець зазначає , що у методі простого експоненціального згладжування застосовується зважене (експоненціально) ковзне усереднення всіх даних попередніх спостережень [31, 32]. Ця модель найчастіше застосовується до даних, в яких необхідно оцінити наявність залежності між аналізованими показниками (тренда) або залежність аналізованих даних. Метою експоненціального згладжування є оцінка поточного стану, результати якого визначать всі наступні прогнози.

Метод експоненціального згладжування був незалежно відкритий Брауном у 1959 році [33] і Хольтом у 1957 році [34]. Експоненціальне згладжування, як і метод ковзних середніх, для прогнозу використовує минулі значення часового ряду.

Експоненціальне згладжування передбачає постійне оновлення моделі за рахунок найбільш свіжих даних. Цей метод ґрунтується на усередненні (згладжування) часових рядів минулих спостережень у низхідному (експоненціально) напрямку. Тобто більш пізнім подіям присвоюється більша вага. Вага присвоюється наступним чином: для останнього спостереження вагою буде величина α, для передостаннього — $(1-\alpha)$, для того, яке було перед ним, — $(1-\alpha)^2$, тощо.

О. Є. Полозюк наголошує, звичайне експоненціальне згладжування використовується, коли в даних відсутній тренд або сезонність [35]. Прогноз у цьому випадку є середньозваженою величиною всіх наявних попередніх значень ряду, причому ваги зменшуються в геометричній прогресії з часом, коли ми рухаємося назад (у минуле). Тому (на відміну від методу ковзного середнього) не існує точки, в якій ваги обриваються, тобто прямують до нуля. Прагматично

зрозумілу модель простого експоненціального згладжування можна записати наступним чином:

$$F_{t+1} = \alpha \cdot X_t + (1 - \alpha) \cdot F_t \quad (1.21)$$

де F_{t+1} — прогнозне (згладжене) значення на період $t + 1$;

α — параметр згладжування (smoothing parameter), або, інакше, фактор демпфірування (damping factor) – $(0 < \alpha < 1)$;

X_t — поточне значення часового ряду;

F_t — прогнозне значення на період t, або згладжене значення для періоду $t - 1$.

В. Михайленко підкреслює, що при рекурсивному застосуванні цієї формули кожне нове згладжене значення (яке також є прогнозом) розраховується як середньозважене значення поточних спостережень і згладженого ряду [36]. Звичайно, результат згладжування залежить від параметра адаптації альфа. Його можна інтерпретувати як коефіцієнт дисконтування, що характеризує ступінь знецінення даних за одиницю часу. Крім того, вплив даних на прогноз зменшується експоненціально зі збільшенням "віку" даних [37].

Слід зауважити, що значення згладжуючого параметра не може дорівнювати 0 або 1, так як в цьому випадку сама ідея експоненціального згладжування відкидається. Так, якщо альфа дорівнює 1, то прогнозне значення F_{t+1} збігається з поточним значенням ряду X_t, при цьому експоненціальна модель прагне до найпростішої «наївної» моделі, тобто в цьому випадку прогнозування є абсолютно тривіальним процесом. Якщо альфа дорівнює 0, то початкове прогнозне значення F_0 (initial value) одночасно буде прогнозом для всіх наступних моментів ряду, тобто прогноз в цьому випадку буде виглядати у вигляді звичайної горизонтальної лінії.

Проте, варіанти згладжуючого параметра, близькі до 1 або 0 є доволі поширеними. Так, якщо альфа близько до 1, то попередні спостереження

часового ряду практично повністю ігноруються. У разі якщо альфа близько до 0, то ігноруються вже поточні спостереження. Значення альфа між 0 і 1 дають проміжні результати. На думку ряду авторів [38, 39], оптимальне значення альфа знаходиться в межах від 0,05 до 0,30.

В цілому краще оцінювати оптимальне альфа за вихідними даними (за допомогою пошуку по сітці), а не використовувати штучні рекомендації. Проте, в разі якщо значення альфа, що перевищує 0,3, мінімізує ряд спеціальних критеріїв, то це вказує на те, що інша техніка прогнозування (із застосуванням тренда або сезонності) здатна забезпечити ще більш точні результати. Для знаходження оптимального значення альфа (тобто мінімізації спеціальних критеріїв) використовується квазіньютонівський алгоритм максимізації правдоподібності (вірогідності), який ефективніше звичайного перебору на сітці.

Відзначимо, що при використанні методів експоненціального згладжування важливим питанням завжди є визначення початкових умов (початкового прогнозного значення F_0). Процес вибору початкового значення згладженого ряду називається ініціалізацією (initializing), або, інакше, «розігрівом» («warming up») моделі. Справа в тому, що початкове значення згладженого процесу може істотно вплинути на прогноз для подальших спостережень. З іншого боку, вплив вибору зменшується з довжиною ряду і стає некритичним при дуже великій кількості спостережень. Браун вперше запропонував використовувати в якості стартового значення середнє динамічного ряду. Інші автори пропонують використовувати в якості початкового прогнозу перше фактичне значення часового ряду.

У середині минулого століття Хольт запропонував розширити модель простого експоненціального згладжування за рахунок включення в неї фактора росту (growth factor), або інакше тренда (trend factor).

Даний метод дозволяє врахувати присутність в даних лінійного тренду. Пізніше були запропоновані інші види трендів: експонентний, демпфірований, тощо.

Вінтерс запропонував удосконалити модель Хольта з точки зору

можливості опису впливу сезонних факторів.

Зокрема, він далі розширив модель Хольта за рахунок включення в неї додаткового рівняння, що описує поведінку сезонної компоненти (складової).

Про фрактальну структуру знакових комбінаторних просторів детально описала Н. К. Тимофієва. Існує кілька напрямків представлення даних у сучасних структурних моделях, найпопулярнішими з яких є моделі на основі штучних нейронних мереж (ШНМ).

Поняття штучної нейронної мережі (ШНМ) вперше було описано в 1943 році математиком Волтером Фітцем і нейропсихологом Уорреном Мак-Каллоком. ШНМ — це обчислювальна модель з великою кількістю паралельних процесів, об'єднаних великою кількістю зв'язків.

Структура тришарової нейронної мережі виглядає наступним чином (рис. 1.2).

На рис. 1.2 показано структуру тришарової нейронної мережі прямого поширення. У цій мережі сигнали проходять через шари в одному напрямку, від входу до виходу. Нейрони першого вхідного шару поширюють сигнал без змін до нейронів другого прихованого шару, де сигнал обробляється для формування вихідного нейрона.

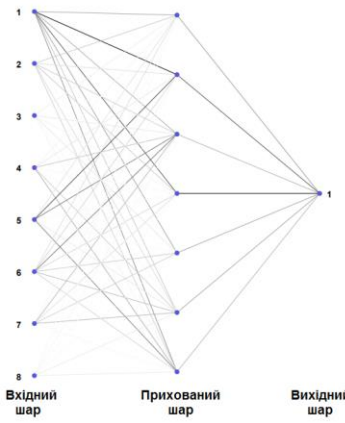


Рисунок 1.2 — Структура тришарової нейронної мережі

Якщо ви використовуєте більше шарів, логіка презентації буде схожою. Всі знання зберігаються у вигляді ваг.

Наступний крок — навчання нейронної мережі. О. Лопатько розповідає,

що навчання нейронної мережі — це процес, заснований на прикладах подій, для яких відомо, як встановити ваги та мінімізувати функцію втрат. Навчання є найважливішим процесом у використанні нейронної мережі, оскільки саме від навчання найбільше залежить кінцевий результат.

Перш ніж використовувати нейронну мережу для запланованої задачі, потрібно знайти найкращий набір ваг. Козирський писав, що нам потрібна певна міра, яка показує, який набір ваг найкраще працює на функції втрат.

Середньоквадратична помилка зазвичай використовується як опція втрат для нейронних мереж.

Квадратична функція втрат:

$$\chi^2 = \sum_j w_j x_j^2 = \sum_j \frac{1}{2} \sum_i \left(T_{ji} - O_{ji}\right)^2 \quad (1.22)$$

де T_{ji} — вихід мережі номер j;

O_{ji} — вірна відповідь мережі номер j.

Однак середньоквадратична помилка не є єдиним варіантом, і існує тенденція до використання середньої перехресної ентропії всіх навчальних вибірок.

Функція ентропії:

$$E_D = \sum_j w_j E_D^j = \sum_j \sum_i - log\left(\frac{1}{2}\left(1 + T_{ji} O_{ji}\right)\right) \quad (1.23)$$

Таким чином, функція ентропії дозволяє формально визначити завдання навчання класифікатора. Процедура навчання нейронної мережі полягає в мінімізації функції втрат у ваговому просторі.

У найпростішому випадку класифікація може бути представлена подією, яка вимагає окремого вихідного вузла, і якщо вихід більший за вибране значення, то це клас 1 (сигнал); якщо він більший за значення, то це клас 1 (сигнал); інакше

— клас 2 (фон).

Н. Савка довів, що визначення значення умовної щільності ймовірності відбувається, коли у вихідному шарі є багато вузлів, кожен з яких відповідає на питання, чи є ймовірність більшою за порогове значення. Вихідним вектором є інтегральна щільність ймовірності.

Таким чином, штучні нейронні мережі можуть бути використані для моделювання нелінійного зв'язку між майбутнім значенням часового ряду та фактичним значенням, а також значенням зовнішніх факторів.

Однак, для часових рядів з мінливою фрактальною структурою, ШНМ, яка була навчена для однієї частини часового ряду, в більшості випадків не здатна забезпечити задовільні результати прогнозування для іншої частини часового ряду, де характеристики часового ряду змінилися. Тому нейронну мережу необхідно перенавчати, якщо структура часового ряду змінюється.

Іншим основним недоліком ШНМ є спосіб інтерпретації результатів, вони незрозумілі для людини, але використання нечіткої логіки та нечітких моделей часових рядів позбавлене цього основного недоліку.

Моделі нечітких часових рядів детально описав О. С. Якушенко. Науковець зазначає, що на відміну від інших методів прогнозування, в умовах, коли більшості реальних подій властивий високий ступінь невизначеності, методи прогнозування на основі моделей нечітких часових рідів дозволять здійснити найбільш адекватну оцінку майбутніх змін в соціально-економічних системах.

Визначення моделей нечітких часових рядів першими дали Song і Chissom. У 2006 році S. Chen в розглянув стаціонарну модель часового ряду, засновану на використанні нечітких множин. Надалі, дана модель була поліпшена за допомогою використання генетичних алгоритмів для «тонкої» настройки параметрів нечітких моделей прогнозування.

У був запропонований метод прогнозування на основі часових рядів першого порядку, який показав найкращі результати у порівнянні з вищезгаданої розробкою S. Chen. Було виявлено, що кількість нечітких множин, використаних

для опису часового ряду, значно впливає на точність прогнозу.

С. О. Субботін зазначає, що під нечітким часовим рядом (НЧР) розуміється упорядкована послідовність даних про досліджуваному процесі, для якого характерна мінливість у часі, якщо значення обраного параметра в момент часу виражаються через нечітку мітку.

У контексті цієї лінгвістичної моделі ключовим показником є область моментів часу.

На основі теореми FAT (Fuzzy Approximation Theorem), яка доводить, що будь-яка математична система може бути апроксимована системою, заснованою на нечіткій логіці, можна стверджувати, що будь-яке відношення може бути описане з нескінченною точністю за допомогою формалізованих висловлювань природної мови «ЯКЩО-ТО».

Перехід до нечіткої моделі часового ряду в контексті лінгвістичного представлення усуває проблему штучних нейронних мереж щодо форми інтерпретації отриманих результатів.

На основі вищевикладеного, варто зазначити, що масштабність наявності викладених у сучасній науковій літературі моделей аналізу фінансових часових рядів дозволяє варіювати особливостями часового ряду та підбирати модель прогнозування, найбільш підходящу до конкретної ділянки часового ряду, що особливо актуально для часових рядів з фрактальною структурою.

1.3 Огляд та аналіз аналогів

В даний час для прогнозування економічних показників найчастіше застосовуються статистичні пакети програм, математичні пакети програм, аналітичні платформи, нейропакети, програми для автоматизації бізнесу та табличний процесор MS Excel. Ринок цього програмного забезпечення є досить динамічним: деякі програмні продукти з різних причин припиняють розвивати, і вони залишають ринок; натомість з'являються нові продукти; деякі статистичні програмні продукти припинили самостійне існування та складовими частинами увійшли до програм для автоматизації бізнесу або аналітичних платформ.

Аналіз останніх оглядів статистичного програмного забезпечення, що використовується показав наступне:

1) в даний час у якості комп'ютерного інструменту для вирішення економетричних завдань найчастіше використовується табличний процесор MS Excel;

2) серед статистичних програмних продуктів у якості альтернативи розглядаються статистичні пакети загального призначення (STATISTICA, SPSS), спеціалізовані програми для аналізу часових рядів;

3) у якості компромісу за критеріями «можливості програмного продукту» і «вартість» в даний час рекомендується використовувати програми для аналізу часових рядів, а також табличний процесор MS Excel;

4) у зазначених оглядах відсутнє порівняння можливостей статистичних програм із можливостями аналітичних платформ.

Загальні відомості про порівнюване програмне забезпечення, що має інструменти для прогнозування, представлено в табл. 1.2.

Таблиця 1.2 — Програмне забезпечення для прогнозування

Програмний продукт	Виробник	Рік
Статистичні пакети загального призначення		
IBM SPSS Statistics	IBM (США) (до 2009 г. SPSS Inc.)	1967
STATISTICA	Statsoft Inc. (США)	1985
Програми для автоматизації бізнесу		
Forecast PRO	Business Forecasting System Inc. (США)	1986
Нейропакет		
NeuroShell 2	Ward Systems Croup Inc. (США)	1988
Табличний процесор		
Microsoft Excel	Microsoft Corporation (США)	1985

Відомості про позиціонування програмних продуктів, що порівнюються, наведені в табл. 1.3.

Для прогнозування економічних показників, переважно, використовуються методи регресійного аналізу, аналізу часових рядів, і навіть штучні нейронні мережі. Порівняння програмного забезпечення, що

розглядається, за реалізованими в них методами економетрики та прогнозування наведено в табл. 1.4 — 1.6.

Таблиця 1.3 — Позиціювання програмного забезпечення для прогнозування

Програмний продукт	Позиціонування
IBM SPSS Statistics	Повнофункціональна статистична система, призначена для вирішення дослідницьких та бізнес-завдань за допомогою аналізу даних
STATISTICA	Потужні та зручні у використанні інструменти для статистичного та графічного аналізу, прогнозування, data mining, створення власних додатків користувача, інтеграції, спільної роботи, web-доступу.
Forecast PRO	Програмне забезпечення для розрахунку прогнозів та розробки планів. Механізм прогнозування продукту Forecast PRO Expert Selection у 2020 р. показав найкращі результати на незалежному конкурсі з прогнозування, в якому порівнювали 24 програмні продукти при прогнозуванні 3003 часових рядів попиту у різних галузях.
NeuroShell 2	Універсальний пакет, призначений для нейромережевого аналізу даних.
Microsoft Excel	Потужний інструмент, який використовується для створення та форматування електронних таблиць, аналізу даних та обміну інформацією для прийняття управлінських рішень.

Таблиця 1.4 — Методи регресійного аналізу у програмному забезпеченні

Метод регресійного аналізу	IBM SPSS Statistics	STATISTICA	Forecast PRO	NeuroShell 2	Microsoft Excel
1. Лінійний регресійний аналіз					
1.1. Класична лінійна модель множинної регре (КЛММР)	+	+	-	-	+
1.2. Гребенева регресія (рідж-регресія) (КЛММР в умовах мультиколінеарності)	+	+	-	-	-

Продовження таблиці 1.4

1.3. Ортогоналізація пояснюючих змінних методом головних компонентів (КЛММР в умовах мультиколінеарності)	-	-	-	-	-
1.4. Покрокова регресія (КЛММР в умовах мультиколінеарності)	+	+	-	-	-
1.5. Узагальнена лінійна модель множинної регресії					
— лінійна модель множинної регресії з гетероскедастичними залишками (узагальнений МНК або його окремий випадок — зважений МНК)	+	+	-	-	-
— лінійна модель множинної регресії з автокорельованими залишками (узагальнений МНК)	+	-	-	-	-
1.6. Лінійна модель множинної регресії зі стохастичними змінними, що пояснюють (метод інструментальних змінних)	+	-	-	-	-
1.7. Лінійна модель множинної регресії зі змінною структурою (фіктивні змінні)	+	+	-	-	-
2. Нелінійний регресійний аналіз					
	+	+	-	-	+
3. Системи одночасних рівнянь					
— двокроковий МНК	+	-	-	-	-

У сучасному прогнозуванні спостерігається тенденція використання комбінованих моделей прогнозування, у яких відбувається поєднання прогнозів індивідуальних моделей прогнозування. Вирізняють два класи таких моделей прогнозування [3, с. 126]:

1) селективна модель, в якій помилки прогнозів використовуються для формування адаптивного критерію, що дозволяє вибирати з набору моделей в даний час найкращу модель;

2) гібридна модель, в якій проводиться об'єднання прогнозів, отриманих за моделями, що входять до набору, з вагою залежно від точності прогнозів.

Відомості про наявність комбінованих моделей прогнозування в програмному забезпеченні подано в табл. 1.7. Гібридна модель на основі усереднення прогнозів використовується у програмі STATISTICA при поєднанні прогнозів множини (ансамблю) нейронних мереж.

Таблиця 1.5 — Методи аналізу часових рядів у програмному забезпеченні

Метод аналізу часових рядів	IBM SPSS Statistics	STATISTICA	Forecast PRO	NeuroShell 2	Microsoft Excel
1. Спектральний аналіз	+	+	-	-	+
2. Вейвлет-аналіз	-	-	-	-	-
3. Трендові моделі					
3.1. Аналітичні методи					
— криві зростання	+	+	+	-	+
— Грей-метод	-	-	-	-	-
— фільтр Ходрика-Прескотта	-	-	-	-	-
— LRX-фільтр	-	-	-	-	-
— фільтр Бакстера-Кінга	-	-	-	-	-
3.2. Алгоритмічні методи					
— методи ковзного середнього	+	+	+	-	+
— методи експоненційного ковзного середнього	-	+	-	-	+
— метод ковзної медіани	+	+	-	-	-
3.3. Ковзна регресія	-	-	-	-	-
4. Моделі стаціонарних часових рядів (для моделювання регресійних залишків)					
— моделі авторегресії (AR)	+	+	-	-	-
— моделі ковзного середнього (MA)	+	+	-	-	-
— модель авторегресії — ковзного середнього (ARMA)	+	+	-	-	-
— авторегресійна умовно гетероскедастична модель (ARCH-модель)	-	-	-	-	-
- узагальнена авторегресійна умовно гетероскедастична модель (GARCH-модель)	-	-	-	-	-
5. Моделі нестаціонарних однорідних тимчасових рядів					
5.1. Модель авторегресії — проінтегрованого ковзного середнього (ARIMA)	+	+	+	-	-
5.2. ARIMA з інтервенціями	-	+	-	-	-
5.3. Сезонна модель ARIMA	+	+	-	-	-
5.4. Сезонна декомпозиція					
— метод Census 1	+	+	+	-	-
— метод X11 (Census 2)	-	+	-	-	-
— модель TRAMO	-	-	-	-	-
— Модель SEATS	-	-	-	-	-

Продовження таблиці 1.5

5.5. Регресійні моделі з розподіленими лагами					
— звичайні лаги	-	+	-	-	-
— поліноміальна лагова структура Алмон	-	+	-	-	-
— геометрична лагова структура Койка	-	-	-	-	-
6. Адаптивні моделі прогнозування на основі експоненційного згладжування					
— модель Брауна нульового порядку	+	+	+	+	-
— модель Хольта	+	+	+	+	-
— модель Хольта-Уінтерса	+	+	+	+	-
— Інші адаптивні моделі	+	+	-	-	-
7. Моделі прогнозування рідкісного продажу (модель Кростона, метод Bootstrapping або інші)	-	-	-	+	-

Вибір програмного забезпечення для прогнозування економічних показників є складним багатокритеріальним завданням. Адекватний вибір методів прогнозування (і, відповідно, програмного забезпечення) для організацій залежить багатьох чинників:

— цілей прогнозу;

— горизонту прогнозування;

— кількості наявних ретроспективних даних;

— особливостей даних.

Таблиця 1.6 — Нестатистичні методи інтелектуального аналізу даних у програмному забезпеченні

Нестатистичний метод інтелектуального аналізу даних	IBM SPSS Statistics	STATISTICA	Forecast PRO	NeuroShell 2	Microsoft
1. Штучні нейронні мережі					
— багатошаровий персептрон (MLP)	+	+	-	-	-
— мережі типу радіальної базової функції (RBF)	+	+	-	-	-
— узагальнено-регресійна нейронна мережа (GRNN)	-	+	-	-	-
— Лінійна мережа	-	+	-	-	-

Продовження таблиці 1.6

2. Еволюційне програмування					
— генетичні алгоритми	-	-	-	-	-
— метод групового обліку аргументів	-	-	-	-	-

Таблиця 1.7 — Комбіновані моделі прогнозування у програмному забезпеченні

Вид комбінованої моделі	IBM SPSS Statistics	STATISTICA	Forecast PRO	NeuroShell 2	Microsoft Excel
Селективна модель	+	+	+	-	-
Гібридна модель	-	+	-	-	-

Умови застосування моделей прогнозування залежно від кількості ретроспективних даних, узагальнені за інформацією різних джерел, полягають у наступному [6, 28]:

— моделі на основі експоненційного згладжування дозволяють обчислювати прогнозні значення за наявності 1 або 2 спостережень;

— для достовірної ідентифікації регресійних моделей число спостережень має у 6-7 разів перевищувати кількість оцінюваних параметрів при незалежних змінних;

— авторегресійні моделі призначені для прогнозування часових рядів, що мають 50 і більше рівнів;

— на навчання нейронної мережі потрібні сотні спостережень.

2 МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ЯК ІНСТРУМЕНТ ДЛЯ РОЗРОБКИ ВЕБ-ДОДАТКУ ДЛЯ ПРОГНОЗУВАННЯ В ДАТА-МАЙНІНГУ

2.1 Аналіз підходів до прогнозування в дата-майнінгу

Починаючи з середини XX століття використання нелінійних математичних методів знаходить все ширше застосування в економіці. Цей підхід особливо затребуваний у сфері аналізу динаміки ціноутворення фондових ринків.

З наукової точки зору гіпотеза фрактального ринку є альтернативною парадигмою, що описує взаємодію учасників ринку та ціноутворення активів як результат цієї взаємодії. Причини гіпотези фрактального ринку є більш наближеними до реального стану справ, ніж традиційні положення гіпотези ефективного ринку.

З практичної погляду дослідження фрактальних властивостей ціноутворення активів дозволяє отримувати більш точну оцінку ринкових ризиків. За рахунок цього можуть бути отримані рекомендації, необхідні для роботи приватних і інституційних інвесторів.

У той же час залишається відкритим питання про практичну застосовність у їхньому нинішньому вигляді теоретичних методів прогнозування часових рядів у рамках гіпотези фрактального ринку.

Аналіз методів прогнозування, викладених у сучасній науковій літературі, фінансових часових рядів дозволяє стверджувати про відсутність методу, який би відрізнявся високою точністю прогнозу, проте їх (методів) кількість дозволяє варіювати особливостями часового ряду та підбирати модель прогнозування, найбільш підходящу до конкретної ділянки часового ряду, що особливо актуально для часових рядів з фрактальною структурою.

У рамках даного дослідження запропоновано методику розробки системи для прогнозування часових рядів методом фрактального аналізу, який передбачає масштабування часового ряду та виділення корозмірності.

Загальний метод аналізу та прогнозування фінансових часових рядів можна описати в декілька кроків, які відбуваються в шість етапів, як показано на рисунку 2.1.

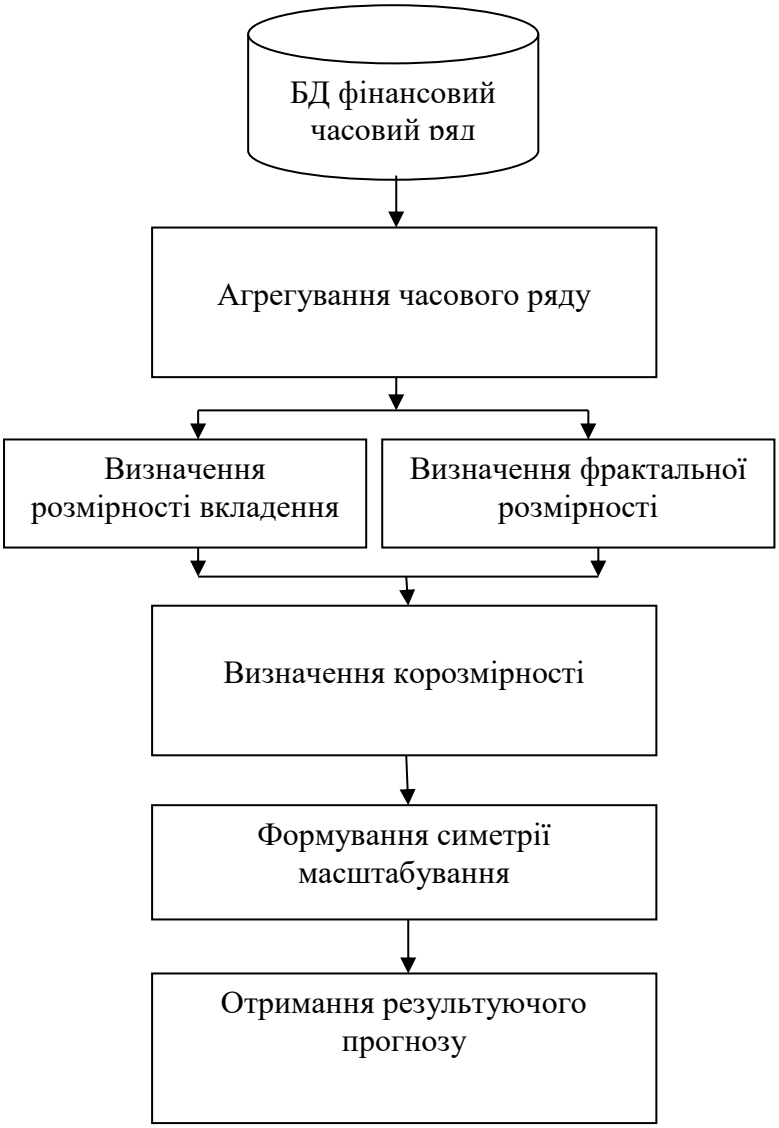


Рисунок 2.1 — Структурна схема реалізації загального способу аналізу та прогнозування фінансового часового ряду

Перший етап — використання статистичних методів та візуалізації для попереднього аналізу фінансових часових рядів з метою виявлення наявності або відсутності сильних хвостів, трендів, циклічних або сезонних компонентів тощо.

Другий етап — агрегування фінансових часових рядів з метою отримання агрегованих економічних показників шляхом їх групування. Агреговані

показники - це узагальнені, синтетичні показники, які об'єднуються в загальний індикатор.

Третій етап — фрактальний аналіз фінансових часових рядів з метою визначення характеристик і тенденцій, таких як стійкість тренду або, навпаки, хаотичність, персистентність або антиперсистентність. Обчислювальна частина фрактального аналізу базується на визначенні кореляції. І. С. Сиротенко зазначає, що оцінки, отримані наприкінці цього етапу, мають числовий характер.

Четвертий етап — виявляється симетрія масштабів грошових часових рядів. Виявлення симетрій обертання, зсуву та масштабування в числових послідовностях фінансових послідовностей для отримання числових показників трансформацій, а також для оцінки ступеня порушення симетрії.

П’ятий етап — формується прогноз для досліджуваної фінансової послідовності шляхом застосування розрахунків на основі розробленого алгоритму.

Шостий етап — оцінюється похибка прогнозу для фінансової історії, що розглядається.

Запропоновано модель часових рядів для фінансових питань. В якості базової моделі для аналізу та прогнозування обрано дані про фінансові результати діяльності банків, які мають фрактальні властивості на короткому горизонті, але асимптотично подібні до рівноважного стану. Визначено метод аналізу та прогнозування фінансових часових рядів з використанням лінгвістичного моделювання, що передбачає масштабування часових рядів та виділення кореляційних зв'язків. Фінансове агрегування здійснюється з метою консолідації економічних показників в єдину групу. Формування симетрії масштабування фінансових часових рядів здійснюється з метою отримання числових значень перетворень, а також оцінки ступеня порушення симетрії. Характеристики запропонованого методу дозволяють аналізувати та прогнозувати фінансові часові ряди в умовах, коли існуючі методи, що використовують окремі методи математичної статистики або інформаційні технології, є неефективними.

В основу розробки системи для прогнозування часових рядів методом фрактального аналізу покладено Model-view-controller.

Model-view-controller — патерн використання декількох шаблонів проектування, де модель даних додатку, користувацький інтерфейс і користувацький досвід розділені на три окремі компоненти так, що зміна одного компонента має мінімальний вплив на інші. Цей патерн проектування часто використовується для створення архітектурної основи при переході від теорії до застосування в домені.

Концепція MVC була описана Трігве Реенскаугом у 1979 році, який працював над мовою програмування Smalltalk в Xerox PARC в той час. Початкова реалізація описана в статті "Програмування додатків на Smalltalk-80: Як використовувати Model-View-Controller". Пізніше Джим Альтхофф та його команда розробників реалізували версію MVC для бібліотеки класів Smalltalk-80.

В оригінальній концепції була описана сама ідея та роль кожного з елементів:

— моделі;

— уявлення;

— контролера.

Але зв'язок між ними було описано без конкретизації. Крім того, розрізняли дві основні модифікації: активна та пасивна модель.

Пасивна модель — модель не може ніяк впливати на подання або контролер, але використовується як джерело даних для відображення. Будь-які зміни в моделі відстежуються контролером і відповідають за перемальовування представлення, якщо це необхідно. Цей тип моделі частіше використовується в структурному програмуванні. Оскільки в цьому випадку модель просто представляє структуру даних, без методів її обробки.

Активна модель — модель сповіщає подання про те, що в ній відбулися зміни, а подання, зацікавлені в отриманні сповіщень, підписуються на ці повідомлення. Це дозволяє моделі залишатися незалежною як від контролера,

так і від представлень. Класичною реалізацією концепції MVC можна вважати версію з активною моделлю.

З розвитком концепції об'єктно-орієнтованого програмування та патернів проектування було створено кілька модифікацій концепції MVC, які можуть відрізнятися від оригіналу при реалізації різними авторами. Наприклад, Еріан Верме описав узагальнений приклад MVC у 2004 році.

Основна мета реалізації цієї концепції — відокремити бізнес-логіку (модель) від її візуалізації (представлення). Таке відокремлення збільшує можливість повторного використання. Найбільш корисне застосування цієї концепції, коли користувачеві потрібно бачити одні й ті ж дані одночасно в різних контекстах та/або з різних точок зору.

До моделі можна додати кілька подань, не впливаючи на її реалізацію. Наприклад, деякі дані можуть бути представлені одночасно в таких форматах: таблиця, гістограма і кругова діаграма.

Не торкаючись реалізації видів, можна змінити реакцію дії користувача (натискання мишею на кнопці, введення даних), цього достатньо використовувати інший контролер.

Багато розробників спеціалізуються лише в одній області, де вони або розробляють графічний інтерфейс користувача, або розробляють бізнес-логіку. Тому цілком можливо, що програмісти, які беруть участь у розробці бізнес-логіки (моделі), не знають, яке представлення використовується.

MVC дозволяє розділити обробку даних, презентації та дій користувача на три окремі компоненти (рис. 2.2).

Модель надає знання: дані та методи роботи з цими даними. реагує на запити, змінюючи свій стан. Вона не містить інформації про те, як ці знання можна візуалізувати.

Відповідає за відображення інформації (візуалізацію). Часто у якості уявлення виступає форма (вікно) із графічними елементами.

Контролер забезпечує зв'язок між користувачем і системою: перевіряє вхідні дані користувача і використовує модель і подання для реалізації необхідної реакції.

Варто зазначити, що і представлення, і контролер є залежними від моделі. Однак, ні представлення, ні контролер не є залежними від моделі. Таким чином досягається мета розділення, що дозволяє створювати модель, незалежну від візуального представлення, а також створювати декілька різних представлень однієї і тієї ж моделі.

На основі вище викладеного варто зазначити, що запропонований спосіб аналізу та прогнозування фінансового часового ряду дозволить виконати масштабування часового ряду та виділення корозмірності, що у свою чергу допоможе здійснити прогнозування часового ряду.

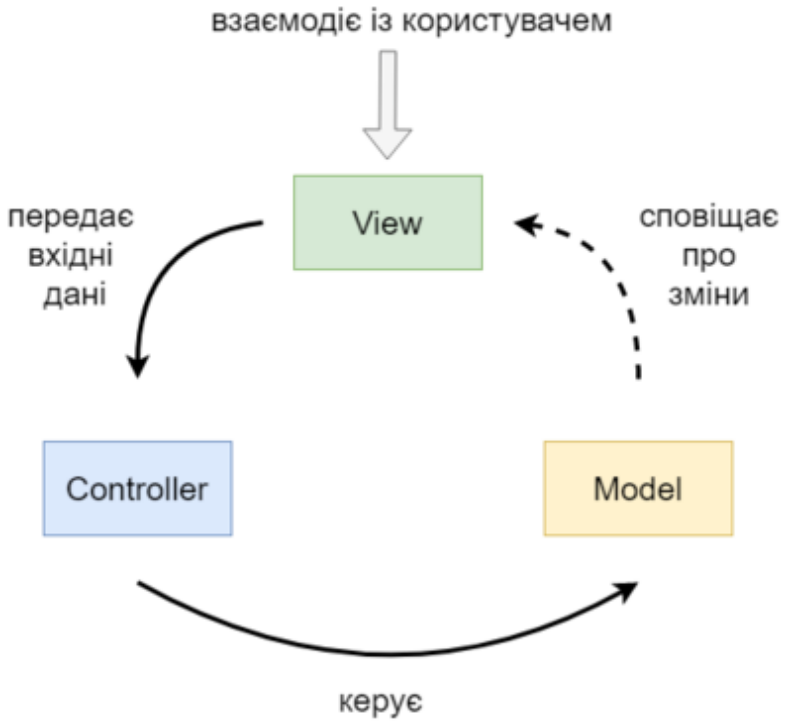


Рисунок 2.2 — Схема паттерна MVC

2.2 Дослідження математичних моделей для прогнозування в дата-майнінгу

Створення модифікованих моделей, що поєднують інтелектуальні та статистичні методи аналізу та прогнозування фінансових часових рядів, є

актуальним завданням сьогодення, що зумовлено виявленими недоліками існуючих методів у галузі сучасної економіки.

Інтеграція нечіткої логіки та детермінованої теорії хаосу в рамках лінгвістичного моделювання сприяє розвитку інноваційних підходів до прогнозування фінансових часових рядів з фрактальною структурою.

Фінансові часові ряди мають високу нелінійну мінливість, особливо на високих частотах, і часто проявляють фрактальні властивості. Коли фрактальна розмірність часового ряду дорівнює нулю, з нею пов'язані дві характеристики:

— фрактальні процеси демонструють гетерогенність — високу ймовірність екстремальних або віддалених флуктуацій, зазвичай з нерегулярними інтервалами;

— фрактальні процеси також виявляють симетрію експоненціальність пропорційність зв'язку між коливаннями на різних відстанях.

Фракталізація на фінансових ринках не означає хаотичну поведінку, подібну до випадковості, породженої невеликою кількістю детермінованих рівнянь. Фрактальність у великомасштабних багатовимірних системах, таких як фінансові ринки, є стохастичною. Цей тип дробовості зазвичай є результатом мультиплікативної взаємодії двох або більше стохастичних процесів.

Значна частина сучасної літератури про нелінійну волатильність [38] на фінансових ринках ґрунтується на дифузії волатильності під впливом декількох випадкових факторів. Дифузія волатильності з мультиплікативною взаємодією між стохастичними факторами має тенденцію генерувати фракціоналізацію.

На великих відстанях фінансові тертя можна моделювати структурними рівняннями, які, як правило, не є фрактальними. У цьому сенсі економічні часові ряди доцільно характеризувати як такі, що мають фрактальні властивості на короткому горизонті, але асимптотично подібні до рівноваги. У деяких випадках структурні рівняння, що лежать в основі, також індукують фрактальність [9].

Найбільш прямим прикладом є обмінний курс, на який, як очікується, впливають відмінності в реальних доходах. Нехай X_t — валютний курс, I_t —

процентна ставка, e - індекс очікувань, ε_t — залишок монетарного часового ряду та F — іноземна валюта. Структурне рівняння має такий вигляд:

$$X_t = \omega_0 + \omega_1 X_{t-1} + \omega_2[(I_t - \pi_1^e) - (I_{Fl} - \pi_{Ft}^e)] + \varepsilon_l \qquad (2.1)$$

Як зазначено, використання коефіцієнтів стохастичного процесу може призвести до нелінійної мінливості. Відповідно, вираз $[(I_t - \pi_1^e) - (I_{Fl} - \pi_{Ft}^e)]$ вимагає фрактальної поведінки. Крім того, номінальна і реальна процентні ставки також можуть бути фрактальними, так що різниця в реальних процентних ставках сама по собі є різницею в двох незалежних фрактальних процесах.

У дослідженні запропоновано алгоритм прогнозування, що базується на двох методах:

— застосування моделі переходу станів для умовного прогнозування. ймовірність настання екстремальних подій;

— побудови моделі симетрії на короткому часовому горизонті.

Методологія моделювання часових рядів базується на розкладанні часового ряду на складові та моделюванні значень кожної складової окремо. У цьому дослідженні пропонується застосувати агрегування до кожного ряду, тобто розбити ряд на короткі інтервали, значення яких є певним чином схожими. З математичної точки зору, агрегування розглядається як перетворення вихідної моделі в модель з меншою кількістю змінних і обмежень, яка забезпечує наближений (порівняно з вихідною) опис досліджуваного процесу або об'єкта.

Головним фактором дослідження фрактальної структури фінансових часових рядів є знаходження фрактальних показників у рамках досліджуваного ряду. Так у роботі доведено, що фрактальна розмірність близька до 1,5 ($\mu = 0.5$) є справедливою лише для 15—25% фінансових часових рядів, що є незадовільним показником, решту часу поведінка рядів істотно відрізняється від броунівського руху. На думку автора, даний факт пояснюється тим, що ділянки «не нормальної» (або «аномальної») поведінки мають коротку тривалість. У той же час для перевірки на нормальність відомими методами необхідні великі

масиви даних (від кількох сотень до кількох тисяч точок). Всередині інтервалу, що використовується для перевірки на нормальність стандартними методами, виявиться велика кількість ділянок з різною поведінкою. При розрахунку тестових значень характеристики персистентних і антиперсистентних ділянок компенсуються і підсумкові значення виявляться близькими до нормальних.

На основі вищевикладеного задамося, що N — кількість подій, L — характерна довжина, а D — фрактальна розмірність; зафіксуємо значення для фінансового часового ряду $D = 1$.

Імовірнісна міра розмірності визначається за формулою: $[N(|Y_t - Y_{t-1}|) > L/N(|Y_t - Y_{t-1}|)]/L$. Частка спостережень, що лежать за межами порога L щодо загальної вибірки, варіюється в залежності від:

$$[N(|Y_t - Y_{t-1}|) > L/N(|Y_t - Y_{t-1}|)] \approx L^d \quad (2.2)$$

де знаком ≈ позначається асимптотична рівність. У цьому сенсі розмірність — це міра ентропії або хаотичності.

Вимірювання змінюється залежно від порогу і тому часто вимірюється як асимптотична межа, коли поріг наближається до нуля. Модифікацією цього методу є оцінка середньої неоднорідності процесу за допомогою кореляції C, що призводить до різниці в діапазоні між розміром маскування d і дробовим розміром D:

$$C = D - d \quad (2.3)$$

Якщо $C \neq 0$, то процес називається фрактальним.

Для низьких часток (близьких до нуля) процес є більш однорідним: екстремальних коливань мало. І навпаки, вищі значення C характеризуються більшою кількістю екстремальних подій або коливань вище порогового значення. Процес стає менш однорідним, коротшим, менш частим і менш стабільним.

Взаємозв'язок між кореляцією і симетричним масштабуванням показано в наступних рівняннях, де τ — часовий масштаб від 1 до T, де T — найбільший часовий масштаб; μ — індекс; q — кількість індикаторів масштабування. Симетричне масштабування відбувається наступним чином::

$$\mu(|Y_t - Y_{t-1}|^q) \approx [\mu(|Y_t - Y_{t-1}|^q)]\left[(\tau/T)^{\zeta(q)}\right] \quad (2.4)$$

де ζ — функція, яка включає в себе три параметри C_1, H, α:

$$\zeta(q) = qH - \{[C_1/(\alpha - 1)](q^\alpha - q)\}, \text{де } \alpha \neq 1 \quad (2.5)$$

$$\zeta(q) = qH - (C_1 q ln q), \text{де } \alpha = 1 \quad (2.6)$$

Параметр C_1 — кодування, пов'язане з масштабуванням засобів вибірки. Коли $C_1 = 0, \zeta(q)$ є лінійним трендом. Коли $C_1 \neq 0$, кривизна осі $\zeta(q)$ залежить від корозмірності і розподілу ймовірностей. Коефіцієнт α характеризує розподіл ймовірності. Випадок $\alpha = 2$ відповідає гаусівському розподілу, тоді як α=1 відповідає розподілу Коші. Для більшості економічних процесів 1<α<2 . У цьому випадку розподіл має важчі хвости, ніж у стандартному розподілі, і дисперсія змінюється з часом. Цікавою властивістю як послідовностей 1<α<2, так і 0<C1<1 є те, що інтегрування зазвичай не призводить до згладжування. Замість цього, інтегрування показує дискретні стрибки.

Коефіцієнт H характеризує індекс дробовості, тобто величину, яка зменшується зі збільшенням часового лагу між двома парами однакових значень. Позначення H насправді бере початок з експонента Херста, або відкликаного коефіцієнта діапазону [9]. Проте в цих рамках H оцінюється як один з низки коефіцієнтів масштабування. Ця статистика пов'язана з порядком інтеграції адитивної константи. Для процесу $I(0)$ $H = 0.5$. Для нестаціонарного процесу

порядок інтеграції може бути відновлений шляхом оцінки темпів змін [21]. Цей метод надійний як для нелінійності, так і для дробових порядків інтеграції.

Враховуючи подвійні властивості екстремальних подій та симетрії масштабу, природно розробити моделі, які можуть охопити обидві властивості, і алгоритм аналізу фінансової сфери базується на лінгвістичному моделюванні. Перехід стану визначається як перехід від низького до екстремального стану з низькою або середньою волатильністю і позначається множиною. Модель переходу стану валютного курсу може бути описана наступним чином:

$$X_{t+1} = \omega_0 + \omega_1 X_t \tag{2.7}$$

Стан є можливість передбачати ітеративно за допомогою форми регресії:

$$S_{Et+1} = \omega_0 + \omega_1 S_{Et} + \omega_2 X_t + \omega_3 (X_t - X_{l-1}) + \varepsilon_l \tag{2.8}$$

в якому реальна ціна золота оцінюється на основі як темпів зміни, так і рівня фінансових показників, оскільки було виявлено, що це покращує результати практичного застосування.

Цей тип моделі, звичайно, має один внутрішній ризик: переходи стану можуть бути подані неправильно. Для захисту від цієї можливості слід оцінювати моделі з перехідними станами і без них. Ще одна проблема в реалізації, в тому числі стану всіх змінних фактичної ціни на золото (ФЦЗ) [33] можуть зробити модель занадто складною. Обмежене число переходів станів за окремими причинно-наслідковими факторами може бути кращим. Остання проблема полягає в тому, як вказати крайню подію. Це можна визначити емпірично, протестувавши значення, які дають найкращі прогнози.

Ідея використання симетрії масштабів для прогнозування спочатку була запропонована для фізичних процесів, які демонструють сильну симетрію між великими і малими масштабами. Оскільки пропорційні співвідношення у фінансових журналах зазвичай обмежуються короткими інтервалами, моделі тут

використовують лише наближену симетрію. Для прогнозованого темпу змін існує симетрія по відношенню до запізнілого значення:

$$(X_{t+1} - X_t) = \lambda_{1t}(X_{t+1} - X_{t-1}) \quad (2.9)$$

де λ_{1t} — коефіцієнт пропорційності, числовий коефіцієнт позначає відстань поділу на ФНЗ, а t — коефіцієнт вказує на зміну часу. Аналогічним чином, для темпів змін:

$$(X_t - X_{t-1}) = \lambda_{2t}(X_t - X_{t-2}) \quad (2.10)$$

рішення для λ_{2t}:

$$\lambda_{2t} = [(X_t - X_{t-1})/(X_t - X_{t-2})] \quad (2.11)$$

Однією з проблем практичної реалізації є те, що знаменники коефіцієнтів можуть містити нульові значення. Для обчислення відсутніх значень можна використовувати будь-яку кількість процедур інтерполяції. Якщо коефіцієнт є невизначеним, дані згладжуються експоненціально і підставляється коефіцієнт, заснований на згладжених даних. Прогностична модель для всієї процедури:

$$X_{t+1} = \omega_0 + \omega_1 X_t + \omega_2[(I_t - \pi_t^e) - (I_{Ft} - \pi_{Fl}^e)] + \omega_3\lambda_{2t} + \varepsilon_l \quad (2.12)$$

Як зазначалося вище, більшість структурних рівнянь в макроекономіці не враховують довгострокові фрактали. Це виправдовує використання структурних рівнянь і побудову таблиць залишкових коефіцієнтів.:

$$(\varepsilon_l - \varepsilon_{t-1}) = \gamma_{2t}(\varepsilon_t - \varepsilon_{t-2}) \quad (2.13)$$

$$\gamma_{2t} = [(\varepsilon_t - \varepsilon_{t-1})/(\varepsilon_t - \varepsilon_{1-2})] \quad (2.14)$$

де γ_{2t} — коефіцієнт пропорційності для залишку. Таким чином, модель прогнозування стає:

$$X_{t+1} = \omega_0 + \omega_1 X_t + \omega_2[(I_t - \pi_t^e) - (I_{Ft} - \pi_{Fl}^e)] + \omega_3\gamma_{2t} + v_t \qquad (2.15)$$

де v_t — залишок від цієї регресії.

Основним недоліком моделі є те, що масштабні коефіцієнти відомі лише для поточного періоду. Крім того, під час практичних тестів часто виявляється, що масштабний коефіцієнт є занадто нестабільним для ефективного прогнозування. Розумним рішенням є розкладання масштабного коефіцієнта на дві складові: систематичний процес та залишки [20]. Нехай γ_{2St} позначає системну складову у γ_{2t}. Зазвичай форма цієї компоненти не відома прогнозисту, але її можна оцінити за допомогою лагової регресії [23]:

$$\gamma_{2St} = \omega_0 + \omega_1\gamma_{2l-1} + \omega_2\gamma_{2t-2} + v_t \qquad (2.16)$$

Прогнозоване значення, γ_{2St+1}, потім включається в рівняння прогнозування. Отримуємо:

$$X_{t+1} = \omega_0 + \omega_1 X_t + \omega_2[(I_t - \pi_t^e) - (I_{Ft} - \pi_{Fl}^e)] + \omega_3\gamma_{2St+1} + v_t \qquad (2.17)$$

Оцінка фрактальних параметрів ґрунтується на масштабуванні середнього логістичного значення абсолютної логістичної різниці зі збільшенням відстані розділення [30]. По суті, це зв'язок між симетрією масштабування та функцією ζ(q). Фрактальні індекси є локалізованими фрактальними особливостями фінансових часових рядів. На основі описаних методів і алгоритмів аналізу та прогнозування створено узагальнену методологію аналізу та розроблено програмний продукт.

2.3 Моделювання оптимізаційного рішення

Для реалізації запропонованих рішень наукових задач дослідження

пропонується система для прогнозування фінансових показників банку методом фрактального аналізу, який передбачає масштабування часового ряду та виділення корозмірності. Система дозволить спостерігати коливання чистого прибутку та здійснювати прогнозування відносно даних. Система дозволяє реалізувати модель автоматизованої системи для аналізу та прогнозування фінансових часових рядів з використанням лінгвістичного моделювання. На основі отриманих даних можна визначити ефективність методики автоматизованого аналізу та прогнозування фінансового часового ряду за допомогою лінгвістичного моделювання.

Система прогнозування фінансових результатів діяльності банків з використанням фрактального аналізу призначена для розв'язання задачі прогнозування часових рядів з фрактальними характеристиками (в якості фрактальних даних пропонується показник чистого прибутку), фактично є експертною системою підтримки прийняття рішень (СППР), що розробляється, побудована на основі інструментарію, спрямованого на оцінку стану складної системи (об'єкта), в якій присутня велика кількість значущих факторів [6] або систем фінансових ризиків. Вона передбачає наявність взаємозв'язків і взаємозалежностей між факторами, які можуть бути не задані в явному вигляді на початку використання системи. Схема інформаційного зв'язку при прогнозуванні часових рядів з використанням фрактального аналізу наведена на рис. 2.3.

Система прогнозування фінансових показників банку з використанням фрактального аналізу - це програмний комплекс (ПК), призначений для аналізу та прогнозування фінансових часових рядів, представлених показником чистого прибутку (ЧП), за допомогою лінгвістичної моделі. Користувачів програми прогнозування фінансових показників банку з використанням фрактального аналізу можна поділити на три категорії: системні адміністратори та окремі особи або групи осіб як кінцеві користувачі.

Складовими додатку для прогнозування фінансових показників банку методом фрактального аналізу (рис. 2.4) є наступні програмні модулі:

Модуль авторизації використовується для надання користувачеві дозволу на виконання певних операцій і підтвердження отримання дозволу, коли користувач намагається виконати ці операції.

Рисунок 2.3 — Схема інформаційних зв’язків системи для прогнозування фінансових показників банку методом фрактального аналізу

Модуль управління використовується для нагляду за цілісністю інформаційних полів і даних, забезпечення дотримання стандартів безпеки та прийнятного рівня продуктивності комп'ютерів і програмного забезпечення.

Модуль обліку результатів має дуже важливі функції, а саме:

— адміністратори та користувачі можуть ознайомитися з чинними стандартами;

— користувачі можуть вибирати власні прийнятні часові стандарти та розуміти результати, сформовані за прогнозованими вибірками.

Модуль обробки результатів використовується для оцінки результатів.

Модуль відображення результатів призначений для того, щоб безпосередні користувачі системи могли ознайомитися з поточним курсом, ввести свої

персональні дані, а потім переглянути результати після того, як вони завершили операцію.

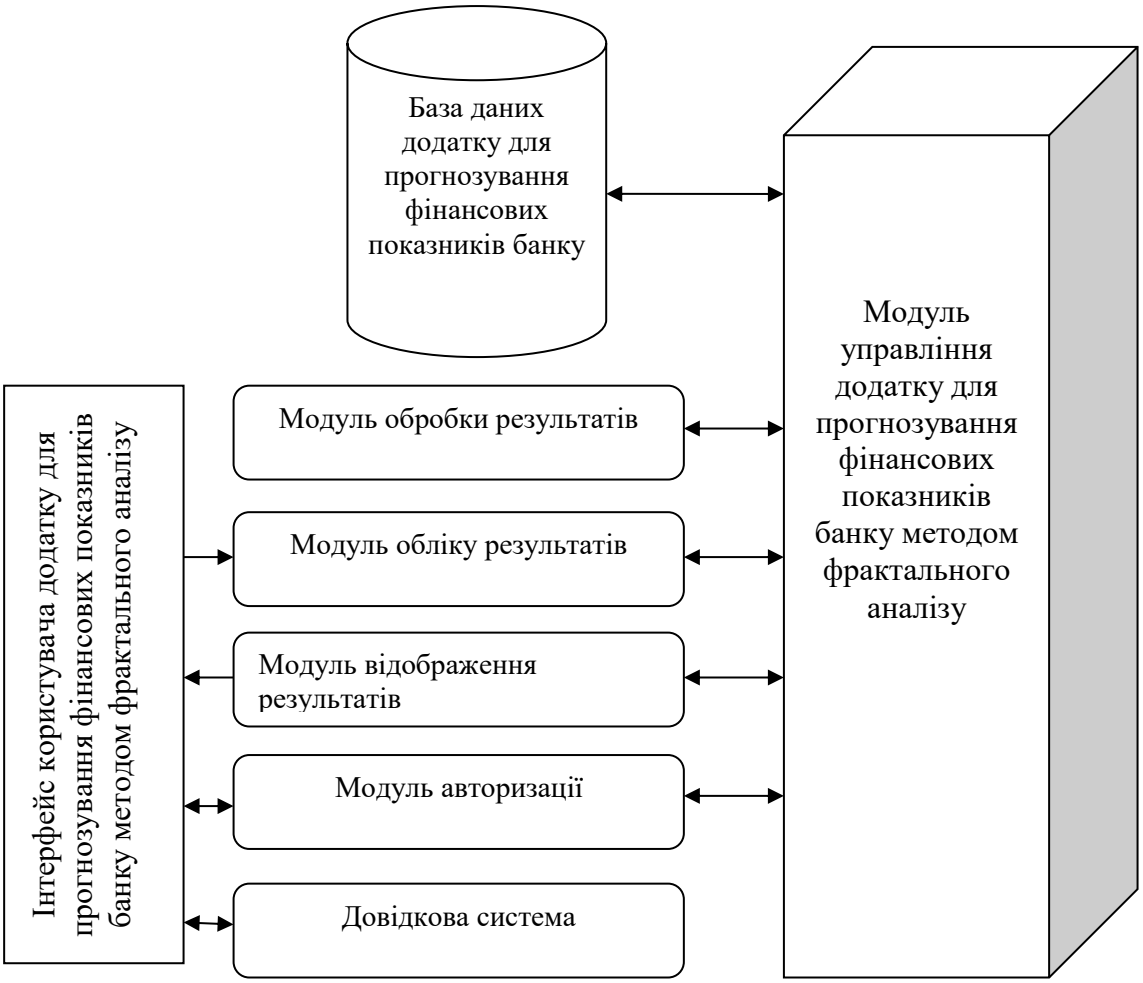


Рисунок 2.4 — Узагальнена структура програмних модулів додатку для прогнозування фінансових показників банку методом фрактального аналізу

Довідкова система призначена для інтерактивного обміну інформацією між усіма модулями програми, що прогнозує часові ряди за допомогою фрактального аналізу.

Для зберігання даних (фінансові дані та результати, паролі доступу тощо) рекомендується використовувати відповідну базу даних для прогнозування часових рядів за допомогою фрактального аналізу відповідної структури.

Базові принципи роботи, які повинен забезпечувати додаток для прогнозування часових рядів методом фрактального аналізу є такими.

На першому етапі, залежно від типу користувача (адміністратор системи,

користувач або група користувачів) та його персональних даних, модуль авторизації користувача визначає права користувача (за допомогою логіну та паролю) та відкриває доступ до відповідних робочих модулів системи (модуль управління, модуль контролю результатів обробки фінансових часових рядів або модуль введення фінансових специфікацій).

Після авторизації в системі може працювати лише персонал з підтвердженими правами доступу до відповідних модулів системи.

Після авторизації користувач може отримати доступ до відповідного робочого модуля системи та виконувати відповідні операції. Результати роботи в системі автоматично обробляються модулем обробки результатів фінансових часових рядів, редагуються та відображаються в базі даних.

При цьому користувач має постійний доступ до довідкової системи програми для прогнозування часових рядів з використанням фрактального аналізу.

Засобом програмування в роботі обрано розробку системного макету додатку для прогнозування часових рядів методом фрактального аналізу, середовищем розробки слугує Zend Studio, мовою програмування PHP. Модель для зберігання даних сформована з урахуванням базової реляційної моделі в системі Embarcadero ER Studio; а база даних отримана завдяки використанню мови SQL Server.

Запропоновані програмні засоби, які використані для розробки додатку для прогнозування часових рядів методом фрактального аналізу надають можливість швидкого програмування та логічного розгортання розробленого додатку.

Інноваційні підходи до розробки програмного забезпечення засновані на створенні програмних продуктів, які складаються з окремих модулів. Такий алгоритм пояснюється зручністю і простотою програмного забезпечення, що планується розробляти. Ще однією значною перевагою слугує те, що окремі модулі програм можуть бути використані у складі інших програмних продуктів або окремо.

Тому макет додатку для прогнозування часових рядів методом

фрактального аналізу розроблено у вигляді взаємопов’язаного набору модулів, згідно моделі макету.

Показано процес аналізу та прогнозування фінансових часових рядів за допомогою лінгвістичного моделювання (включаючи масштабування часових рядів та виділення кореляцій), а також місце запропонованого застосування прогнозування часових рядів за допомогою фрактального аналізу зображено на рис. 2.6.

Запропонований додаток для прогнозування часових рядів методом фрактального аналізу дозволяє здійснювати аналіз та прогнозування фінансових показників банку, моделювання за допомогою мови передбачає масштабування часових рядів та виділення кореляцій. Тому було розроблено низку дій, виконання яких дозволить отримати необхідну інформацію щодо доцільності використання лінгвістичного моделювання для аналізу та прогнозування фінансових часових рядів.

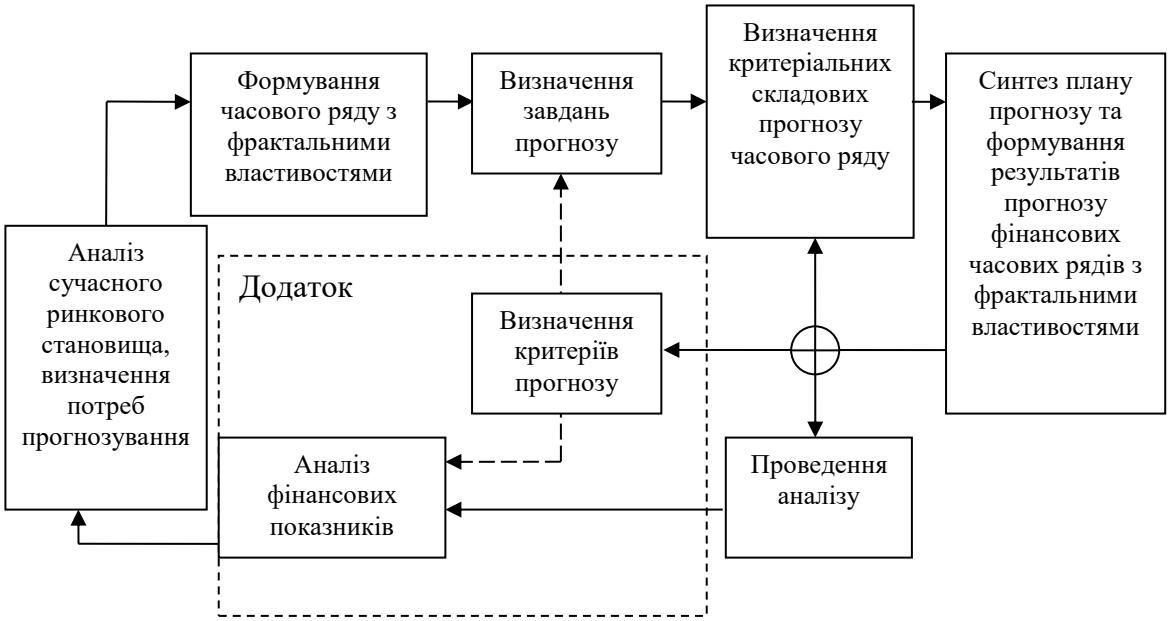


Рисунок 2.6 — Формування складових процесу аналізу та прогнозування фінансового фінансових показників банку

Виходячи з того, що сучасний економічний розвиток бізнесу характеризується впровадженням систем управління ресурсами та

автоматизацією прогнозування, для прогнозування часових рядів рекомендується використовувати застосування фрактального аналізу.

Дослідницька компанія Forrester вважає, що економетрика, як механізм аналізу часових рядів, відіграє другорядну роль в процесі аналізу та прогнозування ряду фінансових рядів, підвищуючи значення експертних оцінок та сценарних методів. методів у формуванні загального прогнозу..

Існуючі автоматизовані системи аналізу та прогнозування, у своїй більшості, мають однакову структуру, яка виходить на три технологічні рівні, кожен з яких має відповідні функціональні блоки, що виконують покладені на них функції. Блок управління процесом формування фінансових рядів(рис. 2.7) на рівні, призначеному для користувача, виконує управління принципами формування та структуризації.

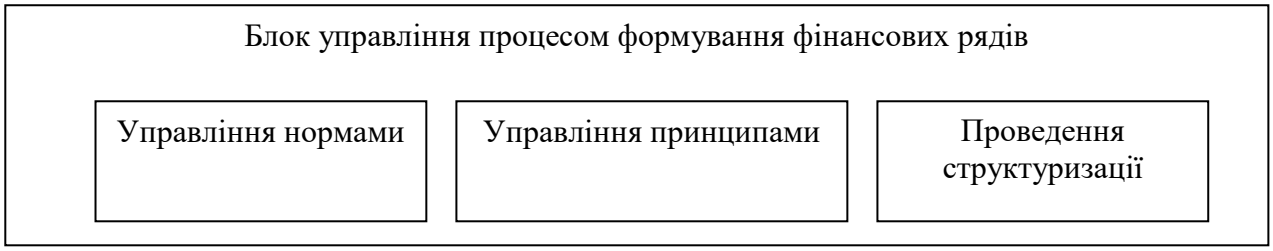


Рисунок 2.7 — Структура сучасної системи прогнозування (за даними Forrester Research)

Аналітики центру TAdviser виділяють 29 функціональних модулів для рішень найвищого класу. У напрямку розвитку фінансової аналітики на сучасних економічних ринках таких як Forex (рис. 2.8) виділяють управління: процесом, що породжує часовий ряд, управління процесом формування статичного ряду, прогнозуванням, закономірними складовими (тренд, сезонна складова, циклічна складова).

Процес управління прогнозуванням включає процеси планування, організації, контролю, обліку та аналізу. Таким чином, можна визначити місце запропонованого застосування прогнозування часових рядів за допомогою фрактального аналізу в системі економічного прогнозування (рис. 2.9).

Завдяки формуванню єдиної системи прогнозування на основі сучасних алгоритмів прогнозування та механізмів економічного аналізу, з урахуванням узгодженості всіх модулів та ортогональності методів, автоматизована система аналізу та прогнозування фінансових часових рядів з використанням мовних моделей є ефективним та перспективним комплексом в сучасній системі економічного прогнозування.

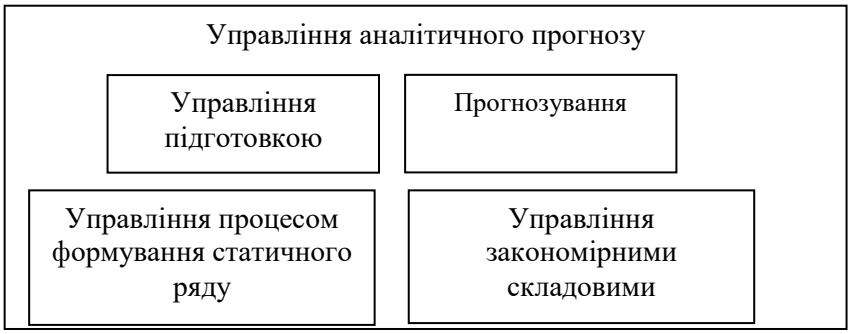


Рисунок 2.8 — Структура сучасної системи прогнозування (згідно з даними TAdviser)

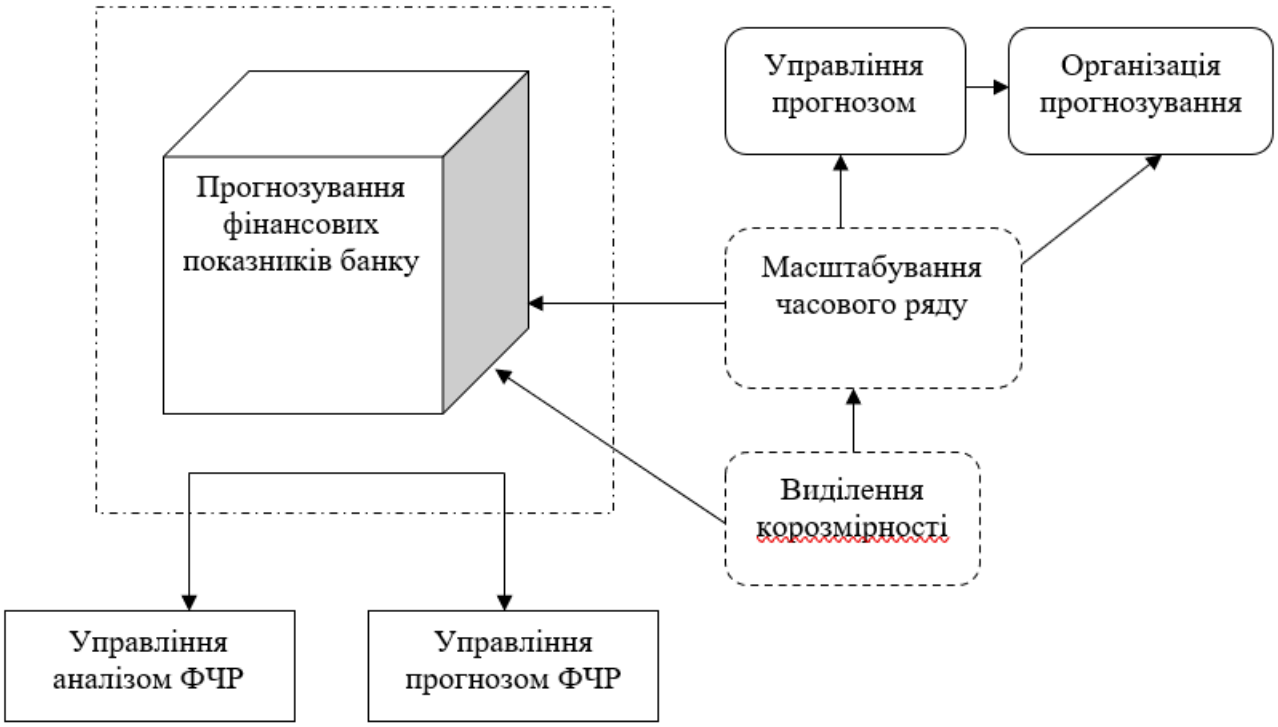


Рисунок 2.9 — Місце запропонованого додатку для прогнозування часових рядів методом фрактального аналізу в сучасних системах економічного прогнозування

3 ПРОГРАМНА РЕАЛІЗАЦІЯ ВЕБ-ДОДАТКУ ДЛЯ ПРОГНОЗУВАННЯ В ДАТА-МАЙНІНГУ

3.1 Постановка задачі на проектування веб-додатку для прогнозування в дата-майнінгу

Основними цілями створення програми для прогнозування в дата-майнінгу є:

— виконання аналізу часових рядів різного виду;

— автоматизація процесу виведення результату;

— автоматизація процесу проведення аналізу та прогнозування часових рядів.

Основою методики створення інформаційної технології для прогнозування в дата-майнінгу є Model-view-controller. Схема патерна наведена на рисунку 3.1.

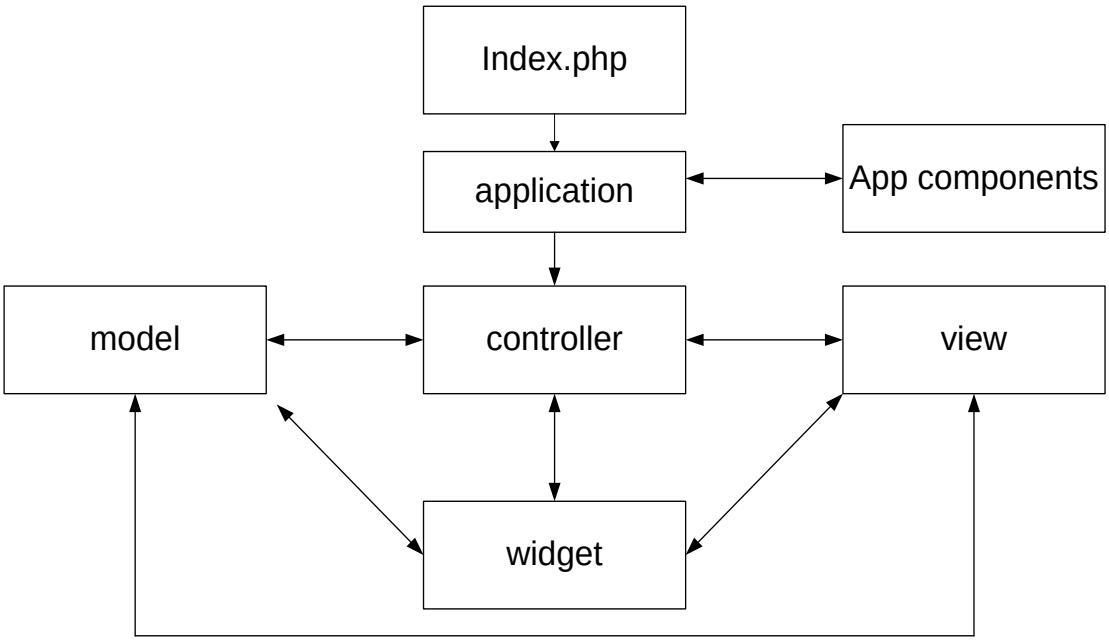


Рисунок3.1 — СхемаModel-view-controller

Для вирішення завдання прогнозування часового ряду пропонується використовувати схему, представлену на рис. 3.2.

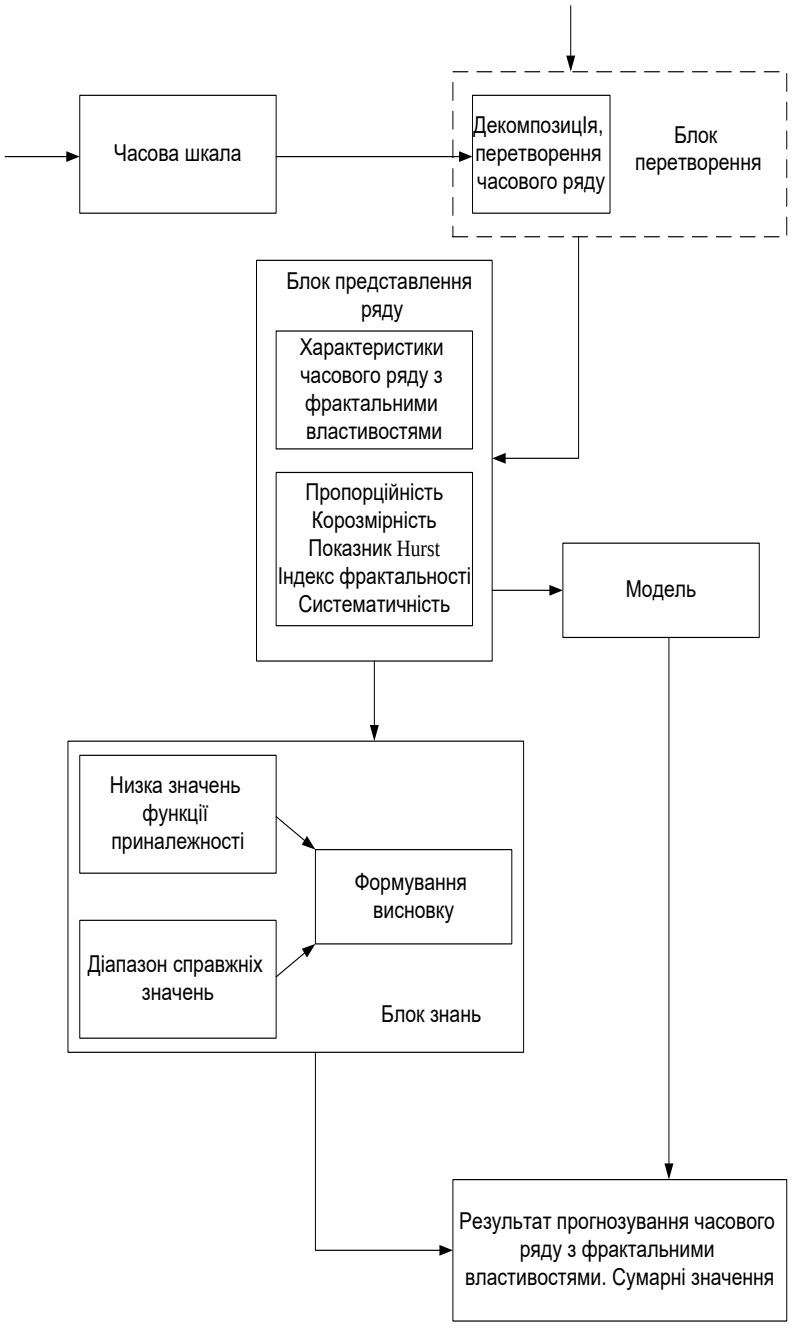


Рисунок 3.2 — Схема реалізації запропонованої моделі прогнозування часового ряду

3.2 Інформаційне забезпечення системи

Попередню модель інформаційної системи для аналізу та прогнозування в дата-майнінгу можна представити у вигляді функціональних схем претендентів.

Головна діаграма прецедентів, інформаційної системи показана на рис. 3.3.

Суть структури системи полягає у поділі функції на дві основи.

Перша база призначена для генерації часових рядів та надання загальної моделі ряду (рис. 3.4).

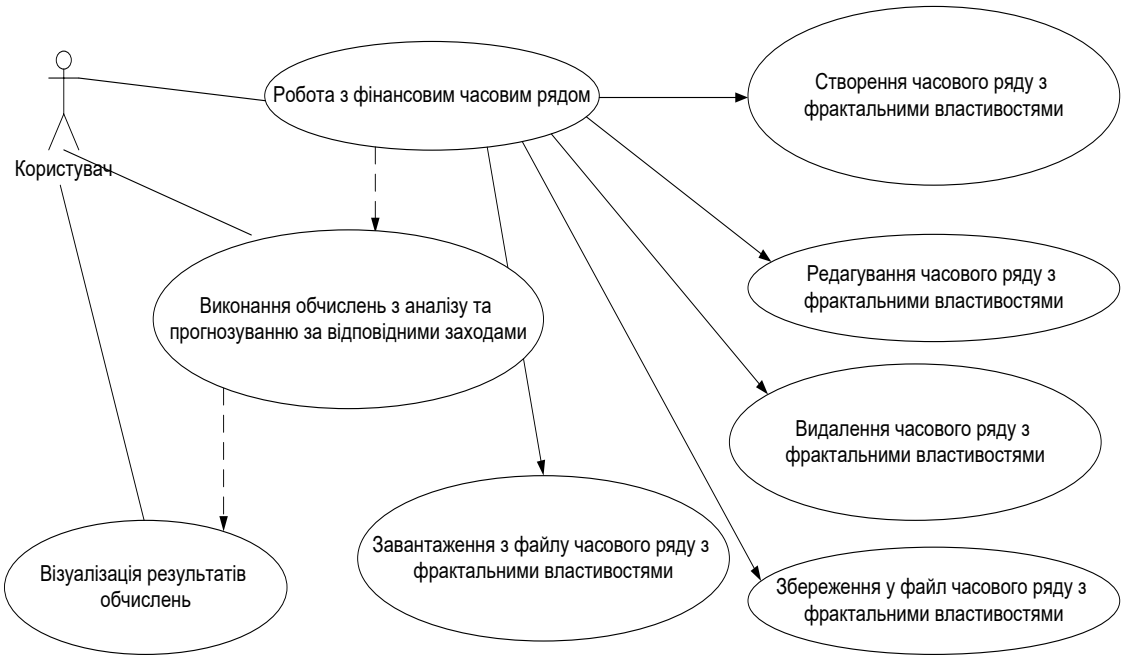


Рисунок 3.3 — Діаграма прецедентів інформаційної системи для аналізу та прогнозування в дата-майнінгу

Другий фундамент базується на математичній секції і виконує всі необхідні розрахунки (рис. 3.5).

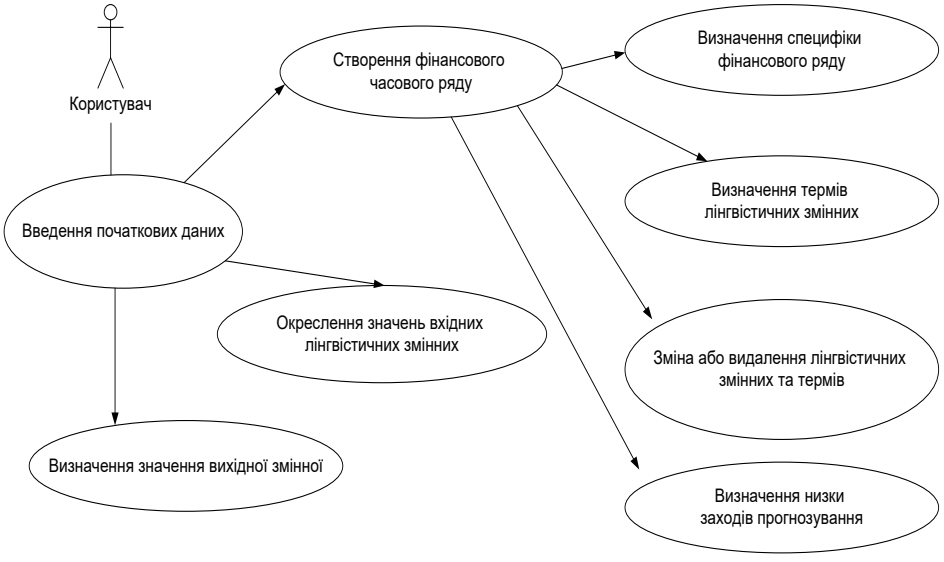


Рисунок 3.4 — Діаграма формування часового ряду та представлення загальної моделі ряду

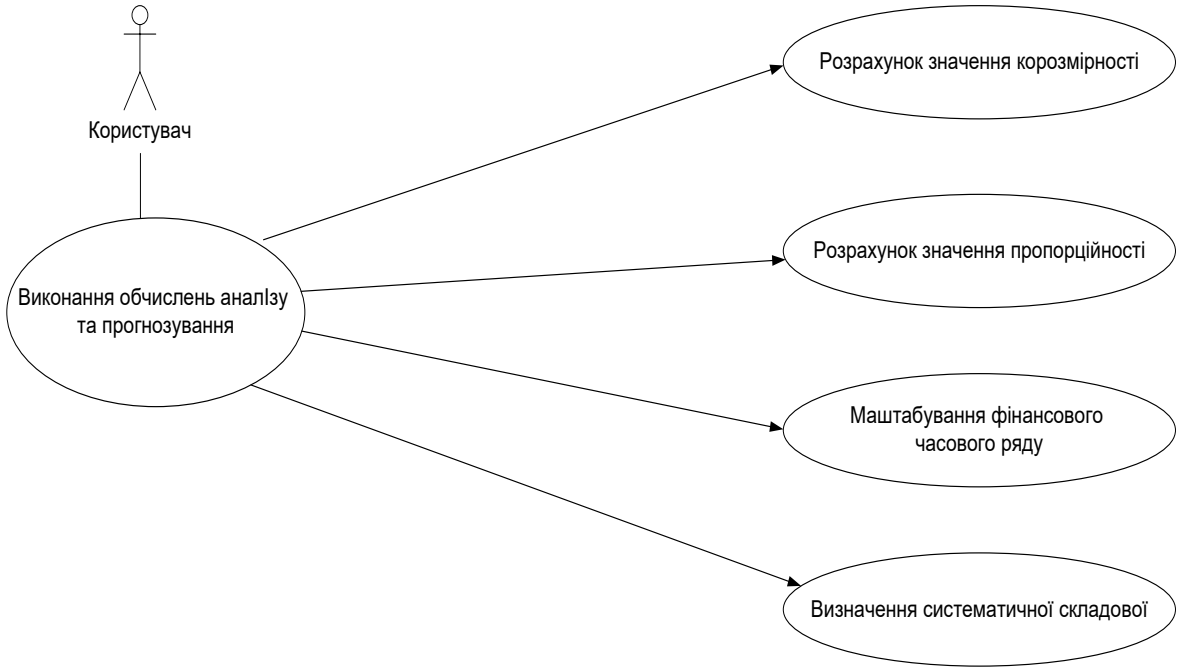


Рисунок 3.5 — Діаграма принципів проведення математичних обчислень та формування результату прогнозу

Механізм виведення прогнозних значень базується на моделі, яка визначає вхідний ряд. У разі визначення структури часового ряду прогнози виводяться у вербальній формі з урахуванням тренду на всьому часовому проміжку. Числова форма прогнозів визначається на основі індивідуальної оцінки характеристик кожної окремої одиниці в числовій послідовності. Виходячи з цього, інформаційна система аналізу та прогнозування даних інтелектуального аналізу характеризується здатністю робити точні багаторівневі прогнози всіх можливих типів та інтенсивності основних тенденцій часового ряду.

3.3 Розробка структури веб-додатку для прогнозування в дата-майнінгу

У рамках даного дослідження розробляється веб-сайт для прогнозування в дата-майнінгу. Побудова веб-додатку починається з розробки БД яка буде зберігати всі вхідні дані.

Так як система проектується на локальному сервері, здійснюємо побудову БД із використанням PhpMyAdmin. Щоб працювати з великою кількістю даних і

мати можливість змінювати та зберігати ці дані на веб-сайті, потрібно використовувати бази даних. MySQL — найбільш зручна та поширена система для роботи з базами даних. У базі даних вся інформація зберігається у якості пов'язаних між собою таблиць. PhpMyAdmin — це зручний веб-додаток, написаний мовою PHP, що дозволяє працювати з базами даних MySQL. PhpMyAdmin дозволяє легко створювати бази даних MySQL, таблиці, будувати запити до баз даних MySQL

Побудуємо БД у PhpMyAdmin у Open Server Panel, аналогічна робота проводиться на реальному хостингу. Щоб створити базу даних, потрібно в Open Server Panel увійти до PhpMyAdmin. Вибираємо вкладку "Додатково" і запускаємо PhpMyAdmin.

Вводимо дані:

Користувач: root

Пароль: порожній

При виконанні будь-якої операції в phpmyadmin потрібно натискати кнопку "Вперед". Щоб у PhpMyAdmin створити базу даних, тиснемо "Створити БД" у лівому верхньому кутку для бази даних та вибираємо кодування. Як правило, використовується кодування utf8_general_ci. Натискаємо "Створити".

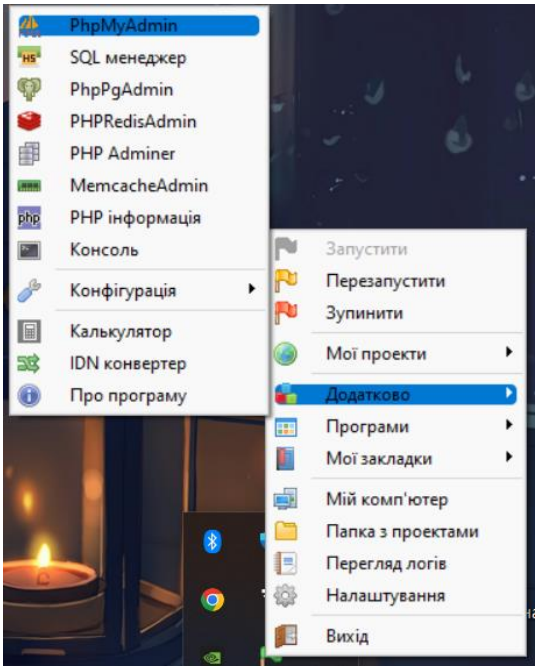


Рисунок 3.6 — Запуск PhpMyAdmin

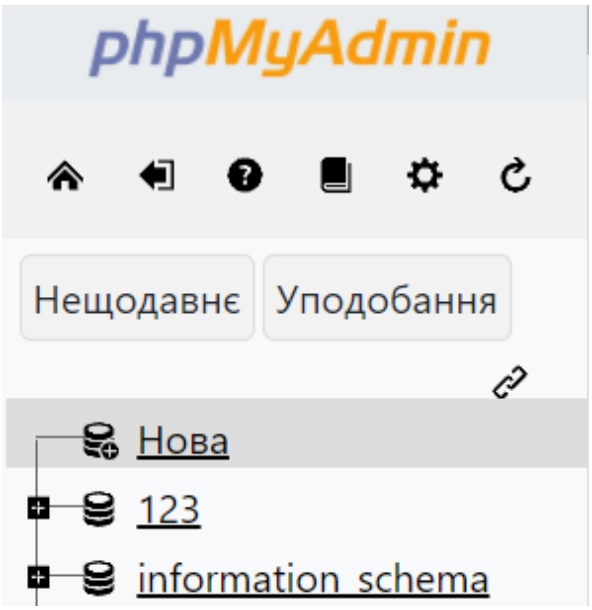


Рисунок 3.7 — Створення БД

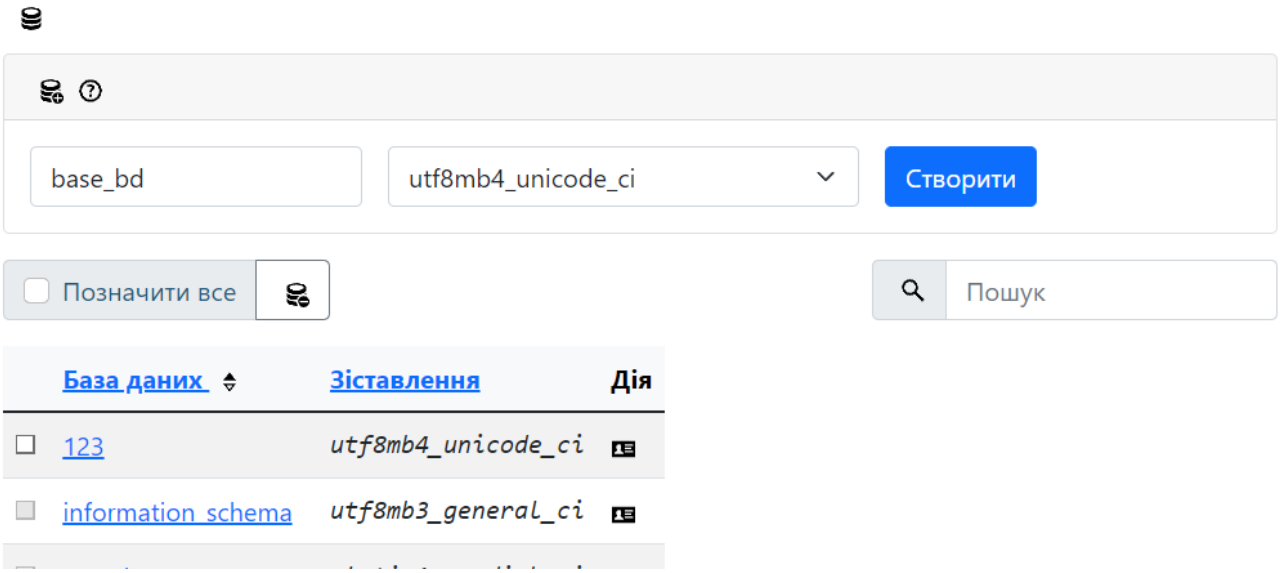


Рисунок 3.8 — Створення таблиць БД

База даних складається із таблиць. Щоб створити таблицю в базі даних PhpMyAdmin, потрібно вибрати базу даних та натиснути "Створити таблицю". Вводимо її ім'я та кількість стовпців. Після цього тиснемо "Вперед".

Для першого стовпця більшості таблиць у базі даних MySQL ставиться унікальний номер запису ID. Він має індексуватися, тому вибираємо індекс PRIMARY. Ставимо галочку на A_I, щоб номер запису при додаванні автоматично збільшився. Другий стовпець відповідає за ім'я користувача, має назву "name" і тип VARCHAR. Можна обмежити його довжину, поставимо 255.

У результаті отримаємо таблицю.

Після завершення створення таблиці натискаємо "Зберегти".

Щоб додати запис до таблиці бази даних PhpMyAdmin, виділяємо потрібну нам таблицю і переходимо у вкладку "Вставити". Заповнюємо всі поля, крім ID(воно заповниться автоматично) і натискаємо "Вперед".

Для того щоб запустити розроблений веб додаток необхідно запустити локальний сервер, у нашому випадку OpenServer.

Перед початком тестування веб-додатку для прогнозування в дата-майнінгу варто здійснити налаштування всіх необхідних складових.

Веб Сервер**:** можна встановити локальний сервер (на своєму комп'ютері) або використовувати віддалений (але для цього треба купити домен та хостинг). Рекомендовано встановити локальний веб-сервер OpenServer. Він безкоштовний, має великий набір функцій. Одна з переваг — він портативний, його можна розгорнути на флешку.

Необхідно перейти в меню « Мої сайти » і запустити свій сайт («db») на установку.

Веб-сайт дозволить автоматизувати роботу зі створення прогнозу чистого прибутку банку. Інтерфейс користувача направлено на підтримку однаковості у відображенні даних, що робить систему зрозумілою та легкою для освоєння, інформаційну мінімізацію, що дозволить структурувати основні дані та сконцентрувати увагу на найголовнішому.

Загалом сайт має чотири основні сторінки:

— графіки;

— прогноз;

— довідники;

— дії.

На сторінці «Графіки» користувач має можливість обрати показник банку та переглянути графік коливання чистого прибутку банку.

На сторінці «Прогноз» користувач має можливість обрати показник банку та переглянути графік прогнозування у відповідності до обраного показника.

На сторінці «Довідники» користувач має можливість переглянути різні компанії / банки, які працюють за аналогічною схемою.

На сторінці «Дії» користувач має можливість оновити дані, навчити нейронну мережу та оновити показник Херста.

4 ПЕРЕВІРКА ПРАЦЕЗДАТНОСТІ ТА ТЕСТУВАННЯ

4.1 Тестування веб-додатку для прогнозування в дата-майнінгу

Проведемо експериментальне дослідження ефективності розробленого веб-додатку для прогнозування в дата-майнінгу.

Запустимо веб-сайт на локальному сервері, завантажемо БД та протестуємо роботу зі зведеними даними щодо фінансової діяльності Банку. Результати виведемо на рисунках (4.1 — 4.5)

Рисунок 4.1 — Результат роботи розробленого веб-додатку для прогнозування в банку Восток

Рисунок 4.2 — Результат прогнозування

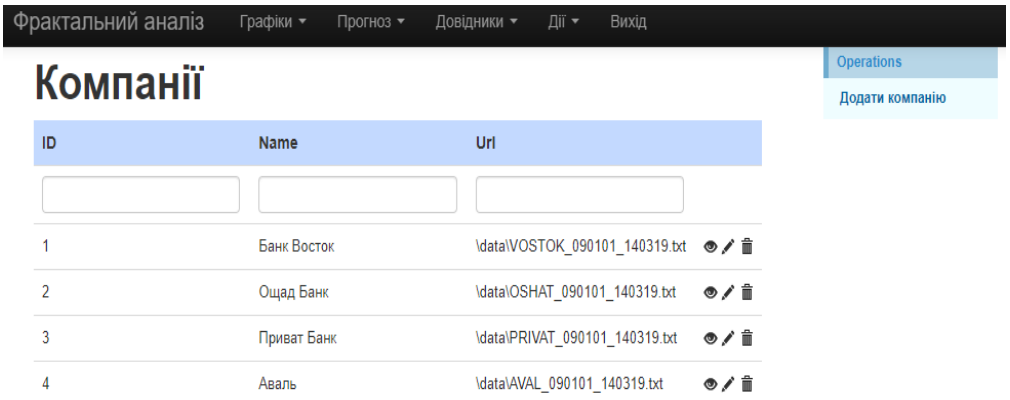


Рисунок 4.3 — Банки, які працюють за аналогічною схемою

Рисунок 4.4 — Результат роботи розробленого веб-додатку для прогнозування в Ощад Банку

Рисунок 4.5 — Результати Результат прогнозування в Ощад Банку

При проведенні тестування збоїв та недоліків у роботі веб-сайту не виявлено.

4.2 Верифікація результатів

Порівняємо відомі методи прогнозування з розробленим у рамках даної кваліфікаційної роботи.

Прогноз за методом експоненціального згладжування званий методом Брауна — найкраще рішення для прогнозування, коли дані про продажі доступні лише за кілька періодів (місяць, день, тиждень, квартал) і ще не зрозуміло, чи є тенденція висхідною або низхідною.

Формула розрахунку прогнозу:

$$\hat{Y}_{t+1} = k * Y_t + (1-k) * \hat{Y}_t \quad (4.1)$$

де $\hat{Y}_{t+1}$ — прогноз на наступний період t + 1;

Y_t — дані для прогнозу за поточний період t (наприклад, продажу по місяцях);

k — коефіцієнт згладжування ряду, k задається в діапазоні від 0 до 1, $0 < k < 1$

$\hat{Y}_t$ — значення прогнозу на поточний період t. Причому в перший період (місяць, день ...) $\hat{Y}_1 = Y_1$, тобто $\hat{Y}_t$ в перший період рівні продажу в цей період.

Прогноз за методом експоненціального згладжування = коефіцієнт згладжування * Останнім фактичне значення продажів + (1 коефіцієнт згладжування) * попередній прогноз за методом експоненціального згладжування.

Зокрема, модель передбачає періодичний перерахунок прогнозів після закінчення попереднього періоду та отримання нових даних для прогнозів попереднього періоду.

k — коефіцієнт згладжування ряду, його значення задається вручну від 0 до 1. Чим k більше, тим більше вплив останніх періодів на прогноз.

Зелений графік — значення прогнозу при коефіцієнті згладжування k = 0,1.

Видно, що періодично дана модель стикається з фактичними продажами (синім графіком), але набагато рідше, ніж червоний графік. Також зелена модель більш згладжена і повільніше реагує на сплески в останніх періодах, ніж червоний графік.

=$D$3*C11+(1-$D$3)*D11

C	D	E	F	G
	Експонентна модель 1			Експонентна м
	k=	Точность	0,01%	k=
	0,8			0,1
Сума (тис. грн.)	Прогноз Ŷ=	Похибка моделі	среднеквадра	Прогноз Ŷ=
778 949,28	778949,28	0,00	0,00	778949,28
870 963,44	778949,28	92014,16	0,01	778949,28
874 286,38	852560,61	21725,77	0,00	788150,70
819 776,59	869941,22	-50164,63	0,00	796764,27
767 763,37	829809,52	-62046,14	0,01	799065,50
739 061,21	780172,60	-41111,40	0,00	795935,29
763 468,55	747283,49	16185,06	0,00	790247,88
731 872,86	760231,54	-28358,68	0,00	787569,95
853 940,80	737544,60	116396,20	0,02	782000,24
960 662,73	830661,56	130001,18	0,02	789194,29
901 756,90	934662,50	-32905,60	0,00	806341,14
1 088 090,36	908338,02	179752,35	0,04	815882,71

Рисунок 4.6 — Процес розрахунку методом Брауна

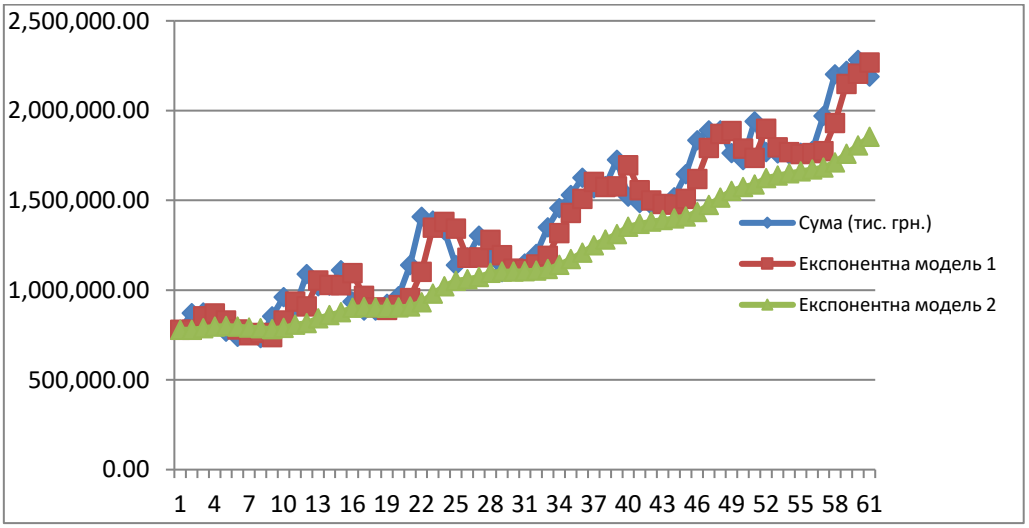


Рисунок 4.7 — Графік прогнозування методом Брауна

Розрахуємо точність прогнозу при k = 0.8 і k = 0.1. Для кожної моделі визначимо: помилку моделі, відхилення, середньоквадратичне відхилення та

точність прогнозу.

Помилку моделі = для цього в кожному періоді з фактичних продажів віднімаємо прогноз продажів на цей період.

Відхилення = для кожного періоду розрахуємо відношення квадрата помилки моделі до квадрату прогнозу на цей період

=$G$3*C9+(1-$G$3)*G9

	D	E	F	G	H
	Експонентна модель 1			Експонентна модель 2	
	k=	Точность	0,01%	k=	Точность
	0,8			0,1	
грн.)	Прогноз Ŷ=	Похибка моделі	среднеквадра	Прогноз Ŷ=	Похибка моделі
49,28	778949,28	0,00	0,00	778949,28	0,00
63,44	778949,28	92014,16	0,01	778949,28	92014,16
86,38	852560,61	21725,77	0,00	788150,70	86135,68
76,59	869941,22	-50164,63	0,00	796764,27	23012,32
63,37	829809,52	-62046,14	0,01	799065,50	-31302,13
61,21	780172,60	-41111,40	0,00	795935,29	-56874,08
68,55	747283,49	16185,06	0,00	790247,88	-26779,33
72,86	760231,54	-28358,68	0,00	787569,95	-55697,08
40,80	737544,60	116396,20	0,02	782000,24	71940,56
62,73	830661,56	130001,18	0,02	789194,29	171468,44

Рисунок 4.8 — Процес розрахунку похибки моделі

=(E6^2)/(D6^2)

	D	E	F	G
	Експонентна модель 1			Експонентна моде
	k=	Точность	0,01%	k=
	0,8			0,1
. грн.)	Прогноз Ŷ=	Похибка моделі	среднеквадра	Прогноз Ŷ=
949,28	778949,28	0,00	0,00	778949,28
963,44	778949,28	92014,16	0,01	778949,28
286,38	852560,61	21725,77	0,00	788150,70
776,59	869941,22	-50164,63	0,00	796764,27
763,37	829809,52	-62046,14	0,01	799065,50
061,21	780172,60	-41111,40	0,00	795935,29
468,55	747283,49	16185,06	0,00	790247,88

Рисунок 4.9 — Процес розрахунку відхилення

Середньоквадратичне відхилення = середнє значення квадратичних відхилень за весь аналізований період.

Точність прогнозу = 1 - Середньоквадратичне відхилення

=(H6^2)/(G6^2)

	D	E	F	G	H	I
	Експонентна модель 1			Експонентна модель 2		
	k=	Точность	0,01%	k=	Точность	2,52%
	0,8			0,1		
. грн.)	Прогноз Ŷ=	Похибка моделі	среднеквадра	Прогноз Ŷ=	Похибка моделі	среднеквадрат
949,28	778949,28	0,00	0,00	778949,28	0,00	0,00
963,44	778949,28	92014,16	0,01	778949,28	92014,16	0,01
286,38	852560,61	21725,77	0,00	788150,70	86135,68	0,01
776,59	869941,22	-50164,63	0,00	796764,27	23012,32	0,00
763,37	829809,52	-62046,14	0,01	799065,50	-31302,13	0,00

Рисунок 4.10 — Процес розрахунку точності прогнозу

Розрахувавши точність прогнозу для моделей з k = 0.8 і k = 0.1 видно, що точність моделі 1 = 99% вище, ніж точність моделі 2 = 96%, отже для цього ряду з двох коефіцієнтів оптимальніше для прогнозу буде використовувати k = 0.8.

Для оцінки оптимального значення k послідовно обчислюються прогнози при k, що дорівнює 0,1; 0,2; 0,3; ... 0,9 і вибирається k, значення точності прогнозу, якого буде найближче до 100%.

Метод Холта використовується для прогнозування часових рядів, значення яких мають висхідну або низхідну тенденцію. Метод Холта також застосовується до рядів, дані для яких не є повним циклом і для яких сезонність ще не визначена (наприклад, неповні роки для помісячного прогнозування).

Якщо часовий ряд має висхідну або низхідну тенденцію, тенденцію слід визначати одночасно з поточним рівнем ряду (наприклад, просте експоненціальне згладжування). Для контролю рівня та нахилу модель Холта вводить два коефіцієнти згладжування — коефіцієнт згладжування ряду та коефіцієнт згладжування тренду.

Розраховуємо експоненціально-згладжений ряд:

$$L_t = k * Y_t + (1-k) * (L_{t-1} - T_{t-1}), \quad (4.2)$$

де L_t — згладжена величина на поточний період;

k — коефіцієнт згладжування ряду;

Y_t — поточні значення ряду (наприклад, обсяг продажів);

L_{t-1} — згладжена величина за попередній період;

T_{t-1} — значення тренда за попередній період;

L_t (Згладжена величина поточний період) = k (коефіцієнт згладжування ряду) * Y_t (поточні значення ряду (наприклад, обсяг продажів)) + (1-коефіцієнт згладжування ряду) * (L_{t-1}(згладжена величина за попередній період) —T_{t-1} (тренд за попередній період)

Коефіцієнт згладжування ряду k знаходиться в діапазоні від 0 до 1.

Для першого періоду на початку даних експоненціально-згладжений ряд дорівнює першому значенню ряду $L_1 = Y_1$;

Використовуючи Excel здійснимо розрахунок. Вводимо значення L першого періоду і розраховуємо згладжену величину для всього масиву даних:

E10 =$E$2*(D10-D9)+(1-$E$2)*E9

	B	C	D	E	U
			k=	0,60	
			b=	0,70	
	Дата	Середнє за пері	Lt = експоненційно згладжений ряд	Tt = значення тренду	
3	січень	5,05	5,05	0	
	лютий	5,05	5,05	0	
	березень	4,85137273	4,930823638	-0,083423453	
	квітень	4,84245484	4,911171741	-0,038783364	
	травень	4,84454828	4,88671101	-0,028757521	
	червень	4,85296333	4,87796541	-0,014749176	
	липень	5,04698387	4,985276156	0,07069277	
	серпень	6,02674333	5,581879353	0,438830068	
	вересень	7,57456452	6,601958426	0,845704372	
	жовтень	7,7	6,922501622	0,478091549	

Рисунок 4.11 — Процес розрахунку методом Хольта

Визначаємо значення тренда:

$$T_t = b * (L_t - L_{t-1}) + (1 - b) * T_{t-1}, \quad (4.3)$$

де T_t — значення тренда на поточний період;

b — коефіцієнт згладжування тренда;

L_t — експоненціально згладжена величина за поточний період;

L_{t-1} — експоненціально згладжена величина за попередній період;

T_{t-1} — значення тренда за попередній період.

T_t (значення тренда на поточний період) = b (коефіцієнт згладжування тренда) * (L_t(експоненціально згладжена величина за поточний період) — L_{t-1} (експоненціально згладжена величина за попередній період)) + (1-b (коефіцієнт згладжування тренда)) * T_{t-1} (значення тренда за попередній період)

Коефіцієнт згладжування тренда b знаходиться в діапазоні від 0 до 1.

Значення тренда на першому періоді дорівнює 0 ($T_1 = 0$);

Результати представлено у таблиці Б.1 додаток Б.

Робимо прогноз за методом Хольта.

Прогноз на p періодів вперед дорівнює:

$$\hat{Y}_{t+p} = L_t + p * T_t, \qquad (4.4)$$

де $\hat{Y}_{t+p}$ — прогноз по методу Хольта на p період;

L_t — експоненціально згладжена величина за останній період;

p — порядковий номер періоду, на який робимо прогноз;

T_t — тренд за останній період.

$\hat{Y}_{t+p}$ (Прогноз по методу Хольта) = L_t (експоненціально згладжена величина за останній період) + p (кількість періодів вперед, на який робимо прогноз) * T_t (тренд за останній період)

Здійснимо прогноз на три місяці. Для цього заповнимо номера періодів, на 3 місяці вперед.

Вводимо формулу прогнозу в клітинку. Для цього до експоненціального ряду за останній період додаємо значення тренда за останній період, помножене на номер періоду для прогнозу. Також зафіксуємо посилання на останнє значення експоненціально згладженого ряду і тренда — для цього виділяємо

посилання і натискаємо F4.

			b=	0,70
год	Дата	Середнє за пері	Lt = експоненційно згладжений ряд	Tt = значення тренду
	липень	23,33553457	22,61244937	0,701324165
	серпень	23,491182	22,85915928	0,383094188
	вересень	24,27320752	23,55435055	0,601562144
	жовтень	26,41259283	25,02867106	1,212493001
	листопад	26,35716371	25,34076945	0,582216773
	грудень	25,59542986	25,26067899	0,118601708
2021	січень	25,2198593	25,18874649	-0,014772234
	лютий	24,9501771	25,05151375	-0,100494589
	березень	24,81979723	24,95268167	-0,09933083
	квітень	25,09075858	25,07526015	0,056005684
	травень	26,26321467	25,76563059	0,500061012
	червень	25,74917219	25,55573114	0,003088693
	липень	25,70562786	25,6444337	0,063018394
	серпень	26,21993977	25,96452998	0,242972919
	вересень	27,245807	26,63610703	0,542995806
	жовтень	27,05494414	26,67021097	0,186771504
	листопад	26,99483071	26,79027421	0,14007572
	грудень	26,8452706	26,76724176	0,025899997
2022	січень	26.5238698	26.61085858	-0.101698222

Рисунок 4.12 — Прогноз на три місяці

Прогнози Хольта слід перераховувати, коли з'являються нові дані. Крім того, при підготовці прогнозних даних завжди слід очищати дані від факторів, які не повторюватимуться протягом прогнозованого періоду (наприклад, зростання продажів від рекламних заходів), або враховувати заплановані фактори, які призведуть до додаткового зростання продажів (наприклад, введення продукту в роздрібну мережу).

=$D$112+F113*$E$112

	D	E	F	G	H	I	J
	k=	0,60					
	b=	0,70			Для оцінки точності прогнозу		
ері	Lt = експоненційно згладжений ряд	Tt = значення тренду	p - номер періода для прогнозу	Прогноз по методу Хольта Ŷt+p = Lt + p *Tt	Прогноз на 1 період для аналізу	Помилка моделі	відхилення помилки моделі від прогнозної моделі
786	25,6444337	0,063018394			25,55881984	0,146808023	3,26169E-05
977	25,96452998	0,242972919			25,70745209	0,512487679	0,000382035
807	26,63610703	0,542995806			26,2075029	1,038304098	0,001452278
414	26,67021097	0,186771504			27,17910283	-0,124158691	2,10601E-05
071	26,79027421	0,14007572			26,85698248	0,137848234	2,6076E-05
706	26,76724176	0,025899997			26,93034993	-0,085079333	1,00441E-05
598	26,61085858	-0,101698222			26,79314175	-0,269271954	0,000103064
			1	26,50916036			
			2	26,40746214			
			3	26,30576392			

Рисунок 4.13 — Перерахунок прогнозу Хольта

Далі виконаємо оцінку точності моделі Хольта і підберемо оптимальні коефіцієнти згладжування ряду і тренда.

Розраховуємо прогноз на 1 період вперед для кожного місяця, коли продажі відомі.

Прогноз на 1 період для аналізу = значення експоненціально-згладженого ряду за попередній період + значення тренда за попередній період. (Значення тренда не множимо на p, тому що прогноз робимо на 1 період, а в цьому випадку p = 1).

D	E	M	N	O	P	Q
k=	0,60					
b=	0,70	Для оцінки точності прогнозу на 6 періодів				
Lt = поненційно джений ряд	Tt = значення тренду	Прогноз на 3 період для аналізу	Помилка моделі	відхилення помилки моделі від прогнозної моделі	Точність прогнозу	Ŷt+p Ряд с моделлю
5,05	0				98,83%	5,05
5,05	0					5,05
4,930823638	-0,083423453					5,05
4,911171741	-0,038783364					4,847400185
4,88671101	-0,028757521	5,05	-0,20545172	0,001798511		5,05
4,87796541	-0,014749176	4,680553278	0,172410052	0,001262151		4,680553278
4,985276156	0,07069277	4,794821648	0,252162222	0,002496297		4,794821648
5,581879353	0,438830068	4,800438448	1,226304882	0,041402973		4,800438448
6,601958426	0,845704372	4,833717882	2,740846638	0,130934512		4,833717882
6,922501622	0,478091549	5,197354465	2,502645535	0,105637286		5,197354465

Рисунок 4.14 — Процес розрахунку прогнозу на 1 період вперед

Розрахуємо помилку моделі = з фактичних даних віднімаємо прогноз на цей період.

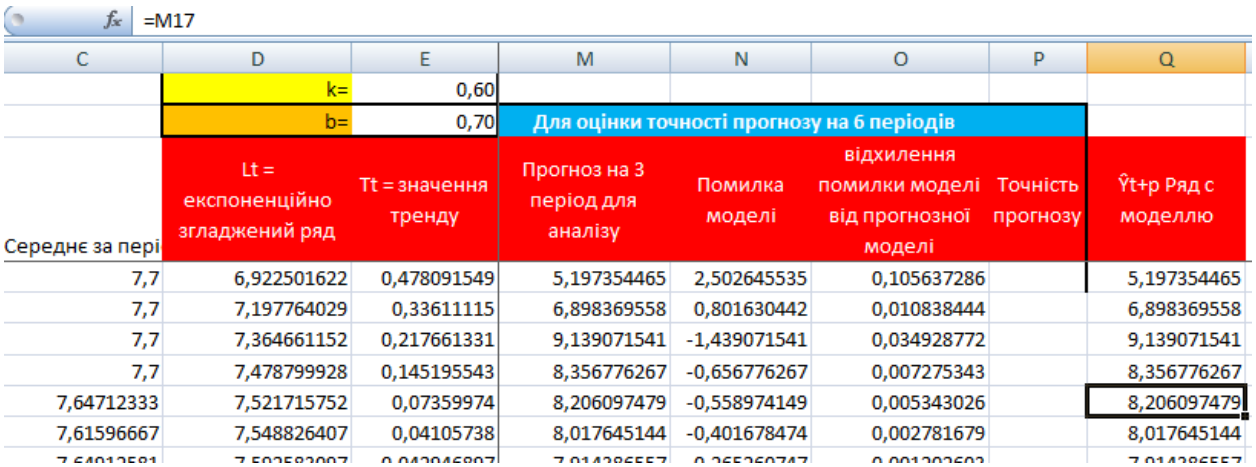
=1-СРЗНАЧ(O4:O112)

	D	E	M	N	O	P	Q
	k=	0,60					
	b=	0,70	Для оцінки точності прогнозу на 6 періодів				
epi	Lt = експоненційно згладжений ряд	Tt = значення тренду	Прогноз на 3 період для аналізу	Помилка моделі	відхилення помилки моделі від прогнозної моделі	Точність прогнозу	Ŷt+p Ряд с моделлю
,05	5,05	0				98,83%	5,05
,05	5,05	0					5,05
273	4,930823638	-0,083423453					5,05
484	4,911171741	-0,038783364					4,847400185
328	4,88671101	-0,028757521	5,05	-0,20545172	0,001798511		5,05

Рисунок 4.15 — Процес розрахунку помилки моделі Хольта

Розрахуємо точність прогнозу.

Для цього визначимо відхилення помилки моделі від прогнозної моделі = відношення помилки моделі в квадраті до фактичного значення в квадраті.



=M17

C	D	E	M	N	O	P	Q
	k=	0,60					
	b=	0,70	Для оцінки точності прогнозу на 6 періодів				
Середнє за пері	Lt = експоненційно згладжений ряд	Tt = значення тренду	Прогноз на 3 період для аналізу	Помилка моделі	відхилення помилки моделі від прогнозної моделі	Точність прогнозу	Ŷt+p Ряд с моделлю
7,7	6,922501622	0,478091549	5,197354465	2,502645535	0,105637286		5,197354465
7,7	7,197764029	0,33611115	6,898369558	0,801630442	0,010838444		6,898369558
7,7	7,364661152	0,217661331	9,139071541	-1,439071541	0,034928772		9,139071541
7,7	7,478799928	0,145195543	8,356776267	-0,656776267	0,007275343		8,356776267
7,64712333	7,521715752	0,07359974	8,206097479	-0,558974149	0,005343026		8,206097479
7,61596667	7,548826407	0,04105738	8,017645144	-0,401678474	0,002781679		8,017645144

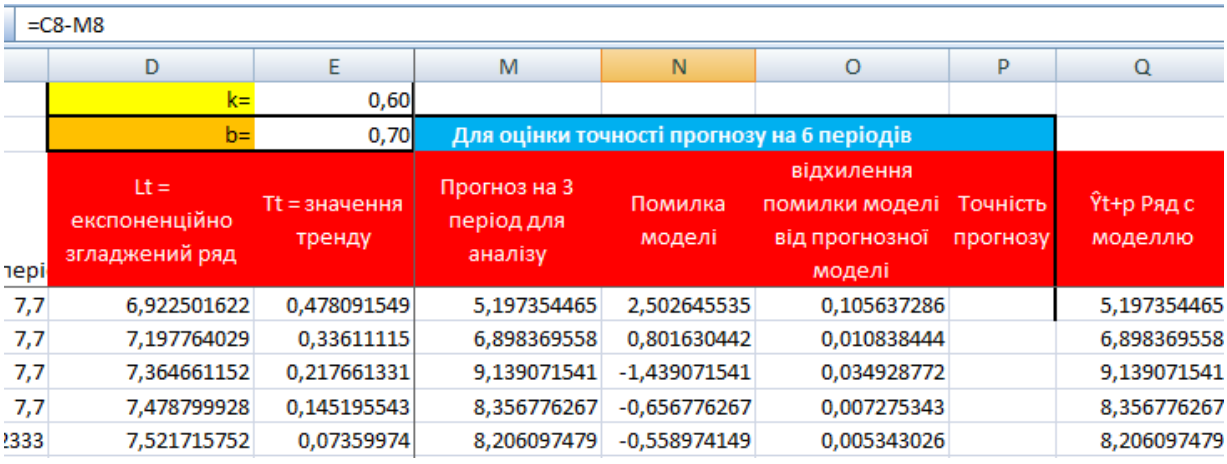
Рисунок 4.16 — Процес розрахунку точності прогнозу методом Хольта

Результати представлено у таблиці В.1 додаток В.

Для вибору послідовності та коефіцієнтів згладжування тренду k і b, які зроблять прогноз максимально точним, необхідно перебрати всі значення k і b послідовно за шкалою від 0 до 1 і знайти комбінацію значень, яка зробить прогноз максимально точним, якомога ближчим до 100%.

Здійснено прогноз на 3 періоди за такою формулою:

$$\hat{Y}_{t+3} = L_t + 3 * T_t \qquad (4.5)$$



=C8-M8

D	E	M	N	O	P	Q
k=	0,60					
b=	0,70	Для оцінки точності прогнозу на 6 періодів				
Lt = експоненційно згладжений ряд	Tt = значення тренду	Прогноз на 3 період для аналізу	Помилка моделі	відхилення помилки моделі від прогнозної моделі	Точність прогнозу	Ŷt+p Ряд с моделлю
6,922501622	0,478091549	5,197354465	2,502645535	0,105637286		5,197354465
7,197764029	0,33611115	6,898369558	0,801630442	0,010838444		6,898369558
7,364661152	0,217661331	9,139071541	-1,439071541	0,034928772		9,139071541
7,478799928	0,145195543	8,356776267	-0,656776267	0,007275343		8,356776267
7,521715752	0,07359974	8,206097479	-0,558974149	0,005343026		8,206097479

Рисунок 4.17 — Прогноз на 3 періоди

Далі визначаємо помилку моделі для кожного періоду, середньоквадратичне відхилення і точність прогнозу, як було описано вище.

Послідовно перебираючи коефіцієнти k і b від 0 до 1 з кроком 0,1, знаходимо таке поєднання, яке дає максимальне значення точності прогнозу.

Результат моделювання виводимо графічно рис. 4.18.

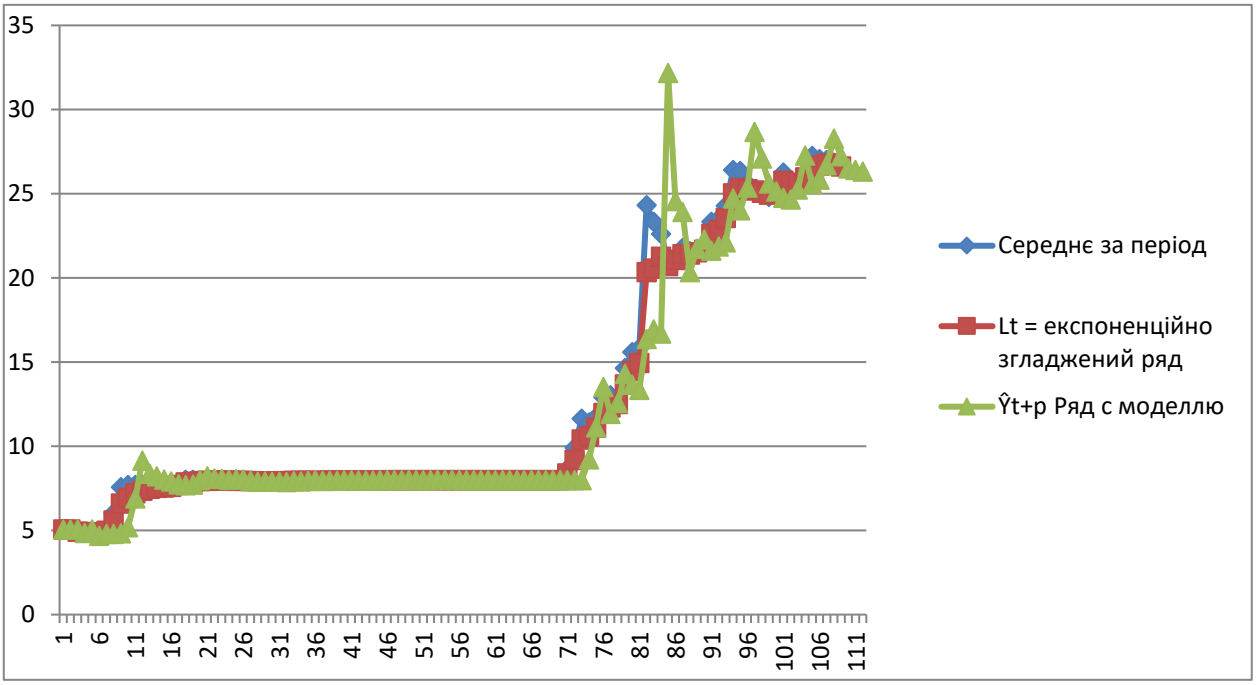


Рисунок 4.18 — Результати прогнозування методом Хольта

Вінтерс розвинув модель експоненціального згладжування з трендом Хольта і додав в неї сезонність. Перевага цього методу полягає в тому, що він дозволяє робити довгострокові прогнози. Однак, щоб зробити прогноз, наприклад, на 1 рік, вам потрібні дані щонайменше за 2 роки, а краще за 3-5 років.

Метод Вінтерса для прогнозування часових рядів з трендами та сезонністю в структурах даних.

Розраховуємо експоненціально-згладжений ряд:

$$L_t = k * Y_t / S_{ts} + (1-k) * (L_{t-1} + T_{t-1}) \quad (4.6)$$

Визначаємо значення тренда:

$$T_t = b * (L_t - L_{t-1}) + (1-b) * T_{t-1} \quad (4.7)$$

Оцінюємо сезонність:

$$S_t = q * Y_t / L_t + (1-q) * S_{ts} \quad (4.8)$$

Робимо прогноз:

$$\hat{Y}_{t+p} = (L_t + p * T_t) * S_{t-s} + p \quad (4.9)$$

Розраховуємо експоненціально-згладжений ряд:

$$L_t = k * Y_t / S_{ts} + (1-k) * (L_{t-1} + T_{t-1})$$

де L_t — згладжена величина на поточний період;

k — коефіцієнт згладжування ряду;

S_{ts} — коефіцієнт сезонності попереднього періоду;

Y_t — поточне значення ряду (наприклад, обсяг продажів);

L_{t-1} — згладжена величина за попередній період;

T_{t-1} — значення тренда за попередній період;

L_t (Згладжена величина поточний період) = k (коефіцієнт згладжування ряду) * Y_t (поточне значення ряду (наприклад, обсяг продажів)) / S_{ts} (коефіцієнт сезонності за цей же період в попередньому сезоні)) + (1—коефіцієнт згладжування ряду) * (L_{t-1} (згладжена величина за попередній період) -T_{t-1} (тренд за попередній період)

Коефіцієнт згладжування ряду k задається в діапазоні від 0 до 1.

Для першого періоду на початку даних експоненціально-згладжений ряд дорівнює першому значенню ряду L1 = Y1;

Сезонність в першому і другому періоді S_{ts} дорівнює 1.

У файлі Excel вводимо значення L:

D6 =$D$3*C6/$F$1+(1-$D$3)*(D5+E5)

A	B	C	D	E	F
		Коефіцієнт сезонності для 1-го року			1
			k=	B=	q=
			0,5	0,9	0,5
Год	Месяц	номер	Lt = експоненційно згладжений ряд	Tt = значення тренду	St-s — коефіцієнт сезонності попереднього періоду;
2016	січень	70	70		1
2017	лютий	43	56,5	-12,15	1
2017	березень	52	48,175	-8,7075	1
2017	квітень	69	54,23375	4,582125	1
2017	травень	50	54,4079375	0,61498125	1
2017	червень	67	61,01145938	6,004667813	1
2017	липень	69	68,00806359	6,897410578	1

Рисунок 4.19 — Процес розрахунку методом Вінтерса

Визначаємо значення тренда:

$$T_t = b * (L_t - L_{t-1}) + (1-b) * T_{t-1}$$

де T_t — значення тренда на поточний період;

b — коефіцієнт згладжування тренда;

L_t — експоненціально згладжена величина за поточний період;

L_{t-1} — експоненціально згладжена величина за попередній період;

T_{t-1} — значення тренда за попередній період.

T_t (значення тренда на поточний період) = b (коефіцієнт згладжування тренда) * (L_t(експоненціально згладжена величина за поточний період) - L_{t-1} експоненціально згладжена величина за попередній період)) + (1-b (коефіцієнт згладжування тренда)) * T_{t-1} (значення тренда за попередній період)

Коефіцієнт згладжування тренда b знаходиться в діапазоні від 0 до 1

Значення тренда для першого періоду дорівнює 0 (T1 = 0);

У доданому файлі розрахуємо значення тренда.

Оцінюємо сезонність:

$$S_t = q * Y_t / L_t + (1-q) * S_{ts}$$

де S_t — коефіцієнт сезонності для поточного періоду;

q — коефіцієнт згладжування сезонності;

Y_t — поточне значення ряду (наприклад, обсяг продажів));

L_t — згладжена величина за поточний період;

S_{ts} — коефіцієнт сезонності за цей же період в попередньому сезоні;

S_t (коефіцієнт сезонності для поточного періоду) = q (коефіцієнт згладжування сезонності) * Y_t(поточне значення ряду (наприклад, обсяг продажів)) / L_t (Згладжена величина за поточний період) + (1-q (коефіцієнт згладжування сезонності) *) * S_{ts} (коефіцієнт сезонності за цей же період в попередньому сезоні)

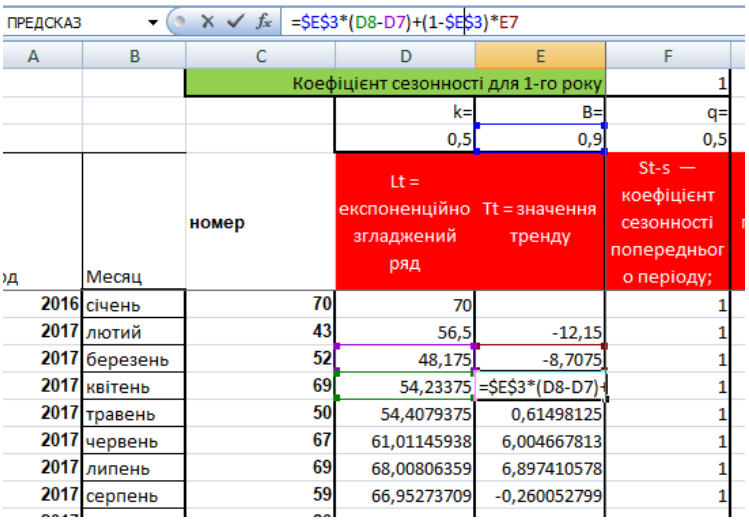


Рисунок 4.20 — Визначення значення тренду

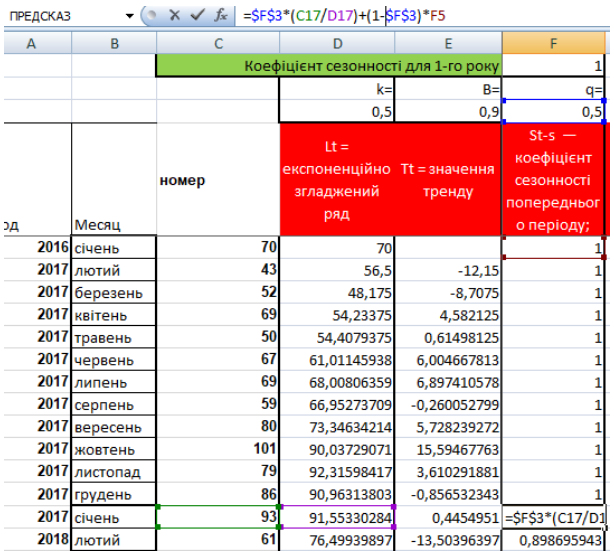


Рисунок 4.21 — Процес розрахунку сезонності

Коефіцієнти сезонності для першого сезону (року) = 1.

Зробимо прогноз по методу Вінтерса.

Прогноз на p періодів вперед дорівнює:

$$\hat{Y}_{t+p} = (L_t + p * T_t) * S_{t-s} + p$$

де $\hat{Y}_{t+p}$ — прогноз по методу Вінтерса на p періодів вперед;

L_t — експоненціально згладжена величина за останній період;

p — порядковий номер періоду, на який робимо прогноз;

T_t — тренд за останній період;

$S_{t-s} + p$ — коефіцієнт сезонності за цей же період в останньому сезоні;

$\hat{Y}_{t+p}$ (Прогноз по методу Вінтерса) = (L_t (експоненціально згладжена величина за останній період) + p (кількість періодів вперед, на яке робимо прогноз) * T_t (тренд за останній період)) * $S_{t-s} + p$ (коефіцієнт сезонності за цей же період в останньому сезоні).

Зробимо прогноз на 12 місяців вперед. Для цього заповнимо номера періодів, на скільки будемо робити прогноз

Tt = значення тренду	St-s — коефіцієнт сезонності попереднього періоду;	p - номер періода для прогнозу	Прогноз по методу Хольта Ŷt+p = Lt + p *Tt	Проноз для оцінки моделі	По мо
7,005812402	1,023088141			74,17847007	27,8
-3,090173369	0,798063037			85,11377469	-19,1
		1	52,7932058	101,2795319	-10,2
		2	37,29717629	0	
		3	11,78487045	0	
		4	-5,534481564	0	
		5	-23,03511119	0	
		6	-44,31179636	0	
		7	-54,33963713	0	
		8	-66,18289499	0	
		9	-91,12820373	0	
		10	-97,78640041	0	
		11	-128,9422647	0	
		12	-114,2008517	0	
				0	

Рисунок 4.22 — Прогноз на 12 місяців

Вводимо формулу прогнозу в клітинку. Для цього суму значень експоненціального ряду і тренда за останній період, помножене на номер періоду для прогнозу, множимо на коефіцієнт сезонності.

Щоб простягнути формулу прогнозу на 12 періодів вперед, зафіксуємо посилання на експонентний ряд і значення тренда за останній період — для цього виділяємо посилання і натискаємо F4:

G	H	I	J	K
р - номер періода для прогнозу	Прогноз по методу Хольта Ŷt+p = Lt + p *Tt	Проноз для оцінки моделі	Похибка моделі	відхилення помилки моделі від прогнозної моделі
		74,17847007	27,82152993	0,074398071
		85,11377469	-19,11377469	0,083869693
1	52,7932058	101,2795319	-10,27953194	0,012760388
2	37,29717629	0	103	1
3	11,78487045	0	97	1
4	-5,534481564	0	79	1
5	-23,03511119	0	93	1
6	-44,31179636	0	74	1
7	-54,33963713	0	103	1
8	-66,18289499	0	96	1
9	-91,12820373	0	58	1
10	-97,78640041	0	76	1
11	-128,9422647	0	94	1
12	-114,2008517	0	56	1
		0	77	

Рисунок 4.23 — Процес розрахунку прогнозу на 12 періодів вперед

Коли з'являються нові дані, доцільно перерахувати прогноз Winters для уточнення рядів, трендів і сезонності. Крім того, при підготовці даних для прогнозу завжди слід вилучати фактори, які не повторюватимуться протягом прогнозованого періоду, або враховувати заплановані фактори, які призведуть до додаткового зростання продажів.

Оцінимо точність моделі Вінтерса і підібремо оптимальні коефіцієнти згладжування для ряду, тренда і сезонності.

Розраховуємо прогноз на 1 період вперед для кожного місяця.

Прогнозні значення, використані для оцінок моделі для першого та другого року (сезонні) = експоненціально згладжені значення ряду за попередній період + значення тренду за попередній період (ми не множимо значення тренду на p, оскільки ми прогнозуємо на один період, і в цьому випадку p = 1).

- номер ріода для прогнозу	Прогноз по методу Хольта Ŷt+p = Lt + p *Tt	Проноз для оцінки моделі	Похибка моделі	відхилення помилки моделі від прогнозної моделі	Точність прогнозу
		70	0	0	0,719518
		70	-27	0,394267171	
		44,35	7,65	0,021642936	
		39,4675	29,5325	0,183190203	
		58,815875	-8,815875	0,031087861	
		55,02291875	11,97708125	0,031955998	
		67,01612719	1,983872813	0,000826665	
		74,90547417	-15,90547417	0,072675699	
		66,69268429	13,30731571	0,027669477	
		79,07458142	21,92541858	0,047125182	
		105,6319683	-26,63196834	0,113645528	
		95,92627605	-9,926276052	0,013322195	
		90,10660568	2,893394317	0,000967942	
		91,99879794	-30,99879794	0,258243879	

Рисунок 4.24 — Процес розрахунку моделі для першого та другого року

Прогноз для третього року (сезону) = (значення експоненціально-згладженого ряду за попередній період + значення тренда за попередній період) помножити на коефіцієнт сезонності цього періоду в попередньому кварталі.

			k=	B=	q=	
			0,5	0,9	0,5	
од	Месяц	номер	Lt = експоненційно згладжений ряд	Tt = значення тренду	St-s — коефіцієнт сезонності попереднього періоду;	p - но період прогн
2020	жовтень	97	83,88182436	10,76812529	1,044503026	
2020	листопад	102	100,9390189	16,4282876	0,980877203	
2020	грудень	80	98,61713781	-0,446864203	0,906441841	
2020	січень	64	78,64277995	-18,02160849	0,948218364	
2021	лютий	53	61,68393397	-17,06512224	0,851942426	
2021	березень	65	51,64582503	-10,74081026	1,183205105	
2021	квітень	86	56,62962389	3,411337949	1,353618226	
2021	травень	83	69,45982607	11,88831575	1,123592018	
2021	червень	63	77,06493055	8,033425609	0,84154722	
2021	липень	66	73,83476405	-2,10380729	0,974343359	
2021	серпень	97	83,57734682	8,557943765	1,088560067	
2021	вересень	104	104,6601445	19,83031226	0,940589053	
2021	жовтень	69	95,27529064	-6,463337225	0,884360091	
2021	листопад	91	90,79302684	-4,680371139	0,991578409	
2021	грудень	75	84,42688058	-6,197568746	0,897392253	
2021	січень	102	92,89973533	7,005812402	1,023088141	

Рисунок 4.25 — Процес розрахунку прогнозу для третього року

Розрахуємо помилку моделі = з фактичних даних віднімаємо прогноз на цей період.

д	Месяц	номер	Lt = експоненційно згладжений ряд	Tt = значення тренду	St-s — коефіцієнт сезонності попереднього періоду;	p - номер періода для прогнозу	Прогноз по методу Хольта Ŷt+p = Lt + p *Tt	Проноз для оцінки моделі
2021	квітень	86	56,62962389	3,411337949	1,353618226			48,6195641
2021	травень	83	69,45982607	11,88831575	1,123592018			63,17802456
2021	червень	63	77,06493055	8,033425609	0,84154722			70,41511227
2021	липень	66	73,83476405	-2,10380729	0,974343359			89,76164793
2021	серпень	97	83,57734682	8,557943765	1,088560067			72,91584917
2021	вересень	104	104,6601445	19,83031226	0,940589053			81,76874476
2021	жовтень	69	95,27529064	-6,463337225	0,884360091			130,0306588
2021	листопад	91	90,79302684	-4,680371139	0,991578409			87,11362047
2021	грудень	75	84,42688058	-6,197568746	0,897392253			78,05611421
2021	січень	102	92,89973533	7,005812402	1,023088141			74,17847007
2022	лютий	66	88,68778577	-3,090173369	0,798063037			85,11377469
2022	березень	91				1	52,7932058	101,2795319
2022	квітень	103				2	37,29717629	0

Рисунок 4.26 — Процес розрахунку помилки моделі

Визначимо відхилення помилки моделі від прогнозної моделі = Ставлення помилки моделі в квадраті до фактичного значення в квадраті.

ПРЕДСКАЗ =(J9^2)/(C9^2)

	A	B	C	D	E	F	G	H	I	J	K	L
			Коефіцієнт сезонності для 1-го року			1						
				k=	B=	q=						
				0,5	0,9	0,5						
	Год	Месяц	номер	Lt = експоненційно згладжений ряд	Tt = значення тренду	St-s — коефіцієнт сезонності попереднього періоду;	p - номер періода для прогнозу	Прогноз по методу Хольта Ŷt+p = Lt + p *Tt	Проноз для оцінки моделі	Похибка моделі	відхилення помилки моделі від прогнозної моделі	Точність прогнозу
	2016	січень	70	70		1			70	0	0	0,719518
	2017	лютий	43	56,5	-12,15	1			70	-27	0,394267171	
	2017	березень	52	48,175	-8,7075	1			44,35	7,65	0,021642936	
	2017	квітень	69	54,23375	4,582125	1			39,4675	29,5325	0,183190203	
	2017	травень	50	54,4079375	0,61498125	1			58,815875	-8,815875	=(J9^2)/(C9^2)	
	2017	червень	67	61,01145938	6,004667813	1			55,02291875	11,97708125	0,031955998	
	2017	липень	69	68,00806359	6,897410578	1			67,01612719	1,983872813	0,000826665	

Рисунок 4.27 — Відхилення помилки моделі від прогнозної моделі

Розрахуємо точність прогнозу = одиниця мінус середнє значення відхилень.

Для вибору коефіцієнтів згладжування ряду, тренду та сезонності k, b та q, які зроблять прогноз максимально точним, необхідно послідовно перебрати всі значення k, b та q на шкалі від 0 до 1 та знайти комбінацію значень, яка зробить прогноз максимально точним, максимально наближеним до 100%.

p - номер періода для прогнозу	Прогноз по методу Хольта Ŷt+p = Lt + p *Tt	Проноз для оцінки моделі	Похибка моделі	відхилення помилки моделі від прогнозної моделі	Точність прогнозу
		70	0	0	=1-СРЗНА
		70	-27	0,394267171	
		44,35	7,65	0,021642936	
		39,4675	29,5325	0,183190203	
		58,815875	-8,815875	0,031087861	
		55,02291875	11,97708125	0,031955998	

Рисунок 4.28 — Процес розрахунку точності прогнозу методом Вінтерса

Результати розрахунку наведено у таблиці Г.1 додаток Г.

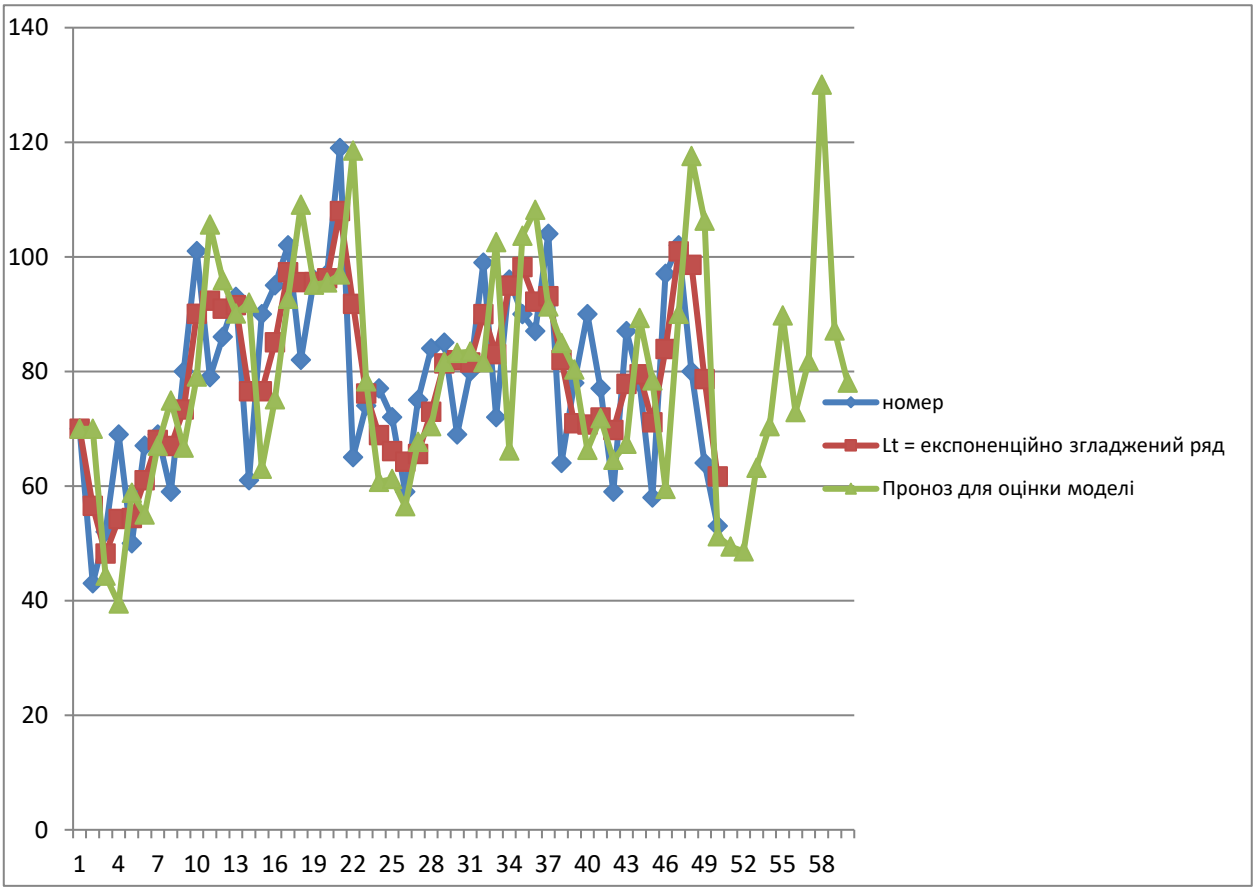


Рисунок 4.29 — Результати прогнозування методом Вінтерса

Далі проведемо аналіз можливостей застосування розроблених веб-додатків у різних галузях економіки та фінансів. Перспективи подальшого розвитку розроблених веб-додатків та їх застосування.

Отримані в кваліфікаційній роботі теоретичні результати являють собою сукупність знань, що лежать в основі методологічної бази побудови прикладних

моделей для дослідження динамічних ефектів в розвитку сучасних економічних процесів. Ця база забезпечує формалізацію і рішення нових прикладних задач, які раніше не цікавили вітчизняну економічну науку. Перш за все, це завдання по дослідженню стабільності економічних процесів, імітаційного моделювання динаміки поведінки мікрооб'єктів в ринковій економіці, обґрунтування надійності економічних регуляторів.

Проведемо порівняння розробленого веб-додатку та розглянутих методів прогнозування.

Таблиця 4.1 — Порівняльна характеристика

Показник	Веб-додаток для прогнозування	Модель Брауна	Модель Хольта	Модель Вінтерса
Швидкодія, с	1,5	8	7	7
Точність прогнозування	98,5	93,6	90,2	88,4
Точність аналізу, %	99,6	97,2	92,5	89,9
Побудова тренду	-	+	+	+
Виділення корозмірності	+	-	-	-
Застосування машинного навчання	+	-	-	-

Таким чином, провівши порівняльний аналіз розглянутих підходів, є можливість стверджувати, що розроблений веб-додаток є кращим за точністю аналізу та точністю прогнозування, що є головними показниками у розглянутому підході, а також швидкодія спрацювання перевищує значення майже у 4 рази. Враховуючи той факт, що при розробці було використано машинне навчання, яке на сьогодні є актуальним напрямком досліджень багатьох вчених, є можливість стверджувати про популярність розгляду даного веб-додатку серед науковців.

Розроблений веб-додаток є універсальним та може знайти своє застосування у прогнозуванні будь-яких фінансових показників підприємств діяльність яких направлена на отримання прибутку.

Щодо подальшого розвитку розробленого веб-додатку варто підкреслити,

що можливе розширення функціональної складової додатку у напрямку максимізації виконуваних функцій:

— порівняння прогнозного значення за роками;

— порівняння прогнозного значення у спектрі інших компаній;

— формування лінії тренду;

— виведення графічного результату та збереження його в окремий файл.

5 ЕКОНОМІЧНА ЧАСТИНА

Для успішного впровадження науково-технічної розробки критично важливо, щоб вона відповідала актуальним вимогам науково-технічного прогресу та ураховувала економічні аспекти. Надання оцінки економічної ефективності результатів науково-дослідної роботи є необхідною частиною цього процесу. Дослідження, яке представлене у магістерській роботі і приурочене розробці та вивченню “Методи машинного навчання для прогнозування у дата-майнінгу”, належить до науково-технічних робіт, спрямованих на виведення на ринок. Рішення про комерціалізацію розробки може бути прийняте протягом проведення самої роботи, відкриваючи перспективи подальшого виведення на ринок. Цей напрямок визначається як пріоритетний, оскільки розроблені результати можуть бути корисними для різних зацікавлених сторін і приносити економічні вигоди. Однак для успішної реалізації цього процесу ключовим є привертання зацікавленого інвестора, який виявить інтерес до втілення даного проекту, і переконання його у доцільності інвестування у цю розробку. Для досягнення цієї мети передбачені наступні етапи виконання робіт:

— проведення комерційного аудиту науково-технічної розробки, включаючи визначення науково-технічного рівня та комерційного потенціалу.

— розрахунок витрат на реалізацію науково-технічної розробки.

— проведення розрахунку економічної ефективності впровадження та комерціалізації науково-технічної розробки для потенційного інвестора, а також обґрунтування економічної доцільності комерціалізації з точки зору інвестора.

5.1 Проведення комерційного та технологічного аудиту науково-технічної розробки

Метою проведення комерційного і технологічного аудиту дослідження за темою “Методи машинного навчання для прогнозування у дата-майнінгу” є оцінювання комерційного потенціалу розробки, створеної в результаті науково-

технічної діяльності.

Оцінювання науково-технічного рівня розробки та її комерційного потенціалу рекомендується здійснювати із застосуванням 5-ти бальної системи оцінювання за 12-ма критеріями, наведеними в табл. 5.1 [40].

Таблиця 5.1 — Рекомендовані критерії оцінювання науково-технічного рівня і комерційного потенціалу розробки та бальна оцінка

Бали (за 5-ти бальною шкалою)					
	0	1	2	3	4
Технічна здійсненність концепції					
1	Достовірність концепції не підтверджена	Концепція підтверджена експертними висновками	Концепція підтверджена розрахунками	Концепція перевірена на практиці	Перевірено працездатність продукту в реальних
Ринкові переваги (недоліки)					
2	Багато аналогів на малому ринку	Мало аналогів на малому ринку	Кілька аналогів на великому ринку	Один аналог на великому ринку	Продукт не має аналогів на великому
3	Ціна продукту значно вища за ціни аналогів	Ціна продукту дещо вища за ціни аналогів	Ціна продукту приблизно дорівнює цінам аналогів	Ціна продукту дещо нижче за ціни аналогів	Ціна продукту значно нижче за ціни аналогів
4	Технічні та споживчі властивості продукту	Технічні та споживчі влас-тивості продукту	Технічні та споживчі властивості продукту на	Технічні та споживчі властивості продукту	Технічні та споживчі властивості продукту
5	Експлуатаційні витрати значно вищі, ніж в аналогів	Експлуатаційні витрати дещо вищі, ніж в аналогів	Експлуатаційні витрати на рівні експлуатаційних витрат аналогів	Експлуатаційні витрати трохи нижчі, ніж в аналогів	Експлуатаційні витрати значно нижчі, ніж в аналогів
Ринкові перспективи					
6	Ринок малий і не має позитивної динаміки	Ринок малий, але має позитивну динаміку	Середній ринок з позитивною динамікою	Великий стабільний ринок	Великий ринок з позитивною динамікою

Продовження таблиці 5.1

7	Активна конкуренція великих компаній на	Активна конкуренція	Помірна конкуренція	Незначна конкуренція	Конкурентів немає
8	Відсутні фахівці як з технічної, так і з комерційної	Необхідно наймати фахівців або витрачати	Необхідне незначне навчання фахівців та	Необхідне незначне навчання фахівців	Є фахівці з питань як з технічної, так і з комерційної
9	Потрібні значні фінансові ресурси, які	Потрібні незначні фінансові ресурси.	Потрібні значні фінансові ресурси. Джерела	Потрібні незначні фінансові ресурси.	Не потребує додаткового фінансування
10	Необхідна розробка нових матеріалів	Потрібні матеріали, що використовуют ь ся у	Потрібні дорогі матеріали	Потрібні досяжні та дешеві матеріали	Всі матеріали для реалізації ідеї відомі та давно
11	Термін реалізації ідеї більший за 10 років	Термін реалізації ідеї більший за 5 років. Термін	Термін реалізації ідеї від 3-х до 5-ти років. Термін	Термін реалізації ідеї менше 3-х років. Термін	Термін реалізації ідеї менше 3-х років. Термін
12	Необхідна розробка регламентних документів та отримання великої кількості дозвільних документів на виробництво та реалізацію продукту	Необхідно отримання великої кількості дозвільних документів на виробництво та реалізацію продукту, що вимагає значних коштів та часу	Процедура отримання дозвільних документів для виробництва та реалізації продукту вимагає незначних коштів та часу	Необхідно тільки повідомлення відповідним органам про виробництво та реалізацію продукту	Відсутні будь-які регламентні обмеження на виробництво та реалізацію продукту

Результати оцінювання науково-технічного рівня та комерційного потенціалу науково-технічної розробки потрібно звести до таблиці. Для оцінки науково-технічного рівня і комерційного потенціалу розробки експертами було запрошено трьох незалежних експертів Вінницького національного технічного університету: к.т.н. доц. кафедри ОТ Кадук Олександр Володимирович, к.т.н.

доц. кафедри ОТ Кожем’яко Андрій Вікторович, к.т.н. доц. кафедри ОТ Колесник Ірина Сергіївна.

Таблиця 5.2 — Результати оцінювання науково-технічного рівня і комерційного потенціалу розробки експертами

Критерії	Експерт (ПІБ, посада)		
	Кадук О.В.	Кожем’яко А.В.	Колесник І.С.
	Бали, виставлені експертами:		
1. Технічна здійсненність концепції	2	2	2
2. Ринкові переваги (наявність аналогів)	0	0	0
3. Ринкові переваги (ціна продукту)	4	3	4
4. Ринкові переваги (технічні властивості)	3	3	3
5. Ринкові переваги (експлуатаційні витрати)	4	4	3
6. Ринкові перспективи (розмір ринку)	2	2	2
7. Ринкові перспективи (конкуренція)	2	2	2
8. Практична здійсненність (наявність фахівців)	2	2	3
9. Практична здійсненність (наявність фінансів)	3	3	2
10. Практична здійсненність (необхідність нових матеріалів)	4	4	3
11. Практична здійсненність	3	4	3
12. Практична здійсненність	4	4	4
Сума балів	$СБ_1$=33	$СБ_2$=34	$СБ_3$=31
Середньоарифметична сума балів $СБ_с$	33		

За результатами розрахунків, наведених в таблиці 5.2, зробимо висновок щодо науково-технічного рівня і рівня комерційного потенціалу розробки. При цьому використаємо рекомендації, наведені в табл. 5.3 [40].

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою “Методи машинного навчання для прогнозування у дата-майнінгу” становить 33 бали, що, відповідно до таблиці 5.3 рівень комерційного потенціалу

розробки вище середнього, що свідчить про комерційну важливість проведення даних досліджень.

Таблиця 5.3 — Науково-технічні рівні та комерційні потенціали розробки

Середньоарифметична сума балів СБ , на основі висновків експертів	Науково-технічний рівень та комерційний потенціал розробки
41…48	Високий
31…40	Вище середнього
21…30	Середній
11…20	Нижче середнього
0…10	Низький

5.2 Визначення рівня конкурентоспроможності розробки

В процесі визначення економічної ефективності науково-технічної розробки також доцільно провести прогноз рівня її конкурентоспроможності за сукупністю параметрів, що підлягають оцінюванню.

В якості аналога для розробки було обрано STATISTICA. Основними недоліками аналога є потреба команди ситемних адміністраторів для контролювання роботи продукту. Також до недоліків можна віднести часткову відсутність системи одночасних рівнянь.

У розробці дана проблема вирішується наявністю повноцінного зв’язку з системами та автономністю роботи проекту без постійного адміністрування проекту.

Одиничний параметричний індекс розраховуємо за формулою [40]:

$$q_i = \frac{P_i}{P_{базі}}. \qquad (5.1)$$

де q_i — одиничний параметричний індекс, розрахований за i-м параметром;

P_i — значення i-го параметра виробу;

$P_{базі}$ — аналогічний параметр базового виробу-аналога, з яким проводиться порівняння.

Загальні технічні та економічні характеристики розробки представлено в таблиці 5.4.

Нормативні параметри оцінюємо показником, який отримує одне з двох значень: 1 — пристрій відповідає нормам і стандартам; 0 — не відповідає.

Груповий показник конкурентоспроможності за нормативними параметрами розраховуємо як добуток частинних показників за кожним параметром за формулою [40]:

$$I_{НП} = \prod_{i=1}^{n} q_i \,, \tag{5.2}$$

де $I_{нп}$ — загальний показник конкурентоспроможності за нормативними параметрами;

q_i — одиничний (частинний) показник за i-м нормативним параметром;

n — кількість нормативних параметрів, які підлягають оцінюванню.

За нормативними параметрами розроблюваний пристрій відповідає вимогам ДСТУ, тому $I_{нп} = 1$.

Таблиця 5.4 — Основні техніко-економічні показники аналога та розробки, що проектується

Показник	Варіанти		Відносний показник якості	Коефіцієнт вагомості параметра
	Базовий (товар-конкурент)	Новий (інноваційне рішення)		
1	2	3	4	5
Надійність, %	95	95	1	30 %
Сумісність, %	96	96	1	20 %
Супровід, %	90	91	1,01	20 %
Брендування, %	85	85	1	15 %
Система одночасних рівнянь, %	45	95	2,1	15 %

Значення групового параметричного індексу за технічними параметрами визначаємо з урахуванням вагомості (частки) кожного параметра [40]:

$$I_{ТП} = \sum_{i=1}^{n} q_i \cdot \alpha_i \ , \qquad (5.3)$$

де $I_{тп}$ — груповий параметричний індекс за технічними показниками (порівняно з виробом-аналогом);

q_i — одиничний параметричний показник i-го параметра;

α_i — вагомість i-го параметричного показника, $\sum_{i=1}^{n} \alpha_i = 1$;

n — кількість технічних параметрів, за якими оцінюється конкурентоспроможність.

Проведемо аналіз параметрів згідно даних таблиці 5.4.

$$I_{mn} = 1 \cdot 0{,}3 + 1 \cdot 0{,}2 + 1{,}01 \cdot 0{,}2 + 1 \cdot 0{,}15 + 2{,}1 \cdot 0{,}15 = 1{,}2.$$

Груповий параметричний індекс за економічними параметрами розраховуємо за формулою :

$$I_{ЕП} = \sum_{i=1}^{m} q_i \cdot \beta_i \ , \qquad (5.4)$$

де $I_{ЕП}$ — груповий параметричний індекс за економічними показниками;

q_i — економічний параметр i-го виду;

β_i — частка i-го економічного параметра, $\sum_{i=1}^{m} \beta_i = 1$;

m — кількість економічних параметрів, за якими здійснюється оцінювання.

Проведемо аналіз параметрів згідно даних таблиці .

$$I_{ЕП} = 0{,}75 \cdot 0{,}5 + 0{,}86 \cdot 0{,}5 = 0{,}80.$$

На основі групових параметричних індексів за нормативними, технічними та економічними показниками розрахуємо інтегральний показник конкурентоспроможності за формулою**:**

$$K_{ІНТ} = I_{НП} \cdot \frac{I_{ТП}}{I_{ЕП}}, \quad (5.5)$$

$$K_{ІНТ} = 1 \cdot 1{,}2 / 0{,}80 = 1{,}5.$$

Інтегральний показник конкурентоспроможності $К_{ІНТ} > 1$, отже розробка переважає відомі аналоги за своїми техніко-економічними показниками.

5.3 Розрахунок витрат на проведення науково-дослідної роботи

Витрати, пов’язані з проведенням науково-дослідної роботи на тему “Методи машинного навчання для прогнозування у дата-майнінгу”, під час планування, обліку і калькулювання собівартості науково-дослідної роботи групуємо за відповідними статтями.

5.3.1 Витрати на оплату праці

До статті «Витрати на оплату праці» належать витрати на виплату основної та додаткової заробітної плати керівникам відділів, лабораторій, секторів і груп, науковим, інженерно-технічним працівникам, конструкторам, технологам, креслярам, копіювальникам, лаборантам, робітникам, студентам, аспірантам та іншим працівникам, безпосередньо зайнятим виконанням конкретної теми, обчисленої за посадовими окладами, відрядними розцінками, тарифними ставками згідно з чинними в організаціях системами оплати праці.

Основна заробітна плата дослідників

Витрати на основну заробітну плату дослідників ($З_о$) розраховуємо у відповідності до посадових окладів працівників, за формулою [40]:

$$З_о = \sum_{i=1}^{k} \frac{M_{ni} \cdot t_i}{T_p}, \quad (5.6)$$

де k — кількість посад дослідників залучених до процесу досліджень;

M_{ni} — місячний посадовий оклад конкретного дослідника, грн;

t_i — число днів роботи конкретного дослідника, дн.;

T_p — середнє число робочих днів в місяці, T_p=21 дні.

$$З_о = 17500 \cdot 55 / 21 = 3977 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 5.5 — Витрати на заробітну плату дослідників

Найменування посади	Місячний посадовий оклад, грн	Оплата за робочий день, грн	Число днів роботи	Витрати на заробітну плату, грн
Керівник проекту	17500	795,5	5	3977
Інженер-дослідник (програміст)	16000	727,3	44	32000
Всього				35977

Основна заробітна плата робітників

Витрати на основну заробітну плату робітників ($З_р$) за відповідними найменуваннями робіт НДР на тему “Методи машинного навчання для прогнозування у дата-майнінгу” розраховуємо за формулою:

$$З_р = \sum_{i=1}^{n} C_i \cdot t_i , \quad (5.7)$$

де C_i — погодинна тарифна ставка робітника відповідного розряду, за виконану відповідну роботу, грн/год;

t_i — час роботи робітника при виконанні визначеної роботи, год.

Погодинну тарифну ставку робітника відповідного розряду C_i можна визначити за формулою:

$$C_i = \frac{M_M \cdot K_i \cdot K_c}{T_p \cdot t_{зм}} , \quad (5.8)$$

де $М_М$ — розмір прожиткового мінімуму працездатної особи, або мінімальної місячної заробітної плати (в залежності від діючого законодавства), приймемо *Мм*=6500 грн;

$К_і$ — коефіцієнт міжкваліфікаційного співвідношення для встановлення тарифної ставки робітнику відповідного розряду;

$К_с$ — мінімальний коефіцієнт співвідношень місячних тарифних ставок робітників першого розряду з нормальними умовами праці виробничих об'єднань і підприємств до законодавчо встановленого розміру мінімальної заробітної плати.

$Т_р$ — середнє число робочих днів в місяці, приблизно $Т_р$ = 21 дн;

$t_{зм}$ — тривалість зміни, год.

$$С_1 = 6500{,}00 \cdot 1 \cdot 1{,}65 / (21 \cdot 8) = 65{,}8 \text{ грн.}$$

$$З_{р1} = 65{,}8 \cdot 75 = 4935{,}3 \text{ грн.}$$

Таблиця 5.6 — Величина витрат на основну заробітну плату робітників

Найменування робіт	Тривалість роботи, год	Розряд роботи	Погодинна тарифна ставка, грн	Величина оплати на робітника грн
1.Підготовчі	75	1	65,8	4935,3
2. Тестувальні	50	1	65,8	3290,2
Всього				8225,4

Додаткова заробітна плата дослідників та робітників

Додаткову заробітну плату розраховуємо як 10 … 12% від суми основної заробітної плати дослідників та робітників за формулою:

$$З_{дод} = (З_о + З_р) \cdot \frac{Н_{дод}}{100\%}, \qquad (5.9)$$

де $Н_{дод}$ — норма нарахування додаткової заробітної плати. Приймемо 11%.

$$З_{дод} = (35977 + 8225{,}4) \cdot 11 / 100\% = 4862{,}3 \text{ грн.}$$

5.3.2 Відрахування на соціальні заходи

Нарахування на заробітну плату дослідників та робітників розраховуємо як 22% від суми основної та додаткової заробітної плати дослідників і робітників за формулою:

$$З_н = (З_о + З_р + З_{дод}) \cdot \frac{Н_{зп}}{100\%} \quad (5.10)$$

де $Н_{зп}$ — норма нарахування на заробітну плату. Приймаємо 22%.

$З_н$ = (35977+ 8225,4+4862,3) · 22 / 100% = 10794,3 грн.

5.3.3 Сировина та матеріали

До статті «Сировина та матеріали» належать витрати на сировину, основні та допоміжні матеріали, інструменти, пристрої та інші засоби і предмети праці, які придбані у сторонніх підприємств, установ і організацій та витрачені на проведення досліджень за темою “Методи машинного навчання для прогнозування у дата-майнінгу”.

Витрати на матеріали ($М$), у вартісному вираженні розраховуються окремо по кожному виду матеріалів за формулою:

$$М = \sum_{j=1}^{n} Н_j \cdot Ц_j \cdot К_j - \sum_{j=1}^{n} В_j \cdot Ц_{вj}, \quad (5.11)$$

де $Н_j$ — норма витрат матеріалу j-го найменування, кг;

n — кількість видів матеріалів;

$Ц_j$ — вартість матеріалу j-го найменування, грн/кг;

$К_j$ — коефіцієнт транспортних витрат, ($К_j$ = *1,1 … 1,15*);

$В_j$ — маса відходів j-го найменування, кг;

$Ц_{вj}$ — вартість відходів j-го найменування, грн/кг.

Проведені розрахунки зведемо до таблиці.

Таблиця 5.7 — Витрати на матеріали

Найменування матеріалу, марка, тип, сорт	Ціна за 1 кг, грн	Норма витрат, кг	Вартість витраченого матеріалу, грн
Папір канцелярський офісний (А4) BUROMAX Premium	174,9	1	174,9
Папір для заміток (А5) BUROMAX Premium Light	110	1	110
Начиння канцелярське BUROMAX Premium	190	1	190
Всього			474,9
З врахуванням коефіцієнта транспортування			522,39

5.3.4 Розрахунок витрат на комплектуючі

Витрати на комплектуючі *($К_в$)*, які використовують при проведенні НДР на тему «Інформаційна система підтримки підприємств малого бізнесу у сфері послуг» не

5.3.5 Спецустаткування для наукових (експериментальних) робіт

До статті «Спецустаткування для наукових (експериментальних) робіт» належать витрати на виготовлення та придбання спецустаткування необхідного для проведення досліджень, також витрати на їх проектування, виготовлення, транспортування, монтаж та встановлення. В роботі спецустаткування не використовувалося.

5.3.6 Програмне забезпечення для наукових (експериментальних) робіт

До статті «Програмне забезпечення для наукових (експериментальних) робіт» належать витрати на розробку та придбання спеціальних програмних засобів і програмного забезпечення, (програм, алгоритмів, баз даних) необхідних для проведення досліджень, також витрати на їх проектування, формування та встановлення. Програмне забезпечення, яке було використано в роботі було

безкоштовним.

5.3.7 Амортизація обладнання, програмних засобів та приміщень

В спрощеному вигляді амортизаційні відрахування по кожному виду обладнання, приміщень та програмному забезпеченню тощо, розраховуємо з використанням прямолінійного методу амортизації за формулою:

$$A_{обл} = \frac{Ц_б}{T_в} \cdot \frac{t_{вик}}{12}, \quad (5.12)$$

де $Ц_б$ — балансова вартість обладнання, програмних засобів, приміщень тощо, які використовувались для проведення досліджень, грн;

$t_{вик}$ — термін використання обладнання, програмних засобів, приміщень під час досліджень, місяців;

$T_в$ — строк корисного використання обладнання, програмних засобів, приміщень тощо, років.

$$A_{обл} = (50000 \cdot 2) / (2 \cdot 12) = 4166{,}67 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 5.8 — Амортизаційні відрахування по кожному виду обладнання

Найменування обладнання	Балансова вартість, грн	Строк корисного використання, років	Термін використання обладнання, місяців	Амортизаційні відрахування, грн
Персональний комп'ютер Core I3-2700	50000	2	2	4166,67
Приміщення	248700	20	2	2072,50
Всього				6239,17

5.3.8 Паливо та енергія для науково-виробничих цілей

Витрати на силову електроенергію ($B_е$) розраховуємо за формулою:

$$B_e = \sum_{i=1}^{n} \frac{W_{yi} \cdot t_i \cdot Ц_e \cdot K_{вni}}{\eta_i}, \quad (5.13)$$

де W_{yi} — встановлена потужність обладнання на визначеному етапі розробки, кВт;

t_i — тривалість роботи обладнання на етапі дослідження, год;

$Ц_e$ — вартість 1 кВт-години електроенергії, грн; (вартість електроенергії визначається за даними енергопостачальної компанії), приймемо $Ц_е = 7,5$ грн;

$K_{вni}$ — коефіцієнт, що враховує використання потужності, $K_{вni} < 1$;

η_i — коефіцієнт корисної дії обладнання, $\eta_i < 1$.

$$В_е = 0,25 \cdot 280,0 \cdot 7,5 \cdot 0,5 / 0,8 = 328,13 \text{ грн.}$$

5.3.9 Службові відрядження

До статті «Службові відрядження» дослідної роботи на тему “Методи машинного навчання для прогнозування у дата-майнінгу” належать витрати на відрядження штатних працівників, працівників організацій, які працюють за договорами цивільно-правового характеру, аспірантів, зайнятих розробленням досліджень, відрядження, пов’язані з проведенням випробувань машин та приладів, а також витрати на відрядження на наукові з’їзди, конференції, наради, пов’язані з виконанням конкретних досліджень.

Витрати за статтею «Службові відрядження» розраховуємо як 20…25% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{св} = (З_о + З_р) \cdot \frac{H_{св}}{100\%}, \quad (5.14)$$

де $H_{св}$ — норма нарахування за статтею «Службові відрядження», приймемо $Н_{св} = 20\%$.

$$В_{св} = (35977 + 8225,4) \cdot 20 / 100\% = 8840,54 \text{ грн.}$$

5.3.10 Інші витрати

До статті «Інші витрати» належать витрати, які не знайшли відображення у зазначених статтях витрат і можуть бути віднесені безпосередньо на собівартість досліджень за прямими ознаками.

Витрати за статтею «Інші витрати» розраховуємо як 50…100% від суми основної заробітної плати дослідників та робітників за формулою:

$$I_в = (З_о + З_р) \cdot \frac{Н_{ів}}{100\%}, \quad (5.15)$$

де $Н_{ів}$ — норма нарахування за статтею «Інші витрати», приймемо $Н_{ів}$ = 50%.

$$І_в = (35977 + 8225{,}4) \cdot 50 / 100\% = 22101{,}36 \text{ грн.}$$

5.3.11 Накладні (загальновиробничі) витрати

До статті «Накладні (загальновиробничі) витрати» належать: витрати, пов’язані з управлінням організацією; витрати на винахідництво та раціоналізацію; витрати на підготовку (перепідготовку) та навчання кадрів; витрати, пов’язані з набором робочої сили; витрати на оплату послуг банків; витрати, пов’язані з освоєнням виробництва продукції; витрати на науково-технічну інформацію та рекламу та ін.

Витрати за статтею «Накладні (загальновиробничі) витрати» розраховуємо як 100…150% від суми основної заробітної плати дослідників та робітників за формулою:

$$В_{нзв} = (З_о + З_р) \cdot \frac{Н_{нзв}}{100\%}, \quad (5.16)$$

де $Н_{нзв}$ — норма нарахування за статтею «Накладні (загальновиробничі) витрати», приймемо $Н_{нзв}$ = 100%.

$$В_{нзв} = (35977 + 8225,4) \cdot 100 / 100\% = 44202,72 \text{ грн.}$$

Витрати на проведення науково-дослідної роботи на тему “Методи машинного навчання для прогнозування у дата-майнінгу”. розраховуємо як суму всіх попередніх статей витрат за формулою:

$$B_{заг} = З_о + З_р + З_{дод} + З_н + М + К_в + В_{спец} + В_{прг} + А_{обл} + В_е + В_{св} + В_{сп} + І_в + В_{нзв}. \quad (5.17)$$

$В_{заг}$=35977+8225,4+4862,3+10794,3+522,39+6239,17+328,14+8840,54+2210

1,36+44202,72= 142093,63 грн.

Загальні витрати *ЗВ* на завершення науково-дослідної (науково-технічної) роботи та оформлення її результатів розраховується за формулою:

$$ЗВ = \frac{B_{заг}}{\eta}, \quad (5.18)$$

де η — коефіцієнт, який характеризує етап (стадію) виконання науково-дослідної роботи, приймемо η=0,7.

ЗВ = 142093,63 / 0,7 = 202990,89 грн.

5.4 Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором

В ринкових умовах узагальнюючим позитивним результатом, що його може отримати потенційний інвестор від можливого впровадження результатів тієї чи іншої науково-технічної розробки, є збільшення у потенційного інвестора величини чистого прибутку.

Результати дослідження проведені за темою “Методи машинного навчання для прогнозування у дата-майнінгу” передбачають комерціалізацію протягом 3-х років реалізації на ринку.

В цьому випадку основу майбутнього економічного ефекту будуть

формувати:

ΔN — збільшення кількості споживачів яким надається відповідна інформаційна послуга у періоди часу, що аналізуються;

N — кількість споживачів яким надавалась відповідна інформаційна послуга у році до впровадження результатів нової науково-технічної розробки, приймемо 1 особа

$Ц_б$ — вартість послуги у році до впровадження інформаційної системи, приймемо 1000,00 грн;

$\pm\Delta Ц_о$ — зміна вартості послуги від впровадження результатів, приймемо зростання на 300,00 грн.

Можливе збільшення чистого прибутку у потенційного інвестора $\Delta П_i$ для кожного із 3-х років, протягом яких очікується отримання позитивних результатів від можливого впровадження та комерціалізації науково-технічної розробки, розраховуємо за формулою :

$$\Delta П_i = (\pm\Delta Ц_о \cdot N + Ц_о \cdot \Delta N)_i \cdot \lambda \cdot \rho \cdot (1 - \frac{\vartheta}{100}), \quad (5.19)$$

де λ — коефіцієнт, який враховує сплату потенційним інвестором податку на додану вартість. У 2021 році ставка податку на додану вартість складає 20%, а коефіцієнт $\lambda = 0{,}8333$;

ρ — коефіцієнт, який враховує рентабельність інноваційного продукту). Приймемо $\rho = 40\%$;

ϑ — ставка податку на прибуток, який має сплачувати потенційний інвестор, у 2023 році $\vartheta = 18\%$.

Збільшення чистого прибутку 1-го року:

$$\Delta П_1 = (1 \cdot 300 + 1000 \cdot 1000) \cdot 0{,}83 \cdot 0{,}4 \cdot (1 - 0{,}18/100\%) = 222125{,}7 \text{ грн.}$$

Збільшення чистого прибутку 2-го року:

$\Delta П_2 = (1 \cdot 300 + 1000 \cdot (1000 + 800)) \cdot 0{,}83 \cdot 0{,}4 \cdot (1 - 0{,}18/100\%) = 400034{,}01$ грн.

Збільшення чистого прибутку 3-го року:

$\Delta П_3 = (1 \cdot 300 + 1000 \cdot (1000 + 800 + 600)) \cdot 0{,}83 \cdot 0{,}4 \cdot (1 - 0{,}18/100\%) = 533278{,}68$ грн.

Приведена вартість збільшення всіх чистих прибутків *ПП*, що їх може отримати потенційний інвестор від можливого впровадження та комерціалізації науково-технічної розробки:

$$ПП = \sum_{i=1}^{T} \frac{\Delta П_i}{(1+\tau)^t}, \tag{5.20}$$

де $\Delta П_i$ — збільшення чистого прибутку у кожному з років, протягом яких виявляються результати впровадження науково-технічної розробки, грн;

T — період часу, протягом якого очікується отримання позитивних результатів від впровадження та комерціалізації науково-технічної розробки, роки;

τ — ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні, τ =18%;

t — період часу (в роках) від моменту початку впровадження науково-технічної розробки до моменту отримання потенційним інвестором додаткових чистих прибутків у цьому році.

$ПП = 222125{,}7 / (1+0{,}18)^1 + 400034{,}01 / (1+0{,}18)^2 + 533278{,}68 / (1+0{,}18)^3 =$

$= 7721951{,}89$ грн.

Величина початкових інвестицій PV, які потенційний інвестор має вкласти для впровадження і комерціалізації науково-технічної розробки:

$$PV = k_{інв} \cdot ЗВ, \tag{5.21}$$

де $k_{інв}$ — коефіцієнт, що враховує витрати інвестора на впровадження науково-

технічної розробки та її комерціалізацію, приймаємо $k_{інв}$=2;

$ЗВ$ — загальні витрати на проведення науково-технічної розробки та оформлення її результатів, приймаємо 202990,89 грн.

$$PV = k_{інв} \cdot ЗВ = 2 \cdot 202990{,}89 = 405981{,}79 \text{ грн.}$$

Абсолютний економічний ефект $E_{абс}$ для потенційного інвестора від можливого впровадження та комерціалізації науково-технічної розробки становитиме:

$$E_{абс} = ПП - PV \tag{5.22}$$

де $ПП$ – приведена вартість зростання всіх чистих прибутків від можливого впровадження та комерціалізації науково-технічної розробки, 7721951,89 грн;

PV – теперішня вартість початкових інвестицій, 405981,79 грн.

$$E_{абс} = ПП - PV = 7721951{,}89 - 405981{,}79 = 366970{,}09 \text{ грн.}$$

Внутрішня економічна дохідність інвестицій $E_в$, які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$E_в = \sqrt[Tж]{1 + \frac{E_{абс}}{PV}} - 1, \tag{5.23}$$

де $E_{абс}$ — абсолютний економічний ефект вкладених інвестицій, грн;

PV — теперішня вартість початкових інвестицій, грн;

$T_ж$ — життєвий цикл науково-технічної розробки, тобто час від початку її розробки до закінчення отримування позитивних результатів від її впровадження, 3 роки.

$$E_в = \sqrt[Tж]{1+\frac{E_{абс}}{PV}} - 1 = (1+366970{,}09\ /405981{,}79)^{1/3} - 1 = 0{,}41.$$

Мінімальна внутрішня економічна дохідність вкладених інвестицій $\tau_{мін}$:

$$\tau_{мін} = d + f, \qquad (5.24)$$

де d — середньозважена ставка за депозитними операціями в комерційних банках; в 2023 році в Україні d =0,1;

f — показник, що характеризує ризикованість вкладення інвестицій, приймемо 0,25.

$\tau_{мін}$ = 0,1+0,25 = 0,35 < 0,41 свідчить про те, що внутрішня економічна дохідність інвестицій $E_в$, які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки вища мінімальної внутрішньої дохідності.

Період окупності інвестицій $T_{ок}$ які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$T_{ок} = \frac{1}{E_в}, \qquad (5.25)$$

де $E_в$ — внутрішня економічна дохідність вкладених інвестицій.

$T_{ок}$ = 1 / 0,41 = 2,4 р.

$T_{ок}$ < 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою “Методи машинного навчання для прогнозування у дата-майнінгу”

становить 33 бали, що, свідчить про комерційну важливість проведення даних досліджень оскільки рівень комерційного потенціалу розробки вище середнього.

При оцінюванні рівня конкурентоспроможності, згідно узагальненого коефіцієнту конкурентоспроможності розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 1,5 рази.

Також термін окупності становить 2,4 роки, що менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

Отже можна зробити висновок про доцільність проведення науково-дослідної роботи за темою “Методи машинного навчання для прогнозування у дата-майнінгу”.

ВИСНОВКИ

У магістерській кваліфікаційній роботі було спроектовано та розроблено веб-систему для прогнозування в дата-майнінгу.

Досліджено вплив алгоритмів машинного навчання на прогнозування в дата-майнінгу. Розглянуто особливості автоматизованого прогнозування в дата-майнінгу.

Вивчено недоліки та переваги наявних веб-систем прогнозування в дата-майнінгу, що дозволило обґрунтувати вимоги до розроблення власної підсистеми.

Розроблено структуру системи прогнозування в дата-майнінгу та наведено інтерфейс. Основними цілями створення програми для прогнозування в дата-майнінгу є: виконання аналізу часових рядів різного виду; автоматизація процесу виведення результату; автоматизація процесу проведення аналізу та прогнозування часових рядів. Основою методики створення інформаційної технології для прогнозування в дата-майнінгу є Model-view-controller.

На базі аналізу існуючих математичних моделей було запропоновано підхід до реалізації побудови системи прогнозування в дата-майнінгу, досліджено механізм роботи прогнозування та на його основі розроблено математичну модель процесу прогнозування.

Обґрунтовано вибір інструментів для програмної реалізації системи прогнозування в дата-майнінгу. Для клієнтської частини та відображення інтерфейсу користувача було вирішено розробляти веб-додаток за допомогою PHP. Сервер бази даних обрано Microsoft SQL Server.

У результаті роботи було розроблено веб-застосунок та базу даних.

Доведено, що узагальненими показниками динаміки розвитку економічних процесів є середній приріст, середній темп зростання і приросту. При виконанні ряду передумов ці показники можуть бути використані в наближених, найпростіших способах прогнозування, що передують більш глибокому кількісному і якісному аналізу.

Здійснено тестування та аналіз роботи розробленї системи. Спочатку було проведено тестування за допомогою розробленого додатку та порівняно його з існуючими методами.

Отримані в кваліфікаційній роботі теоретичні результати являють собою сукупність знань, що лежать в основі методологічної бази побудови прикладних моделей для дослідження динамічних ефектів в розвитку сучасних економічних процесів. Ця база забезпечує формалізацію і рішення нових прикладних задач, які раніше не цікавили вітчизняну економічну науку.

Провівши порівняльний аналіз розглянутих підходів, є можливість стверджувати, що розроблений веб-додаток є кращим за точністю аналізу та точністю прогнозування, що є головними показниками у розглянутому підході, а також швидкодія спрацювання перевищує значення майже у 4 рази. Враховуючи той факт, що при розробці було використано машинне навчання, яке на сьогодні є актуальним напрямком досліджень багатьох вчених, є можливість стверджувати про популярність розгляду даного веб-додатку серед науковців.

Розроблений веб-додаток є універсальним та може знайти своє застосування у прогнозуванні будь-яких фінансових показників підприємств діяльність яких направлена на отримання прибутку.

Щодо подальшого розвитку розробленого веб-додатку варто підкреслити, що можливе розширення функціональної складової додатку у напрямку максимізації виконуваних функцій: порівняння прогнозного значення за роками; порівняння прогнозного значення у спектрі інших компаній; формування лінії тренду; виведення графічного результату та збереження його в окремий файл.

Отже, задачі, поставлені у магістерській кваліфікаційній роботі, було виконано в повному обсязі.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Кіріченко Л. О. Фрактальний аналіз самоподібних і мультифрактальних часових рядів : монографія. – Харків: Панов А. Н., 2019

2. Кучанський О. Ю. Інформаційна система підтримки прийняття рішень у діяльності фінансових установ на основі трендових моделей : автореф. дис. ... канд. техн. наук : 05.13.06. – Київ, 2014

3. Андрієнко В. О. Аналіз фондових ринків : навч. посіб. для студ. екон. спец. – О.: Астропринт, 2011

4. Ісаєнко О. О. Моделювання нестаціонарних часових рядів економічної динаміки на основі рівняння Фоккера – Планка // Бізнес Інформ. – 2014. – № 4.

5. Баклан І. В. Лінгвістичне моделювання часових рядів різної природи з фрактальними властивостями // Систем. технології. – 2016. – № 3.

6. Недашківський Є. Механізми визначення фрактальності у термінах лінгвістичного моделювання // Техн. науки та технології. – 2016. – № 3.

7. Дорош Н. Л. Дослідження часового ряду спеціалізованої інформації // Систем. технології. – 2017. – № 4.

8. Hurst, H.E. (1951), “Long–term storage of reservoirs: an experimental study”, Transactions of the American Society of Civil Engineers, Vol. 116. – pp. 770–799

9. Кучерук І. М., Горбачук І. Т., Луцик П. П. Загальний курс фізики : навч. посібник у 3–х т. – Київ: Техніка, 2006. – Т.2: Електрика і магнетизм.

10. Кравець Т. В. Ефекти синхронізації динаміки фондових індексів та курсів валют при мультифрактальному аналізі з використанням вейвлет технологій // Бізнес Інформ. – 2014. – № 2.

11. Мокін О. Б. До питання вибору оптимальної математичної моделі стаціонарного часового ряду // Вісн. Вінниц. політехн. ін–ту. – 2018. – № 4

12. Семененко Т. О. Прогнозування динаміки макроекономічних показників // Вісн. Сум. держ. ун–ту. Сер. Економіка. – 2019. – № 3.

13. Кушнір М. Я. Використання систем штучного інтелекту у задачах прогнозування фінансових індексів: огляд наукових джерел // Радіоелектрон. і комп'ютер. системи. – 2020. – № 3.

14. Лузіна А.О. Застосування адаптивних моделей для прогнозування чистого доходу від реалізації продукції / А.О. Лузіна, С.П. Кобець // Тези 71–ої наукової конференції професорів, викладачів, наукових працівників, аспірантів та студентів університету (Полтава, 22 квітня – 17 травня 2019 р.). – Полтава : ПолтНТУ, 2019. – Т. 1. – С. 259–260..

15. Математика – історія і сучасність : реком. бібліогр. покажчик : у 2 ч. / укл. О. В. Бобир. – Кривий Ріг : Наукова бібліотека КНУ, 2013. – 22 с..

16. Веровчук В. В. Прогнозування поведінки інформаційного продукту. – 2012 // Вісн. Нац. ун–ту "Львів. політехніка". С. 54–63..

17. Helmert, F. R. (1876b). "Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers directer Beobachtungen gleicher Genauigkeit", Astron. Nachr., 88, 113–32

18. Кащєєва В. Ю. Інформаційна технологія аналізу багатокомпонентних процесів за часовими рядами на основі інтервальних прогнозних моделей : автореф. дис. ... канд. техн. наук : 05.13.06. – Харків, 2014.

19. Степашко В. С. Комп'ютерний експеримент в індуктивному моделюванні : [монографія]. – Київ: Наук. думка, 2014. 576 с.

20. Григорків В. С. Економетрика. Лінійні моделі парної та множинної регресії : навч. посіб. – Чернівці, 2009

21. Козьменко О. В. Економіко–математичні методи та моделі (економетрика) : навч. посіб. – Суми: Унів. кн., 2014

22. Гур'янова Л. С. Економетрика : навч. посіб. – Харків: ХНЕУ ім. С. Кузнеця, 2015

23. Тарасов І. В. Економетрика : навч. посіб. – Миколаїв: НУК, 2019. – 150.

24. Михайловська О. Фрактали як ступінь організованості інвестиційних процесів у сучасних умовах. – 2011 // Економіка України.

25. Роговська–Іщук І. Практичні аспекти технічного аналізу на ринку FOREX. – 2006 // Журн. європ. економіки.

26. Hosking, J. R. M., Fractional Differencing. Biometrica 68, 1981. – p. 165–76.

27. Баранов В. Л. Диференціальні перетворення випадкових процесів. – 2002 // Вісн. Житомир. інж.–технол. ін–ту. Техн. науки

28. Горбачук В. М. Особливості регресійного аналізу часових рядів / В. М. Горбачук, Ю. Г. Кривонос // Комп'ютерна математика. – 2012. – № 2. – С. 3–12

29. Горбачук В. М. Економетричне програмування TSP та EViews. Препр. 96–14 / В. М. Горбачук. – К. : Ін–т кібернетики ім. В. М. Глушкова НАН України, 2000. – 24 с

30. Яковлев А. І. Економічна сутність та методичні основи визначення рівня потенціалу виробничої системи / А. І. Яковлев, О. П. Косенко // Маркетинг і менеджмент інновацій. – 2011. – № 2. – С. 172– 178.

31. Горбачук В. М. Особливості регресійного аналізу часових рядів / В. М. Горбачук, Ю. Г. Кривонос // Комп'ютерна математика. – 2012. – № 2. – С. 3–12.

32. Wooldridge J. M. Introductory econometrics: a modern approach / J. M. Wooldridge. – 4–th edition. – Mason, OH : Cencage Learning, 2009. – 865 p

33. Brown R.G., Meyer R.F. The fundamental theorum of exponential smoothing. Oper. Res. – 1961. – Vol.9. –№ 5.

34. Holt C.C. Forecasting trends and seasonals by exponentially weighted moving averages // O.N.R. Memorandum, Carnegie Inst. of Technology. – 1957. – № 2.

35. Полозюк О. Є. Математичний аналіз. Функції та їх неперервність. Похідна і диференціал. Комплексне число. – Донецк: Юго–Восток, 2002.

36. Міхайленко В. М. Математичний аналіз для економістів : Навч. посіб. для студ. екон. спец. вищ. навч. закл. – 2–е вид. – К., 2002.

37. Яковлев А. І. Економічна сутність та методичні основи визначення рівня потенціалу виробничої системи / А. І. Яковлев, О. П. Косенко // Маркетинг і менеджмент інновацій. – 2011. – № 2. – С. 172– 178.

38. Сергієнко О. А., Татар М. С. Дослідження факторів зовнішнього середовища в процесі формування конкурентних стратегій розвитку підприємства [Текст] / О. А. Сергієнко, М. С. Татар // Просторова економіка: концепції, моделі та регіональні аспекти : монографія / за ред. П. В. Захарченко, Т. П. Несторенко. – Бердянськ : Вид. Ткачук, 2012. – рос. мова, укр. мова: іл. – С. 289–303.

39. Шаповалова О. О. Моделювання попиту із залученням методів аналізу часових рядів / О. О. Шаповалова, Г. В. Солодовник, І. О. Татаров // Комунальне господарство міст. Сер. : Економічні науки. – 2016. – Вип. 127. – С. 10–14. – Режим доступу: http://nbuv.gov.ua/UJRN/kgm_econ_2016_127_4

40. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт / Уклад. : В. О. Козловський, О. Й. Лесько, В. В. Кавецький. – Вінниця : ВНТУ, 2021. – 42 с.

ДОДАТОК А

Міністерство освіти і науки України

Вінницький національний технічний університет

Факультет інформаційних технологій та комп'ютерної інженерії

Кафедра обчислювальної техніки

ЗАТВЕРДЖУЮ

Завідувач кафедри

обчислювальної техніки

_____проф., д.т.н. О. Д. Азаров

«29» вересня 2023 року

ТЕХНІЧНЕ ЗАВДАННЯ

на виконання магістерської кваліфікаційної роботи

"Методи машинного навчання для прогнозування у дата-майнінгу"

08-54.МКР.010.00.000 ТЗ

Науковий керівник к.пед.н., доц. каф. ОТ

_____Добровольська Н.В.

Студента групи 1КІ-22м

_____Мазуренка В.В.

ВНТУ 2023

1 Підставою для виконання магістерської кваліфікаційної роботи є наказ про затвердження теми дипломної роботи, а також актуальність розробки системи для прогнозування в дата-майнінгу.

2 Мета і призначення МКР:

— метою є дослідження методів машинного навчання на прогнозування в дата-майнінгу та розробка веб-додатку для прогнозування часових рядів.;

— призначенням розробки є виконання магістерської кваліфікаційної роботи з можливістю подальшого впровадження та масштабування.

3 Вихідні дані для виконання МКР:

— розробити веб-застосунок для для прогнозування часових рядів;

— основна технологія Model-view-controller;

— мова програмування PHP.

4 Технічні вимоги до виконання МКР — виведення математичної моделі процесу прогнозування.

5 Етапи МКР та очікувані результати (див. табл. А.1).

Таблиця А.1 — Етапи роботи

№ етапу	Назва етапу	Термін виконання		Очікувані результати
		початок	кінець	
1	Огляд та аналіз підходів до прогнозування в дата-майнінгу	21.09.2023	25.09.2023	Розділ 1
2	Дослідження математичних моделей	26.09.2023	04.09.2023	Розділ 2
3	Аналіз та вибір технологій проектування та розробки веб-додатків	03.10.2023	10.10.2023	Розділ 1, 3 частково
4	Програмна реалізація системи	17.10.2022	25.10.2023	Розділ 3 повністю
5	Тестування веб-додатку	26.10.2023	27.10.2023	Розділ 4
6	Підготовка економічної частини	07.11.2023	14.11.2023	Розділ 5
7	Оформлення матеріалів дозахисту МКР	21.11.2023	1.12.2023	Пояснювальна записка, графічний матеріал, презентація

6 Матеріали, що подаються до захисту МКР:

— пояснювальна записка МКР;

— графічні і ілюстративні матеріали;

— протокол попереднього захисту МКР на кафедрі;

— відгук наукового керівника;

— рецензія на виконану роботу;

— анотації до МКР українською та іноземною мовами;

— нормоконтроль про відповідність оформлення МКР діючим вимогам.

7 Порядок контролю виконання та захисту МКР:

— виконання етапів графічної та розрахункової документації КМКР контролюється науковим керівником згідно зі встановленими термінами;

— захист МКР відбувається на засіданні Державної екзаменаційної комісії, затвердженою наказом ректора.

8 Вимоги до оформлення МКР викладені в методичних вказівках до виконання магістерських кваліфікаційних робіт зі спеціальності 123 — «Комп’ютерна інженерія, ДСТУ 3008 : 2015 «Звіти в сфері науки і техніки. Структура та правила оформлювання», ДСТУ 8302 : 2015 «Бібліографічні посилання. Загальні положення та правила складання».

ДОДАТОК Б

Результати розрахунку

Таблиця Б.1 — Результати розрахунку

рік	Дата	Середнє за період, грн	Lt = експоненційно згладжений ряд	Tt = значення тренду
2013	січень	5,05	5,05	0
	лютий	5,05	5,05	0
	березень	4,85137273	4,930823638	-0,083423453
	квітень	4,84245484	4,911171741	-0,038783364
	травень	4,84454828	4,88671101	-0,028757521
	червень	4,85296333	4,87796541	-0,014749176
	липень	5,04698387	4,985276156	0,07069277
	серпень	6,02674333	5,581879353	0,438830068
	вересень	7,57456452	6,601958426	0,845704372
2014	січень	7,7	6,922501622	0,478091549
	лютий	7,7	7,197764029	0,33611115
	березень	7,7	7,364661152	0,217661331
	квітень	7,7	7,478799928	0,145195543
	травень	7,64712333	7,521715752	0,07359974
	червень	7,61596667	7,548826407	0,04105738
	липень	7,64912581	7,592583097	0,042946897
	серпень	7,81361935	7,70802609	0,093694164
	вересень	7,99988333	7,845662768	0,124453924
	січень	7,99890968	7,887829346	0,066852781
	лютий	7,99272667	7,924026628	0,045393932
	березень	7,97714194	7,937738242	0,02321631
2015	січень	7,99887308	7,965132621	0,026140958
	лютий	7,99954643	7,975324523	0,014976619
	березень	7,96599032	7,963733354	-0,003620833
	квітень	7,92572667	7,942377677	-0,016035224
	травень	7,92569667	7,938783162	-0,007326727
	червень	7,91516	7,927539956	-0,010068263
	липень	7,90139677	7,915881349	-0,011181503
	серпень	7,89000714	7,904829425	-0,011090798
	вересень	7,91058333	7,912718087	0,002194824
	січень	7,91047419	7,910493819	-0,00089854
	лютий	7,928	7,921356944	0,007334625
	березень	7,9558129	7,939096667	0,014618194
2016	січень	7,94785	7,938501389	0,003968764
	лютий	7,94082143	7,938305908	0,001053792
	березень	7,9441129	7,941368586	0,002460012
	квітень	7,96494667	7,954531432	0,009951995
	травень	7,96798571	7,958623201	0,005849837
	червень	7,96798564	7,961900729	0,004049221
	липень	7,96798578	7,963932071	0,002636706
	серпень	7,96818573	7,965429584	0,001839271

	вересень	7,9736	7,969596125	0,00346836
	січень	7,975	7,971451106	0,002338995
	лютий	7,978	7,974444845	0,002797315
	березень	7,978	7,975459012	0,001549112
2017	січень	7,98976923	7,983425498	0,006041274
	лютий	7,98848966	7,984047486	0,002247773
	березень	7,98628065	7,984488275	0,000982884
	квітень	7,98673448	7,985442844	0,000963064
	травень	7,99081613	7,98828159	0,002276041
	червень	7,9925	7,98990222	0,001817253
	липень	7,9928871	7,990966247	0,001289995
	серпень	7,993	7,991670501	0,000879976
	вересень	7,993	7,99211621	0,000575989
	січень	7,993	7,992416088	0,000382712
	лютий	7,993	7,992613351	0,000252897
	березень	7,993	7,992744181	0,000167451
2018	січень	7,993	7,992830692	0,000110793
	лютий	7,993	7,99288796	7,33251E-05
	березень	7,993	7,992925854	4,85234E-05
	квітень	7,993	7,992950932	3,21118E-05
	травень	7,993	7,992967528	2,12507E-05
	червень	7,993	7,992978511	1,40632E-05
	липень	7,993	7,992985779	9,30666E-06
	серпень	7,993	7,992990589	6,15891E-06
	вересень	7,993	7,992993772	4,07581E-06
	січень	7,993	7,992995878	2,69726E-06
	лютий	7,993	7,992997272	1,78498E-06
	березень	7,993	7,992998195	1,18125E-06
2019	січень	7,993	7,992998806	7,81723E-07
	лютий	8,64986786	8,387119926	0,275885019
	березень	9,90722903	9,188831381	0,643963524
	квітень	11,63118113	10,39665582	1,038666165
	травень	11,401581	10,58414446	0,442841899
	червень	11,7351346	11,09760179	0,492272696
	липень	12,90308977	11,9839855	0,768150407
	серпень	13,02130904	12,29911946	0,451038896
	вересень	12,94980139	12,50911306	0,282307188
	січень	14,64090228	13,67526372	0,900997616
	лютий	15,59791223	14,46845378	0,825532328
2020	січень	15,8108595	14,94368428	0,58032105
	лютий	24,31495433	20,33431789	3,947539842
	березень	23,32590055	20,55025155	1,335415514
	квітень	22,60931703	21,25152463	0,891515812
	травень	20,93077637	20,70246935	-0,116883954
	червень	21,2392968	21,0713194	0,22312985
	липень	21,76812368	21,40015003	0,297120394
	серпень	21,5744203	21,38586403	0,079135922
	вересень	21,65309764	21,51454983	0,113820833
	січень	21,85386368	21,67260981	0,144788234

	лютий	23,33553457	22,61244937	0,701324165
	березень	23,491182	22,85915928	0,383094188
2021	січень	24,27320752	23,55435055	0,601562144
	лютий	26,41259283	25,02867106	1,212493001
	березень	26,35716371	25,34076945	0,582216773
	квітень	25,59542986	25,26067899	0,118601708
	травень	25,2198593	25,18874649	-0,014772234
	червень	24,9501771	25,05151375	-0,100494589
	липень	24,81979723	24,95268167	-0,09933083
	серпень	25,09075858	25,07526015	0,056005684
	вересень	26,26321467	25,76563059	0,500061012
	січень	25,74917219	25,55573114	0,003088693
	лютий	25,70562786	25,6444337	0,063018394
	березень	26,21993977	25,96452998	0,242972919
2022	січень	27,245807	26,63610703	0,542995806
	лютий	27,05494414	26,67021097	0,186771504
	березень	26,99483071	26,79027421	0,14007572
	квітень	26,8452706	26,76724176	0,025899997

ДОДАТОК В

Результати розрахунку

Таблиця В.1 — Результати розрахунку

рік	Дата	Середнє за період, грн	Lt = експоненційно згладжений ряд	Tt = значення тренда	p - номер періоду для прогнозу	Прогноз по методу Хольта Ŷt+p = Lt + p *Tt	Прогноз на 1 період для аналізу	Похибка моделі	відхилення помилки моделі від прогнозної моделі	Точність прогнозу
2013	січень	5,05	5,05	0,00			5,05			1,00
	лютий	5,05	5,05	0,00			5,05	0,00	0,00	
	березень	4,85	4,93	-0,08			5,05	-0,20	0,00	
	квітень	4,84	4,91	-0,04			4,85	0,00	0,00	
	травень	4,84	4,89	-0,03			4,87	-0,03	0,00	
	червень	4,85	4,88	-0,01			4,86	0,00	0,00	
	липень	5,05	4,99	0,07			4,86	0,18	0,00	
	серпень	6,03	5,58	0,44			5,06	0,97	0,03	
	вересень	7,57	6,60	0,85			6,02	1,55	0,04	
2014	січень	7,70	6,92	0,48			7,45	0,25	0,00	
	лютий	7,70	7,20	0,34			7,40	0,30	0,00	
	березень	7,70	7,36	0,22			7,53	0,17	0,00	
	квітень	7,70	7,48	0,15			7,58	0,12	0,00	
	травень	7,65	7,52	0,07			7,62	0,02	0,00	
	червень	7,62	7,55	0,04			7,60	0,02	0,00	
	липень	7,65	7,59	0,04			7,59	0,06	0,00	
	серпень	7,81	7,71	0,09			7,64	0,18	0,00	
	вересень	8,00	7,85	0,12			7,80	0,20	0,00	
	січень	8,00	7,89	0,07			7,97	0,03	0,00	
	лютий	7,99	7,92	0,05			7,95	0,04	0,00	
	березень	7,98	7,94	0,02			7,97	0,01	0,00	
2015	січень	8,00	7,97	0,03			7,96	0,04	0,00	
	лютий	8,00	7,98	0,01			7,99	0,01	0,00	
	березень	7,97	7,96	0,00			7,99	-0,02	0,00	
	квітень	7,93	7,94	-0,02			7,96	-0,03	0,00	
	травень	7,93	7,94	-0,01			7,93	0,00	0,00	
	червень	7,92	7,93	-0,01			7,93	-0,02	0,00	
	липень	7,90	7,92	-0,01			7,92	-0,02	0,00	
	серпень	7,89	7,90	-0,01			7,90	-0,01	0,00	
	вересень	7,91	7,91	0,00			7,89	0,02	0,00	
	січень	7,91	7,91	0,00			7,91	0,00	0,00	

	лютий	7,93	7,92	0,01			7,91	0,02	0,00	
	березень	7,96	7,94	0,01			7,93	0,03	0,00	
2016	січень	7,95	7,94	0,00			7,95	-0,01	0,00	
	лютий	7,94	7,94	0,00			7,94	0,00	0,00	
	березень	7,94	7,94	0,00			7,94	0,00	0,00	
	квітень	7,96	7,95	0,01			7,94	0,02	0,00	
	травень	7,97	7,96	0,01			7,96	0,00	0,00	
	червень	7,97	7,96	0,00			7,96	0,00	0,00	
	липень	7,97	7,96	0,00			7,97	0,00	0,00	
	серпень	7,97	7,97	0,00			7,97	0,00	0,00	
	вересень	7,97	7,97	0,00			7,97	0,01	0,00	
	січень	7,98	7,97	0,00			7,97	0,00	0,00	
	лютий	7,98	7,97	0,00			7,97	0,00	0,00	
	березень	7,98	7,98	0,00			7,98	0,00	0,00	
2017	січень	7,99	7,98	0,01			7,98	0,01	0,00	
	лютий	7,99	7,98	0,00			7,99	0,00	0,00	
	березень	7,99	7,98	0,00			7,99	0,00	0,00	
	квітень	7,99	7,99	0,00			7,99	0,00	0,00	
	травень	7,99	7,99	0,00			7,99	0,00	0,00	
	червень	7,99	7,99	0,00			7,99	0,00	0,00	
	липень	7,99	7,99	0,00			7,99	0,00	0,00	
	серпень	7,99	7,99	0,00			7,99	0,00	0,00	
	вересень	7,99	7,99	0,00			7,99	0,00	0,00	
	січень	7,99	7,99	0,00			7,99	0,00	0,00	
	лютий	7,99	7,99	0,00			7,99	0,00	0,00	
	березень	7,99	7,99	0,00			7,99	0,00	0,00	
2018	січень	7,99	7,99	0,00			7,99	0,00	0,00	
	лютий	7,99	7,99	0,00			7,99	0,00	0,00	
	березень	7,99	7,99	0,00			7,99	0,00	0,00	
	квітень	7,99	7,99	0,00			7,99	0,00	0,00	
	травень	7,99	7,99	0,00			7,99	0,00	0,00	
	червень	7,99	7,99	0,00			7,99	0,00	0,00	
	липень	7,99	7,99	0,00			7,99	0,00	0,00	
	серпень	7,99	7,99	0,00			7,99	0,00	0,00	
	вересень	7,99	7,99	0,00			7,99	0,00	0,00	
	січень	7,99	7,99	0,00			7,99	0,00	0,00	
	лютий	7,99	7,99	0,00			7,99	0,00	0,00	
	березень	7,99	7,99	0,00			7,99	0,00	0,00	
2019	січень	7,99	7,99	0,00			7,99	0,00	0,00	
	лютий	8,65	8,39	0,28			7,99	0,66	0,01	
	березень	9,91	9,19	0,64			8,66	1,24	0,02	
	квітень	11,63	10,40	1,04			9,83	1,80	0,02	
	травень	11,40	10,58	0,44			11,44	-0,03	0,00	
	червень	11,74	11,10	0,49			11,03	0,71	0,00	
	липень	12,90	11,98	0,77			11,59	1,31	0,01	
	серпень	13,02	12,30	0,45			12,75	0,27	0,00	
	вересень	12,95	12,51	0,28			12,75	0,20	0,00	

	січень	14,64	13,68	0,90			12,79	1,85	0,02	
	лютий	15,60	14,47	0,83			14,58	1,02	0,00	
2020	січень	15,81	14,94	0,58			15,29	0,52	0,00	
	лютий	24,31	20,33	3,95			15,52	8,79	0,13	
	березень	23,33	20,55	1,34			24,28	-0,96	0,00	
	квітень	22,61	21,25	0,89			21,89	0,72	0,00	
	травень	20,93	20,70	-0,12			22,14	-1,21	0,00	
	червень	21,24	21,07	0,22			20,59	0,65	0,00	
	липень	21,77	21,40	0,30			21,29	0,47	0,00	
	серпень	21,57	21,39	0,08			21,70	-0,12	0,00	
	вересень	21,65	21,51	0,11			21,46	0,19	0,00	
	січень	21,85	21,67	0,14			21,63	0,23	0,00	
	лютий	23,34	22,61	0,70			21,82	1,52	0,00	
	березень	23,49	22,86	0,38			23,31	0,18	0,00	
2021	січень	24,27	23,55	0,60			23,24	1,03	0,00	
	лютий	26,41	25,03	1,21			24,16	2,26	0,01	
	березень	26,36	25,34	0,58			26,24	0,12	0,00	
	квітень	25,60	25,26	0,12			25,92	-0,33	0,00	
	травень	25,22	25,19	-0,01			25,38	-0,16	0,00	
	червень	24,95	25,05	-0,10			25,17	-0,22	0,00	
	липень	24,82	24,95	-0,10			24,95	-0,13	0,00	
	серпень	25,09	25,08	0,06			24,85	0,24	0,00	
	вересень	26,26	25,77	0,50			25,13	1,13	0,00	
	січень	25,75	25,56	0,00			26,27	-0,52	0,00	
	лютий	25,71	25,64	0,06			25,56	0,15	0,00	
	березень	26,22	25,96	0,24			25,71	0,51	0,00	
2022 рік	січень	27,25	26,64	0,54			26,21	1,04	0,00	
	лютий	27,05	26,67	0,19			27,18	-0,12	0,00	
	березень	26,99	26,79	0,14			26,86	0,14	0,00	
	квітень	26,85	26,77	0,03			26,93	-0,09	0,00	
2023	січень				1,00	26,51				
	лютий				2,00	26,41				
	березень				3,00	26,31				

ДОДАТОК Г

Результати розрахунку за методом Вінтерса

Таблиця Г.1 — Результати розрахунку за методом Вінтерса

Рік	Місяць	номер	Середнє за період, грн	Lt = експоненційно згладжений ряд	Tt = значення тренда	p - номер періоду для прогнозу	Прогноз по методу Вінтерса Ŷt+p = Lt + p *Tt	Прогноз на 1 період для аналізу	Похибка моделі	відхилення помилки моделі від прогнозної моделі	Точність прогнозу
2016	січень	70	70		1			70	0	0	1
2017	лютий	43	57	-12	1			70	-27	0	
2017	березень	52	48	-9	1			44	8	0	
2017	квітень	69	54	5	1			39	30	0	
2017	травень	50	54	1	1			59	-9	0	
2017	червень	67	61	6	1			55	12	0	
2017	липень	69	68	7	1			67	2	0	
2017	серпень	59	67	0	1			75	-16	0	
2017	вересень	80	73	6	1			67	13	0	
2017	січень	101	90	16	1			79	22	0	
2017	лютий	79	92	4	1			106	-27	0	
2017	березень	86	91	-1	1			96	-10	0	
2017	квітень	93	92	0	1			90	3	0	
2018	травень	61	76	-14	1			92	-31	0	
2018	червень	90	76	-1	1			63	27	0	
2018	липень	95	85	8	1			75	20	0	
2018	серпень	102	97	12	1			93	9	0	
2018	вересень	82	96	0	1			109	-27	0	
2018	січень	96	96	0	1			95	1	0	
2018	лютий	97	96	1	1			96	1	0	
2018	березень	119	108	11	1			97	22	0	
2018	січень	65	92	-14	1			119	-54	1	
2018	лютий	74	76	-15	1			78	-4	0	
2018	березень	77	69	-8	1			61	16	0	
2018	квітень	72	66	-3	1			61	11	0	
2019	травень	59	64	-2	1			56	3	0	
2019	червень	75	66	1	1			68	7	0	
2019	липень	84	73	7	1			70	14	0	
2019	серпень	85	81	8	1			82	3	0	
2019	вересень	69	82	1	1			83	-14	0	
2019	січень	80	82	0	1			83	-3	0	
2019	лютий	99	90	8	1			82	17	0	
2019	березень	72	83	-6	1			103	-31	0	
2019	січень	96	95	10	1			66	30	0	

2019	лютий	90	98	4	1			104	-14	0	
2019	березень	87	92	-5	1			108	-21	0	
2019	квітень	104	93	0	1			91	13	0	
2020	травень	64	82	-10	1			85	-21	0	
2020	червень	78	71	-11	1			80	-2	0	
2020	липень	90	71	-1	1			66	24	0	
2020	серпень	77	72	1	1			72	5	0	
2020	вересень	59	70	-2	1			65	-6	0	
2020	січень	87	78	7	1			67	20	0	
2020	лютий	78	80	2	1			89	-11	0	
2020	березень	58	71	-7	1			78	-20	0	
2020	січень	97	84	11	1			59	38	0	
2020	лютий	102	101	16	1			90	12	0	
2020	березень	80	99	0	1			118	-38	0	
2020	квітень	64	79	-18	1			106	-42	0	
2021	травень	53	62	-17	1			51	2	0	
2021	червень	65	52	-11	1			49	16	0	
2021	липень	86	57	3	1			49	37	0	
2021	серпень	83	69	12	1			63	20	0	
2021	вересень	63	77	8	1			70	-7	0	
2021	січень	66	74	-2	1			90	-24	0	
2021	лютий	97	84	9	1			73	24	0	
2021	березень	104	105	20	1			82	22	0	
2021	січень	69	95	-6	1			130	-61	1	
2021	лютий	91	91	-5	1			87	4	0	
2021	березень	75	84	-6	1			78	-3	0	
2021	квітень	102	93	7	1			74	28	0	
2022	травень	66	89	-3	1			85	-19	0	
2022	червень	91				1	53	101	-10	0	
2022	липень	103				2	37	0	103	1	
2022	серпень	97				3	12	0	97	1	
2022	вересень	79				4	-6	0	79	1	
2022	січень	93				5	-23	0	93	1	
2022	лютий	74				6	-44	0	74	1	
2022	березень	103				7	-54	0	103	1	
2022	січень	96				8	-66	0	96	1	
2022	лютий	58				9	-91	0	58	1	
2022	березень	76				10	-98	0	76	1	
2022	квітень	94				11	-129	0	94	1	
2023	травень	56				12	-114	0	56	1	
2023	червень	77						0	77	1	
2023	липень	89						0	89	1	
2023	серпень	65						0	65	1	
2023	вересень	60						0	60	1	
2023	січень	75						0	75	1	
2023	лютий	83						0	83	1	

ДОДАТОК Д

Протокол перевірки кваліфікаційної роботи

Назва роботи: МЕТОДИ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ У ДАТА-МАЙНІНГУ

Тип роботи: магістерська кваліфікаційна робота
(БДР, МКР)

Підрозділ кафедра обчислювальної техніки
(кафедра, факультет)

Показники звіту подібності Unicheck

Оригінальність 87,4 Схожість 12,6%

Аналіз звіту подібності (відмітити потрібне):

✓ Запозичення, виявлення у роботі, оформлені коректно і не містять ознак плагіату.

□ Виявлені у роботі запозичення не мають ознак плагіату, але їх надмірна кількість викликає сумніви щодо цінності роботи і відсутності самостійності її виконання автором. Роботу направити на розгляд експертної комісії кафедри.

□ Виявлені у роботі запозичення є недобросовісними і мають ознаки плагіату та/або в ній містяться навмисні спотворення тексту, що вказують на спроби приховування недобросовісних запозичень.

Особа, відповідальна за перевірку _____ Захарченко С.М.
(підпис) (прізвище, ініціали)

Ознайомлені з повним звітом подібності, який був згенерований системою Unicheck щодо роботи.

Автор роботи _____ Мазуренко В.В.
(підпис) (прізвище, ініціали)

Керівник роботи _____ Добровольська Н.В.
(підпис) (прізвище, ініціали)