

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Чорноморський національний університет
імені Петра Могили
Факультет комп'ютерних наук
Кафедра інтелектуальних інформаційних систем

ДОПУЩЕНО ДО ЗАХИСТУ

Завідувача кафедри інтелектуальних
інформаційних систем, д-р техн.наук, проф.

_____ Ю. П. Кондратенко

«____» _____ 2024 р.

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

ІНТЕЛЕКТУАЛЬНА СИСТЕМА ПРОГНОЗУВАННЯ
ПРОДАЖУ У СФЕРІ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

Спеціальність 122 «Комп'ютерні науки»

122 – КРМ – 601.21810108

Виконав студент 6-го курсу, групи 601

_____ *А. М. Дарій*

«19» лютого 2024 р.

Керівник: канд. пед. наук, доцент

_____ *Н. М. Болюбаиш*

«19» лютого 2024 р.

Миколаїв – 2024

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Чорноморський національний університет ім. Петра Могили
Факультет комп'ютерних наук
Кафедра інтелектуальних інформаційних систем

Освітньо-кваліфікаційний рівень **магістр**

Галузь знань **12 «Інформаційні технології»**
(шифр і назва)

Спеціальність **122 «Комп'ютерні науки»**
(шифр і назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри інтелектуальних
інформаційних систем, д-р техн. наук, проф.

_____ Ю. П. Кондратенко

« » _____ 20 р.

З А В Д А Н Н Я
на виконання кваліфікаційної роботи магістра

Дарію Артему Миколайовичу

1. Тема кваліфікаційної роботи магістра «Інтелектуальна система прогнозування продажу у сфері електронної комерції».

Керівник роботи Болюбаш Надія Миколаївна, канд. пед. наук, доцент.

Затв. наказом Ректора ЧНУ ім. Петра Могили від «___» _____ 2023 р. № _____

2. Строк подання студентом роботи 21 лютого 2024 р.

3. Вхідні (початкові) дані до роботи: загальні відомості про мережеві сервіси у сфері електронної комерції, база даних транзакцій онлайн магазину комерційної фірми, яка спеціалізується на продажах запчастин до автомобілів Tesla.

Очікуваний результат роботи: інтелектуальна система у сфері онлайн продажу автозапчастин, де передбачено збір і аналіз інформації у процесі її функціонування та прогнозування попиту на товари для прийняття рішень з планування збуту й замовлень асортименту продукції.

4. Перелік питань, що підлягають розробці (зміст пояснювальної записки):

– аналіз предметної сфери електронної комерції та дослідження теоретичних засад прогнозування продажу товарів;

– обґрунтування вибору інструментальних засобів розробки системи прогнозування продажу;

– розробка та здійснення програмної реалізації системи прогнозування продажу запчастин до автомобілів, дослідження якості прогнозної моделі та точності прогнозу.

5. Перелік графічного матеріалу: презентація, рисунки, таблиці.

6. Завдання до спеціальної частини: охорона праці при роботі з екранними пристроями.

7. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис
Спеціальна частина з охорони праці	д-р біол. наук, професор Л. І. Григор'єва	
Методична частина	канд. пед. наук, доцент Н. М. Болюбаш	

Керівник роботи канд. пед. наук, доцент Болюбаш Н. М.
(наук. ступінь, вчене звання, прізвище та ініціали)

_____ (підпис)

Завдання прийнято до виконання Дарій А. М.
(прізвище та ініціали)

_____ (підпис)

Дата видачі завдання « ____ » _____ 2023 р.

КАЛЕНДАРНИЙ ПЛАН

виконання кваліфікаційної роботи магістра

Тема: «Інтелектуальна система прогнозування продажу у сфері електронної комерції»

№	Найменування роботи	Початок	Закінчення	Примітки
1.	Визначення керівника і теми КРМ. Подання заяви на затвердження теми КРМ	01.09.2023	19.10.2023	Виконано
2.	Отримання завдання на виконання КРМ	20.10.2023	29.10.2023	Виконано
3.	Складання календарного плану	30.10.2023	5.11.2023	Виконано
4.	Огляд літератури за темою дослідження. Аналіз існуючих підходів до побудови прогнозних моделей та мережесервісів з онлайн продажу запчастин до автомобілів	5.11.2023	25.11.2023	Виконано
5.	Проходження переддипломної практики, збір та аналіз матеріалів до КРМ	27.11.2023	24.12.2023	Виконано
6.	Аналіз предметної області та розробка технічного завдання	25.12.2023	27.12.2023	Виконано
7.	Проектування та програмна реалізація системи прогнозування для комерційної фірми з продажу автозапчастин	28.12.2023	15.01.2024	Виконано
8.	Робота над розділами фахової частини КРМ	16.01.2024	24.12.2024	Виконано
9.	Розробка методичної частини КРМ та спеціальної частини з охорони праці	25.01.2024	01.02.2024	Виконано
10.	Обговорення отриманих результатів з керівником та попередній захист КРМ	29.01.2024	3.02.2024	Виконано
11.	Корегування роботи за результатами попереднього захисту	4.02.2024	6.02.2024	Виконано
12.	Остаточне оформлення пояснювальної записки та слайдів доповіді до захисту	7.02.2024	9.02.2024	Виконано
13.	Подання рецензенту та рецензування КРМ	9.02.2024	12.02.2024	Виконано
14.	Подання КРМ, її електронної копії та інших документів (відгуку, рецензії) до захисту	15.02.2024	19.02.2024	Виконано
15.	Захист КРМ перед ЕК	27.02.2024	27.02.2024	Виконано

Розробив студент Дарій А. М.

(прізвище та ініціали)

(підпис)

Керівник роботи канд.пед.н., доц. Болубаш Н.М.

(наук. ступінь, вчене звання, прізвище та ініціали)

(підпис)

« » 2023 р.

АНОТАЦІЯ

до кваліфікаційної роботи магістра
студента групи 601 ЧНУ ім. Петра Могили

Дарія Артема Миколайовича

на тему: **«ІНТЕЛЕКТУАЛЬНА СИСТЕМА ПРОГНОЗУВАННЯ ПРОДАЖУ
У СФЕРІ ЕЛЕКТРОННОЇ КОМЕРЦІЇ»**

Кваліфікаційна робота магістра присвячена розробці та здійсненню програмної реалізації системи прогнозування продажу у сфері електронної комерції. Що є актуальним в умовах розширення сфери онлайн торгівлі, оскільки підвищує ефективність електронної комерції шляхом зменшення навантаження на менеджерів при складанні бізнес-планів, плануванні попиту на товари та поповнення асортименту, роботі з залишками товарів.

Об'єкт дослідження – процес купівлі та продажу товарів через мережу Інтернет.

Предмет дослідження – програмні засоби та методи аналізу продажу і побудови прогнозних моделей у сфері електронної комерції.

Мета дослідження – підвищення ефективності системи прогнозування продажу товарів у сфері електронної комерції на основі методів та технологій прогнозування часових рядів.

Кваліфікаційна робота магістра складається з фахової, методичної і спеціальної частини з охорони праці. Пояснювальна записка фахової частини кваліфікаційної роботи складається зі вступу, трьох розділів, висновків та додатків. У першому розділі здійснено аналіз предметної сфери електронної комерції та розкрито теоретичні засади прогнозування продажу товарів. У другому розділі обґрунтовано вибір технологій і засобів розробки системи. У третьому розділі описано проектування та програмну реалізацію системи прогнозування продажу для комерційної фірми з продажу автозапчастин. У спеціальній частині з охорони праці розглядаються питання охорони праці та безпеки у надзвичайних ситуаціях.

Кваліфікаційна робота магістра містить ___ сторінку (без додатків), ___ рисунків, ___ таблиці, ___ джерел, ___ додатки.

Ключові слова: прогнозування продажу, часовий ряд, тренд, авторегресійна модель, методи прогнозування часових рядів.

ABSTRACT

to the master's qualification work
by the student of the group 601 of Petro Mohyla Black Sea National University

Dariya Artema Mykolayovycha

on the subject: «**INTELLIGENT SALES FORECASTING SYSTEM
IN THE SPHERE OF ELECTRONIC COMMERCE**»

The master's thesis is devoted to the development and implementation of the software implementation of the sales forecasting system in the field of electronic commerce. What is relevant in the conditions of the expansion of the sphere of online trade, as it increases the effectiveness of e-commerce by reducing the burden on managers when drawing up business plans, planning the demand for goods and replenishing the assortment, working with remaining goods.

Object of research – the process of buying and selling goods via the Internet.

Subject of research – software tools and methods of sales analysis and construction of predictive models in the field of e-commerce.

The purpose of the study is to increasing the efficiency of the system of forecasting the sale of goods in the field of electronic commerce based on methods and technologies of time series forecasting.

The master's qualification work consists of a professional, methodical and special part on labor protection. The explanatory note of the professional part of the qualification work consists of an introduction, three sections, conclusions and appendices. In the first chapter, an analysis of the subject area of electronic commerce was carried out and the theoretical principles of forecasting the sale of goods were revealed. In the second section, the choice of technologies and means of system development is substantiated. The third chapter describes the design and software implementation of a sales forecasting system for a commercial firm selling auto parts. The special part on labor protection deals with issues of labor protection and safety in emergency situations.

The master's qualification work contains ___ page (without appendices), ___ figures, ___ tables, ___ sources, ___ appendics.

Key words: sales forecasting, time series, trend, autoregression model, time series forecasting methods.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ.....	3
ВСТУП.....	4
1 ТЕОРЕТИЧНІ ЗАСАДИ ПРОНОЗУВАННЯ ПРОДАЖУ У СФЕРІ ЕЛЕКТРОННОЇ КОМЕРЦІЇ.....	7
1.1 Електронна комерція та інтернет-торгівля.....	7
1.2 Огляд існуючих методів для вирішення задачі прогнозування у електронній комерції	11
1.3 Авторегресійні моделі прогнозування.....	18
1.4 Огляд мережевих ресурсів подажу авозапчастин.....	22
1.5 Постановка задачі.....	31
Висновки до розділу 1	32
2 ІНСТРУМЕНТАЛЬНІ ЗАСОБИ РОЗРОБКИ СИСТЕМИ ПРОГНОЗУВАННЯ ПРОДАЖУ	34
2.1 Мова програмування Python	34
2.2 IDE для розробки на Python PyCharm	37
2.3 Бібліотеки Python для аналізу часових рядів та побудови прогнозних моделей.....	39
2.4 Система керування пакетами PIP	43
2.5 Засоби веброзробки системи прогнозування	45
Висновки до розділу 2	48
3 РОЗРОБКА ТА ПРОГРАМНА РЕАЛІЗАЦІЯ СИСТЕМИ ПРОГНОЗУВАННЯ ПРОДАЖУ ЗАПЧАСТИН ДО АВТОМОБІЛІВ	50
3.1 Основні етапи побудови прогнозної моделі.....	50
3.2 Попередня підготовка та обробка даних	51
3.3 Програмна реалізація побудови моделей часового ряду	53
3.4 Аналіз та прогнозування продажу.....	59
Висновки до розділу 3	65
ВИСНОВКИ.....	66

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ..... 69

ПЕРЕЛІК СКОРОЧЕНЬ

API – Application Programming Interface

AR – Autoregressive Model

ARIMA – AutoRegressive Integrated Moving Average Model

ARMA – AutoRegressive Moving Average Model

CSS – Cascading Style Sheets

HTML – Hyper Text Markup Language

MAD – Mean Absolute Derivation

MAE – Mean Absolute Error

MAPE – Mean Absolute Percentage Error

MFE – Mean Forecast Error

MPE – Mean Percentage Error

MSE – Mean Squared Error

SARIMA – Seasonal Auto-Regressive Integrated Moving Average

SARMA – Spatial AutoRegressive Moving Average Model

ВСТУП

Актуальність. Високі темпи інформатизації сучасного суспільства супроводжуються накопиченням великих обсягів цифрового контенту у сфері електронної комерції. Це ускладнює роботу менеджерів з онлайн продаж й обумовлює розвиток технологій, спрямованих на прогнозування діяльності комерційних фірм у режимі реального часу. Перспективним напрямком підвищення ефективності комерційної діяльності фірм є впровадження у роботу інтернет-магазинів систем прогнозування продажу на основі аналізу даних про динаміку зміни обсягів продаж товарів протягом певного періоду часу.

Бурхливий ріст глобальних продаж через мережу Інтернет спричинив до впровадження у глобальні мережеві сервіси онлайн продаж інтелектуальних алгоритмів аналізу цифрового контенту, які використовують для підвищення ефективності комерційної діяльності шляхом залучення клієнтів, просування товарів, підвищення прибутку, прогнозування попиту та продаж. До популярних ресурсів електронної комерції, які широко застосовують це у своїй діяльності, відносять Amazon, eBay, iTunes, найбільший український онлайн-ритейлер Rozetka. Проте Інтернет-магазинах більш дрібних представників роздрібної торгівлі та торгівлі автозапчастинами зокрема такий підхід представлений недостатньо, що вимагає їх вдосконалення. Тому є потреба у створенні системи прогнозування продажу для Інтернет-магазину, який займається продажом запчастин до автомобілів.

Основним підходом до побудови прогнозу є моделювання часових рядів обсягів продажу за певний період часу. Моделювання фінансових часових рядів є складною задачею, оскільки комерційна діяльність містить багато чинників, які можуть викликати ріст та падіння попиту на товари та їх вартість. На вибір методів аналізу та прогнозування впливає стаціонарність та не стаціонарність часового ряду. До методів аналізу часових рядів, які базуються на припущенні про стаціонарність ряду, відносять авторегресійну модель (AR-модель) та її

модифікації. Для підвищення ефективності планування онлайн продаж інтернет-магазину необхідна автоматизована система, яка дозволяє здійснювати аналіз динаміки обсягів продаж та будувати модель часового ряду з високою точністю прогнозу.

Це обумовило **мету роботи**, яка полягає у підвищенні ефективності системи прогнозування продажу товарів у сфері електронної комерції на основі методів та технологій прогнозування часових рядів.

Досягнення поставленої мети обумовлює необхідність вирішення наступних **завдань**:

- здійснити аналіз предметної сфери електронної комерції та дослідити теоретичні засади прогнозування продажу товарів;
- обґрунтувати вибір інструментальних засобів розробки системи прогнозування продажу;
- розробити та здійснити програмну реалізацію системи прогнозування продажу запчастин до автомобілів, дослідити якість прогнозованої моделі та точність прогнозу.

Об'єктом дослідження є процес купівлі та продажу товарів через мережу Інтернету.

Предметом дослідження є програмні засоби та методи аналізу продажу і побудови прогнозних моделей у сфері електронної комерції.

Методологічною основою дослідження є загальнонаукові та статистично-аналітичні методи, які дозволили комплексно вивчити предмет та об'єкт дослідження, дослідити основні підходи до побудови прогнозної моделі обсягу продаж у сфері електронної комерції.

Наукова новизна одержаних результатів дослідження полягає у тому, що автором: запропоновано та обґрунтовано напрями вдосконалення отримання прогнозу продаж з онлайн продажу автозапчастин; одержали подальший розвиток підходи до підвищення ефективності роботи у комерційній фірмі шляхом зменшення навантаження на менеджерів при плануванні попиту на товари.

Результати дослідження обговорювалися на XXVI Всеукраїнської науково-практичної конференції «Могилянські читання – 2023: Досвід та тенденції розвитку суспільства в Україні: глобальний, національний та регіональний аспекти» та отримали схвалення (6-10 листопада 2023 року) та отримали схвалення.

Практичне значення отриманих результатів полягає в тому, що сформульовані теоретичні положення та практичні рекомендації щодо підвищення ефективності комерційної діяльності мали практичне застосування. Розроблена система прогнозування продажу була впроваджено у роботу магазину, який здійснює продаж запчастин до автомобілів.

Структура кваліфікаційної роботи магістра. Відповідно до мети, завдань і предмета дослідження кваліфікаційна робота магістра містить основну, методичну та спеціальну частини. Основна частина кваліфікаційної роботи магістра складається із вступу, трьох розділів, висновку, списку використаних джерел та __ додатків. Загальний обсяг кваліфікаційної роботи магістра – __ сторінок, із них тексту основної частини – __ сторінок, методичної частини – __ сторінок, спеціальної – __ сторінок. Кількість використаних джерел – __.

1 ТЕОРЕТИЧНІ ЗАСАДИ ПРОНОЗУВАННЯ ПРОДАЖУ У СФЕРІ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

1.1 Електронна комерція та інтернет-торгівля

Електронна комерція (англ. e-commerce) — це сфера цифрової економіки, яка включає всі фінансові та торгові транзакції, що здійснюються за допомогою комп'ютерних мереж, і бізнес-процеси, пов'язані з проведенням таких транзакцій.

Електронна комерція, увійшла в наше життя як технологічний феномен, що змінив парадигму торгівлі та взаємодії споживача з ринком. Закладена в основу глобальних технологічних трансформацій, електронна комерція стала не лише каналом обміну товарів і послуг, а й ключовим фактором формування нового економічного ландшафту. Визначаючи електронну комерцію, спрямовують увагу на комплексність процесу купівлі та продажу через мережу Інтернет. Це явище починається зі зміни способів життя та споживацьких звичок, адаптуючись під впливом технологій, що перепрофілюють весь економічний ланцюг, починаючи від виробництва і закінчуючи обслуговуванням споживачів [1].

Завдяки електронній комерції з'явилася можливість купувати та продавати будь-які товари та послуги в онлайн-режимі. Це відкрило нові горизонти для бізнесу та надало споживачам необмежений доступ до різноманіття товарів і послуг, навіть без фізичного виходу з дому. Під тиском стрімкого технологічного розвитку, електронна комерція перетворилася з експериментального явища в справжній катализатор глобальних змін у сучасному бізнесі [2].

Сучасні технологічні та соціокультурні зрушення в електронній комерції свідчать про неперервний розвиток і стрімкий ріст. Від перших етапів передачі даних до новітніх систем штучного інтелекту, електронна комерція продовжує трансформуватися, адаптуючись до потреб і вимог сучасного світу [3].

Електронна комерція виявляється надзвичайно різноманітною, охоплюючи різні типи операцій та взаємодій між різними учасниками ринку. Розглянемо

ключові типи електронної комерції, які формують ландшафт цього цифрового економічного простору.

1. B2C (Бізнес-Споживач). Цей тип електронної комерції включає в себе транзакції між компанією і кінцевим споживачем. Інтернет-магазини, онлайн-сервіси та платформи для кінцевих споживачів є прикладами B2C електронної комерції. Завдяки цьому, споживачі мають можливість швидко і зручно здійснювати покупки, порівнювати ціни та здійснювати онлайн-оплати.

2. B2B (Бізнес-Бізнес). У цьому сценарії електронної комерції, операції здійснюються між двома бізнесами. Компанії можуть заключати угоди, здійснювати постачання та обмін послугами через онлайн-платформи. B2B електронна комерція є стратегічним інструментом для оптимізації ланцюга постачань та забезпечення ефективної взаємодії між компаніями.

3. C2C (Споживач-Споживач). Цей тип включає в себе транзакції між самими споживачами. Онлайн-ринки, де один користувач може продавати товари чи послуги іншому користувачеві, є прикладом C2C електронної комерції. Платформи, що об'єднують індивідів для обміну товарами або послугами, стають динамічними майданчиками для кінцевих користувачів.

4. M-Commerce (Мобільна Комерція). Зі зростанням використання мобільних пристроїв, мобільна комерція стала значущою складовою електронної комерції. Додатки та оптимізовані сайти для смартфонів дозволяють споживачам здійснювати покупки в русі, використовуючи свої мобільні пристрої для доступу до товарів та послуг.

5. P2P (Peer-to-Peer). У цьому варіанті електронної комерції, учасники обмінюються товарами чи послугами без прямої участі бізнес-структур. P2P платформи стають посередниками для споживачів, які бажають обмінюватися товарами чи послугами один з одним.

Ці типи електронної комерції створюють різноманітний ландшафт, в якому різні сторони ринку можуть взаємодіяти. Напрямок розвитку електронної комерції

буде визначати те, як споживачі та бізнеси використовуватимуть ці можливості в майбутньому.

За прогнозами зростання продажів в електронній комерції досягне разючої цифри 265%. У 2017 році роздрібні продажі в секторі електронної торгівлі по всьому світу досягли 2,3 трильйона дол. США, а до 2021 року, лише через чотири роки, за прогнозами, виручка складе 4,88 трильйона \$. Ведення електронної комерції супроводжується накопиченням великих обсягів наборів даних, які можна використовувати для вирішення бізнес-завдань, що раніше здавались надто складними.

В умовах стрімкого зростання ринку інтернет-комерції інструменти НСМ стають все більш важливими та навіть необхідними, оскільки вони дають компаніям можливість:

- відрізнитися від конкурентів;
- охоплювати велику кількість замовників у багатьох регіонах країни та світу;
- скорочувати витрати за рахунок прямого продажу та утримання меншої кількості звичайних магазинів;
- надати можливість замовникам купувати те, що вони хочуть, у будь-який час із будь-якого місця за допомогою будь-якого цифрового пристрою – це те, що важливо для мільйонелів та інших прихильників цифровізації;
- отримати цінні дані про замовників за допомогою онлайн-метрик;
- протестувати на ринку нові продукти, послуги, бренди та проекти з мінімальними початковими інвестиціями;
- надати замовникам можливість самообслуговування, що дозволить меншій кількості співробітників відділу продажу обслуговувати більше замовників;
- швидко здійснювати масштабування без зайвих витрат.

Щоб відповідати зростаючим потребам замовників у плані збільшення вибору, спрощення доступу та прискорення поставок, компанії інтегрують пропозиції звичайних магазинів та електронної комерції для створення

універсальної багатоканальної взаємодії із замовниками під час покупок, коли останні можуть:

- 1) вивчати та дивитися продукти та послуги до здійснення покупки в онлайн-режимі або у звичайному магазині;
- 2) замовляти товари в магазині через інтерактивні кіоски, персональні консьєрж-служби та отримувати інші пропозиції до здійснення покупки онлайн або у звичайному магазині;
- 3) використовувати будь-який пристрій - комп'ютер, смартфон, розумний годинник, цифровий помічник тощо щоб здійснювати покупки;
- 4) отримувати персональні рекомендації, купони та інші онлайн-пропозиції на основі інформації, зібраної онлайн або у звичайному магазині;
- 5) оформляти доставку товарів туди, куди їм хочеться (додому або до найближчого магазину), часто наступного дня або через день;
- 6) замовляти товари онлайн з іншого магазину, коли в даному магазині немає потрібного товару (за стилем, розміром, кольором тощо).

Електронна комерція швидко розвивається, у міру того, як інновації в штучному інтелекті, машинному навчанні та Інтернеті речей (IoT) допомагають компаніям по-новому задовольняти зростаючі вимоги замовників та комерційних фірм.

У даному дослідженні було здійснено розробку системи прогнозування продажу для інтернет-магазину роздрібної торгівлі запчастинами до автомобілів. У даному випадку ми маємо справу з моделлю електронної комерції B2C – бізнес для споживача. Однак для ефективної діяльності менеджерів інтернет-магазину доцільно застосувати аналіз накопиченої у процесі комерційної діяльності інформації стосовно продаж для відслідковування тенденцій у попиті на них та прогнозування продаж. Це дасть можливість ефективніше планувати поповнення асортименту товарів та вирішувати маркетингові завдання з просуванням товарів на ринку.

1.2 Огляд існуючих методів для вирішення задачі прогнозування у електронній комерції

Предиктивна аналітика – це тип аналітики даних, у якому майбутні результати прогнозуються за допомогою історичних даних і таких аналітичних методів, як машинне навчання та статистичне моделювання [4]. Будь-яка організація тепер може надійно прогнозувати тенденції та поведінку на мілісекунди, дні або роки в майбутньому за допомогою складних інструментів і моделей предиктивної аналітики. Предиктивна аналітика була підтримана багатьма організаціями; згідно з опублікованим звітом 2017 року, прогнозується, що світовий ринок досягне приблизно 10,95 млрд доларів США до 2022 року, а сукупний річний темп зростання (CAGR) буде складати близько 21 відсотка в період з 2016 по 2022 рік. фірми Zion Market Research [5].

Для предиктивної аналітики використовується широкий спектр методів і технологій, таких як великі дані, інтелектуальний аналіз даних, статистичне моделювання, машинне навчання та різноманітні математичні методи та моделі. Прогнозна аналітика використовується у електронній комерції для аналізу поточних і історичних даних, щоб визначити тенденції та прогнозувати події та умови на основі наданих параметрів [6].

Предиктивна аналітика допомагає знайти і використовувати шаблони в даних для виявлення ризиків і можливостей. Наприклад, моделі можуть бути розроблені для виявлення, як різні фактори поведінки пов'язані один з одним. Такі моделі дозволяють інформоване ухвалення рішень у різних категоріях ланцюжків поставок і закупівель, оцінюючи або обіцянку, або ризик, пов'язаний з певним набором умов.

На відміну від попередніх методів, прогнозна аналітика робить прогнози на майбутнє більш точними та надійними. Роздрібні продавці часто використовують прогносні моделі для прогнозування потреб у товарних запасах, управління графіками відвантаження та налаштування макетів магазинів для максимізації

продажів. Компанії можуть оптимізувати свою маркетингову політику за допомогою прогнозного аналізу, щоб створити нові відгуки або покупки, а також збільшити можливості перехресних продажів. Компанії можуть залучати, утримувати та збільшувати кількість своїх найцінніших клієнтів за допомогою прогнозних моделей. Предиктивна аналітика також може використовуватися для виявлення вивчення поведінки та дій користувачів [7].

Інструменти предиктивної аналітики дозволяють в режимі реального часу отримувати детальну інформацію про безліч бізнес-операцій. Багато різних типів шаблонів і поведінки можна прогнозувати за допомогою цих інструментів. Наприклад, коли поповнювати запаси, як розподіляти ресурси в певний час або коли найкраще запускати рекламну кампанію на основі прогнозів, зроблених на основі аналізу даних за певний період часу.

Прихильники предиктивної аналітики зазвичай використовують інструменти, надані одним або кількома зовнішніми розробниками. Багато з цих програмних засобів адаптовано до потреб конкретних комерційних компаній і відділів. До основних постачальників прогнозного аналізу відносять: Acxiom, IBM, Information Builders, Microsoft, SAP, SAS Institute, Tableau Software, Teradata, TIBCO Software.

Основою прогносної аналітики є моделі – шаблони, які дозволяють створювати корисні результати з минулих і поточних даних, що призводить до позитивних довгострокових результатів. Для побудови моделей предиктивної аналітики можуть використовувати багато методів прогнозного моделювання: регресія, дерева рішень, нейронні мережі, часові ряди тепер широко підтримуються на великій кількості платформ [8].

Прогнозна аналітика більше не обмежується математичними та статистичними моделями, оскільки програмне забезпечення стає все більш інтерактивним і простим у використанні. Бізнес-експерти та аналітики в сфері електронної комерції також використовують ці технології [9].

Аналітика сприяє розвитку Big Data: компанії збирають велику кількість даних про клієнтів у реальному часі, а аналітика використовує ці історичні дані та огляди клієнтів для прогнозування майбутніх подій. Великі дані, як збережені, так і в режимі реального часу, можуть допомогти компаніям перейти з історичної перспективи на перспективу замовника за допомогою передбачуваної аналітики. Наприклад, інтернет-магазини можуть передбачити акції, у яких клієнти найбільше беруть участь або купують у майбутньому, аналізуючи їхні минулі покупки.

Роздрібна торгівля в Інтернеті розвивається стрімкими темпами, і клієнти активно шукають більш привабливий досвід покупок. Підприємства електронної комерції повинні мати можливість бути на крок попереду своїх клієнтів, щоб досягти успіху на високодинамічному ринку. Вони повинні бути в змозі прогнозувати, що покупці шукають у своєму онлайн-магазині. Завдяки ефективному аналізу настроїв клієнтів щодо ціноутворення, прогнозна аналітика також може допомогти компаніям електронної комерції визначити оптимальні ціни на свої товари. Підприємства електронної комерції можуть використовувати потенціал прогностичної предиктивної аналітики, щоб надавати додаткові поради щодо товарів і акцій.

Інтернет-магазини докладають багато зусиль, щоб покращити можливості рекомендацій продуктів на своїх платформах електронної комерції, щоб гарантувати позитивний коефіцієнт конверсії в своїх магазинах електронної комерції. Інтернет-роздрібні торговці можуть використовувати потенціал аналітики, щоб отримати точні дані про окремих клієнтів. Це допоможе їм надавати цільові рекомендації щодо товарів, які базуються на аналізі минулих історій покупок, моделях перегляду магазинів і найпопулярніших товарів у конкретному діапазоні цін. Таким чином, підприємства електронної комерції можуть використовувати предиктивну аналітику для швидкого прийняття важливих рішень, які впливають на результативність. Компанії електронної комерції також повинні регулярно перевіряти моделі предиктивної аналітики, щоб зменшити ймовірність помилок в аналізі даних.

Одним із основних підходів до побудови прогнозу у сфері електронної комерції є дослідження часових рядів, яке включає в себе вивчення динаміки показників комерційної діяльності в часі та їх аналіз з метою прогнозування майбутніх значень. Цей напрямок досліджень має велике значення для управлінської діяльності комерційних компаній, оскільки відомості про тенденції зміни різних показників, таких як ціни на товари та послуги, обсяги продажів та інші, є важливими для прийняття рішень про інвестування, ризик-менеджмент та планування діяльності підприємств.

Основна мета досліджень у галузі часових рядів полягає у вивченні їх структури, тренду, сезонності, циклів та інших характеристик, що дозволяє зробити висновки про тенденції зміни відповідних показників стосовно комерційної діяльності інтернет-магазинів.

Часові ряди – це послідовності даних, які збираються відповідно до певної часової структури, наприклад щоденно, щотижня або щомісяця. Дані можуть стосуватися цін на товари, обсягів продажу та інших показників. Часові ряди дозволяють досліджувати динаміку зміни показників у часі, аналізувати тренди та сезонність, виявляти кореляції між різними показниками та робити прогнози на майбутнє. Аналізуючи динаміку обсягів продажу певної компанії за допомогою часових рядів, можна визначити потенційні ризики та можливості для планування її подальшої діяльності [10].

Часовий ряд є послідовністю значень обсягів продажу, упорядковану у хронологічному порядку: y_1, y_2, \dots, y_n , де y_t – значення на певний момент часу. До основних компонентів часового ряду відносять [11]:

1) *тренд* – основний структурний компонент, який характеризує наявність загального напрямку зміни обсягів продажу, що формується під впливом загальних та довгострокових тенденцій, які впливають на часовий ряд та який можна представити у вигляді функції $f(t)$;

2) *сезонна компонента* – містить значення, які повторюються відносно основної тенденції протягом невеликих періодів часу: року, місяця або тижня;

3) *циклічна компонента* – містить коливання відносно основної тенденції – тренду протягом великого періоду часу, які можуть бути пов'язані з економічними чи іншими циклічними явищами, такими як бізнес-цикли;

4) *випадкова (залишкова) компонента* – складова часового ряду, обумовлена впливом випадкових факторів або пов'язаних з іншими чинниками, які не враховуються в моделі.

Розрізняють також стаціонарні та нестаціонарні часові ряди. Виходячи з цього, стаціонарність та не стаціонарність – це важливі поняття в аналізі часових рядів, які впливають на вибір методів аналізу та прогнозування [12].

До *стаціонарних часових рядів* відносять такі, значення рівнів яких коливається навколо постійного середнього значення, тому стаціонарні часові ряди не мають тренду, сезонної та циклічної складової. Багато статистичних методів аналізу часових рядів базуються на припущенні про стаціонарність ряду. Наприклад, авторегресійна модель (AR-модель) припускає, що ряд є стаціонарним.

Нестационарний часовий ряд – це ряд, значення рівнів якого коливаються навколо середнього значення ознаки, яке з плином часу змінюється. Такі ряди мають тренд, сезонність та/або циклічні зміни в середньому значенні. Підхід, який базується на виокремленні складових часового ряду, включає різні способи фільтрації шуму та чітке виокремлення регулярних компонент [13].

При побудові прогностичної моделі нестаціонарного часового ряду застосовують такі підходи: моделювання компонент у сукупності; розкладання часового ряду на компоненти та моделювання кожної компоненти окремо.

Серед методів аналізу та прогнозування часових рядів можна виділити наступні [14]: 1) методи на основі виокремлення компонент шляхом побудови адитивної чи мультиплікативної моделі; 2) методи на основі механічного згладжування передбачають усереднення значень рівнів ряду на певному інтервалі й добре дозволяють виокремити сезонну та випадкову компоненти: ковзні середні (MA – Moving Average), експоненційні ковзні середні (EMA – Exponential Moving

Average); 3) методи побудови прогнозу на основі нейронних мереж; 4) авторегресійні моделі.

Модель МА передбачає згладжування значень часового ряду $Y_1, Y_2, \dots, Y_t, \dots, Y_n$ за формулою:

$$S_t = m_{t-2} + \frac{1}{n} (Y_{t-1} - Y_{t-2}), \quad (1.1)$$

де m – ковзне середнє за 2 періоди до прогнозного: $m_t = (Y_{t-1} + Y_t + Y_{t+1})/3$.

Модель ЕМА передбачає згладжування значень часового ряду за формулою:

$$S_t = \alpha Y_t + (1 + \alpha) \cdot S_{t-1}, \quad (1.2)$$

де $\alpha = \frac{2}{n+1}$.

Для моделювання часового ряду, який містить і тренд і сезонність, можна застосувати модель Хольта-Вінера (англ. Holt-Winters). Ця модель широко використовується для прогнозування в галузях, де важливо враховувати тренд і сезонність. Основні компоненти моделі Хольта-Вінтерса:

- рівень (Level): середнє значення часового ряду в певний момент часу, є базовим параметром, який визначається на основі історичних даних;
- тренд (Trend): враховує зміну середнього значення часового ряду в часі, може бути лінійним або експоненційним, відображаючи поступове зростання або спадання значень ряду;
- сезонність (Seasonality): враховує періодичність у часовому ряді.

Основним типом моделі Хольта-Вінтерса є модель, яка передбачає потрібне експоненційне згладжування для часового ряду:

- 1) рівняння згладжування ряду:

$$L_t = \alpha + 1 - \alpha \cdot L_{t-1} + T_{t-1}, \quad (1.3)$$

2) оцінка тренду:

$$T_t = \beta \cdot L_t - L_{t-1} + 1 - \beta \cdot T_{t-1}, \quad (1.4)$$

3) оцінка сезонності на l періодів уперед:

$$S_t = \gamma \frac{y_t}{L_t} + 1 - \gamma \cdot S_{t-s}, \quad (1.5)$$

4) рівняння згладжування ряду:

$$y_{t+l} = L_t + l \cdot T_t \cdot S_{t-s+l}, \quad (1.6)$$

де s – період сезонності, α, β, γ - параметри згладжування.

Ключові аспекти цієї моделі Хольта-Вінтерса моделі:

– визначення параметрів: параметри моделі, такі як ступінь згладжування (smoothing parameter) для рівня, тренду та сезонності, можуть бути визначені шляхом аналізу історичних даних;

– підгонка моделі та прогнозування: після визначення параметрів модель підганяється до даних, після чого може бути використана для прогнозування майбутніх значень часового ряду;

– адаптивність: модель є адаптивною, що дозволяє враховувати зміни в структурі ряду та автоматично адаптуватися до них при прогнозуванні.

Модель Хольта-Вінтерса є ефективним інструментом для аналізу та прогнозування часових рядів з урахуванням тренду, сезонності та зміни середнього значення. Вона широко використовується в практичних задачах прогнозування для різних галузей економіки та бізнесу.

1.3 Авторегресійні моделі прогнозування

Авторегресійна модель (англ. AutoRegressive Model, AR) є моделлю стаціонарного часового ряду, у якій значення ряду у певний момент часу лінійно залежить від попередніх значень цього ж ряду.

Для часового ряду $Y_1, Y_2, \dots, Y_t, \dots, Y_n$ авторегресійний процес порядку P (AR(p)-процес) задається рівнянням:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + e_t, \quad (1.7)$$

де a_i – коефіцієнти авторегресії, зазвичай $a_0 = 0$, e_t – випадкова величина.

У авторегресійній моделі поточне значення часового ряду Y_t є лінійною комбінацією попередніх значень та випадкової похибки.

Основні етапи побудови авторегресійної моделі є наступними.

1. Перевірка часового ряду на стаціонарність. Якщо ряд виявився нестаціонарним, здійснюють перетворення, які приводять його до стаціонарного виду.

2. Ідентифікація порядку авторегресійної моделі P . Її можна здійснити шляхом аналізу послідовності коефіцієнтів автокореляції часового ряду – корелограми. Авторегресійний часовий ряд має плавно гасаючу автокореляційну функцію. На практиці порядок авторегресійної моделі P частіше всього не є більшим за 2.

3. Оцінка параметрів авторегресійної моделі – коефіцієнтів авторегресії, здійснюється за допомогою методу найменших квадратів або інших методів.

4. Перевірка адекватності побудованої моделі. Якщо рівняння авторегресії для моделювання часового ряду підібране правильно, то випадкова компонента є білим шумом.

5. Застосування моделі для прогнозування.

Авторегресійні моделі не призначені для опису часових рядів із трендом, однак їх використовують для опису коливань. При аналізі нестационарних часових рядів авторегресійна модель об'єднується з іншими методами аналізу, що суттєво розширює сферу їх застосування. Одним із способів покращення точності AR-моделі є поєднання її з методами механічного згладжування, такими як метод експоненційного згладжування (ES-модель) та метод низької частоти (LLF-модель).

Модель ARMA(p,q) (англ. AutoRegressive Moving Average Model) є об'єднанням авторегресійної моделі AR(p) з моделлю ковзного середнього (MA). Модель ARMA може бути виражена як ARMA(p, q), де p – порядок авторегресії (AR), а q – порядок ковзних середніх (MA).

Основні компоненти моделі ARMA:

1. Autoregressive (AR) компонент: Ця складова виражає залежність між поточним значенням часового ряду та попередніми значеннями в часі. AR-складова включає в себе параметр p, який вказує на кількість попередніх значень, використаних для прогнозування поточного значення;

2. Moving Average (MA) компонент: Ця складова моделює залежність між поточним значенням часового ряду та попередніми значеннями шуму або помилок моделі. MA-складова включає в себе параметр q, який вказує на кількість попередніх значень шуму, використаних для прогнозування поточного значення.

Ключові аспекти моделі ARMA:

– стаціонарність часового ряду: аналіз ARMA передбачає, що часовий ряд є стаціонарним, тобто має сталий середній та дисперсію. Якщо ряд не є

стаціонарним, може бути використана процедура диференціювання для зроблення його стаціонарним;

– оцінка параметрів моделі: параметри моделі ARMA (р і q) можуть бути визначені за допомогою аналізу автокореляційної та часткової автокореляційної функцій часового ряду;

– підгонка моделі та прогнозування: після визначення параметрів моделі ARMA вона підганяється до навчальних даних і може бути використана для прогнозування майбутніх значень часового ряду.

Модель ARMA є корисним інструментом для аналізу та прогнозування стаціонарних часових рядів із досить простою структурою. Однак, вона може бути менш ефективною у випадках, коли часовий ряд має складну структуру або сезонність. У таких випадках модель ARIMA або модифікації ARIMA можуть бути більш придатними для аналізу та прогнозування.

Модель ARIMA(p,d,q) (англ. AutoRegressive Integrated Moving Average Model) є розширенням моделі ARMA(p,q) для нестаціонарних часових рядів, які перетворюють у стаціонарні з порядком інтеграції d . Тому її можна застосовувати для прогнозу нестаціонарних часових рядів. Порядок інтеграції d – це число, яке показує, скільки раз необхідно застосувати оператор перших різниць, щоб часовий ряд став стаціонарним.

Оператор перших різниць – взяття різниць першого порядку від часового ряду Y_t з утворенням ряду перших різниць: $\nabla^1 y_t = y_t - y_{t-1}$.

Модель ARIMA (Autoregressive Integrated Moving Average) є однією з основних моделей у часових рядах для прогнозування. Вона дозволяє моделювати складність та залежності в часових рядах, узагальнюючи ідеї авторегресійних (AR) та ковзних середніх (MA) моделей. Основні компоненти моделі ARIMA:

– Autoregressive (AR) компонент: визначає залежність між поточним значенням та попередніми значеннями в часовому ряді, описується параметром р, який вказує на кількість попередніх значень, використаних для прогнозу;

– Integrated (I) компонент: вказує на кількість разів, які потрібно взяти різницю між спостереженнями для зроблення часового ряду стаціонарним, цей параметр позначається як d ;

– Moving Average (MA) компонент: виражається через залежність між поточним значенням та попередніми значеннями помилок моделі прогнозування, описується параметром q , який вказує на кількість попередніх помилок, використаних для прогнозу.

Загальна формула моделі ARIMA записується як $ARIMA(p, d, q)$. Ця модель може бути використана для прогнозування часових рядів зі стаціонарними або нестаціонарними даними. Інші важливі аспекти моделі ARIMA:

– стаціонарність часового ряду: для застосування ARIMA, часовий ряд повинен бути стаціонарним, якщо ряд не є стаціонарним, може бути застосована процедура диференціювання для зроблення його стаціонарним;

– визначення параметрів моделі: параметри p , d , q можуть бути визначені шляхом аналізу автокореляційної функції (ACF) та часткової автокореляційної функції (PACF) для часового ряду;

– підгонка моделі та прогнозування: після визначення параметрів моделі ARIMA вона підганяється до вихідних даних, а потім може бути використана для прогнозування майбутніх значень часового ряду.

SARIMA (Seasonal Autoregressive Integrated Moving Average) є розширенням моделі ARIMA для аналізу та прогнозування сезонних часових рядів. Вона включає в себе основні компоненти моделі ARIMA: авторегресії (AR), диференціації (I) і ковзного середнього (MA) для сезонної складової ряду разом з додатковими параметрами, що враховують сезонність. Основні компоненти моделі SARIMA:

– авторегресія (AR): виражає залежність між поточним значенням часового ряду та попередніми значеннями в часі, аналогічно до моделі ARIMA;

– інтегрування (I): виражає кількість різниць, які були застосовані до часового ряду для зроблення його стаціонарним. Інтегрування відбувається шляхом диференціювання часового ряду;

– ковзні середні (MA): моделює залежність між поточним значенням часового ряду та попередніми значеннями шуму або помилок моделі, аналогічно до моделі ARIMA;

– сезонність (S): враховує періодичність у часовому ряді. Вона включає в себе додаткові параметри, що визначають сезонність, такі як період сезонності та порядок авторегресії та ковзних середніх для сезонних компонент.

Модель SARIMA може бути виражена як SARIMA(p, d, q)(P, D, Q, s), де (p, d, q) відповідають параметрам ARIMA, (P, D, Q) - параметрам сезонної ARIMA, а s - період сезонності. Ключові аспекти моделі SARIMA [15]:

– визначення параметрів: параметри моделі можуть бути визначені шляхом аналізу автокореляційних та часткових автокореляційних функцій як для звичайної, так і для сезонної складових часового ряду;

– підгонка моделі та прогнозування: після визначення параметрів модель SARIMA підганяється до навчальних даних та може бути використана для прогнозування майбутніх значень часового ряду;

– сезонність та тренд: дозволяє враховувати як сезонні, так і нетривіальні тренди в часовому ряді, що робить її корисною для аналізу та прогнозування складних часових рядів зі сезонною компонентою.

Оцінка точності прогнозу здійснюється з використанням показників, наведених у таблиці 1.1.

1.4 Огляд мережевих ресурсів подажу авозапчастин

Зробимо аналіз наявних ресурсів для підтримки діяльності з продажу автомобілів та запчастин до них.

До найбільш популярних в Україні ресурсів, які надають інформацію про автомобілі та їх характеристики і містять відгуки про рівень обслуговування й сервісу відносять: AvtoBazar, Auto.Ria, RST, AutoSite, сайти-агрегатори Auto.Meta, AutoMoto, UAvto, AvtoPoisk, AutoSale. Кращі веб-сайти автосалонів – Towne Ford,

Santa Margarita Toyota, “Фалькон-Авто”, “Ю.Р.К”, орієнтовані на надання допомоги покупцям у знаходженні автомобілів, які задовольняють їх потреби, використовуючи при цьому перш за все інформаційну та маркетингову функції.

Таблиця 1. 1 – Показники оцінювання точності прогнозу

№ з/п	Показник	Опис	Формула
1.	MAE	Середня абсолютна похибка (Mean Absolute Error)	$MAE = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $
2.	MSE	Середньоквадратична похибка (Mean Squared Error)	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
3.	RMSE	Корінь квадратний із середньоквадратичної похибки (Root Mean Squared Error)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
4.	MAPE	середня абсолютна похибка у відсотках (Mean Absolute Percentage Error)	$MAPE = \frac{100}{n} \sum_{t=1}^n \left \frac{y_t - \hat{y}_t}{y_t} \right $

Ряд мережесервісів призначені для продажу автомобілів: PLATINUM AUTO (рис. 1.1, <http://pa.od.ua/car>), РІКавто (рис. 1.2, <https://rikauto.com.ua>), AutoBoutique (рис. 1.3, <http://autoboutique.cars.ua>). Вони призначені для продажу авто різних марок та країн.

Після вибору країни та марки авто, надається відфільтрований набір авто, які влаштовують запит користувача.

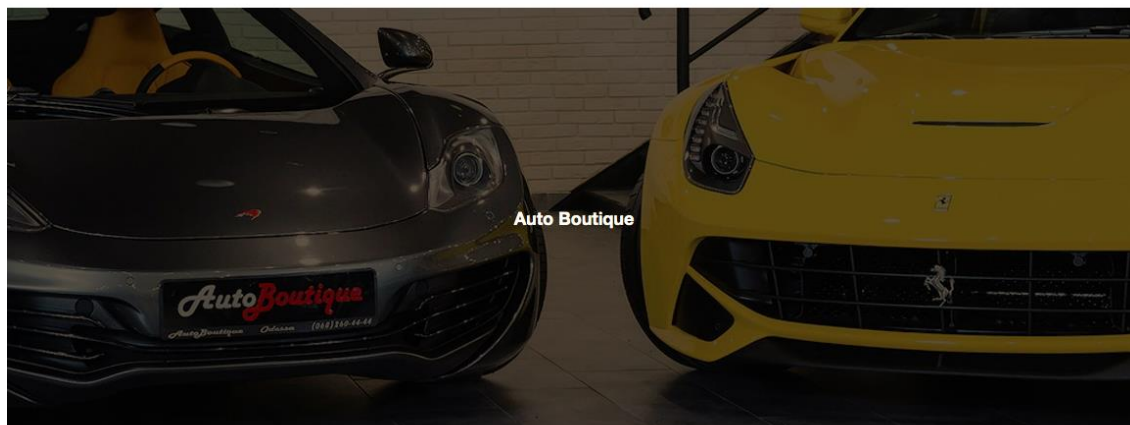
Дана функція є оптимальною та універсальною для більшості додатків даного типу. В ході аналізу даного інтернет-магазинів виявлено, що їх сервіс не здійснює продаж окремих комплектуючих до автомобілів.



Рисунок 1.1 – Сторінка ресурсу PLATINUM AUTO



Рисунок 1.2 – Сторінка ресурсу РІКавто

AutoBoutique

Все марки ▾

Сортировать по: умолчанию ▾

 Показывать проданные

Рисунок 1.3 – Сторінка ресурсу AutoBoutique

В Україні є різноманітні торгові майданчики, такі як: OLX, Ria.com, Prom.ua, Rozetka, Aukro, на яких можна придбати автозапчастини. На цих майданчиках одні люди продають товари, інші їх купляють. Найбільшим торговим майданчиком України є OLX. Не дивлячись на те, що насправді це сайт об'яв, все ж таки його можна назвати торговим майданчиком, оскільки цей ресурс має функцію доставки. У цьому випадку покупець має змогу одразу сплатити товар, як на торговому майданчику. На сайті можна опублікувати оголошення будь-який користувач, після SMS-верифікації. Товари можна виставляти, як на продаж, так і на обмін або віддати безкоштовно. Безкоштовна дошка оголошень RIA.com розширила можливості свого сайту, перетворивши його на торговий майданчик. І тепер, окрім оголошень, можна відкривати власний інтернет-магазин на платформі сайту. Найпопулярнішими напрямками цього майданчика є такі категорії: Продаж автомобілів, Продаж нерухомості та Автотовари. Розглянуті ресурси є торговими майданчиками, які не сфокусовані тільки на автозапчастинах. Їх каталог охоплює великий перелік товарів та не має вбудованих засобів прогнозування продажу.

На даний момент на українському ринку також є спеціалізовані інтернет-магазини продажу автозапчастин. Вони в більшості сфокусовані на певних типах автомобілів, але вбудовані системи прогнозування продажу у них відсутні.

Таким чином, діяльність авторинків та автосалонів розширюється за рахунок впровадження веб-ресурсів, якими широко користуються як покупці, так ті, хто продає автомобілі та запчастини до них. Однак аналіз накопиченої інформації шляхом з метою прогнозування продажу деталей автомобілів, не є реалізованим.

Використовуючи аналітичні інструменти, знайдених платформ для прогнозування продажів товарів, було виявлено недоліки. Основним недоліком платформ є їх закритість програмного коду. Це обмежує можливість використовувати та змінювати внутрішні аналітичні інструменти даних платформ для налаштування відповідних параметрів для вашої компанії.

Прогнозування продажів товарів — це дуже складна робота для людини. Розробники створили інструменти для прогнозування, які використовують нейромережеву технологію, щоб полегшити цю роботу. Існує багато систем та програмних сервісів, що дозволяють здійснювати побудову прогнозів часових рядів. Умовно їх можна поділити на кілька типів: статистичні пакети, табличні процесори, пакети візуального проектування, ERP-системи, системи класу Business Intelligence, когнітивні та академічні системи прогнозування.

MATLAB є одним із інтригуючих програмних інструментів прогнозування, які використовують технологію нечіткого моделювання, зокрема прогнозування χ . Адаптивна система нейронечіткого виведення (ANFIS) є основою для побудови штучної нейронної мережі, яка дозволяє прогнозувати майбутній стан економіки.

Для прогнозування MATLAB має виконати наступні дії: вибір даних для подальшого аналізу; вибір архітектури; визначення структури нейронної мережі; навчання мережі; тестування мережі. Існує кілька алгоритмів навчання [16].

1. Алгоритм Левенберга-Маркуарда використовується для оптимізації параметрів нелінійних регресійних моделей. Середньоквадратична помилка моделі на навчальній вибірці, ймовірно, використовується як критерій оптимізації.

Алгоритм використовує послідовне наближення заданих початкових значень параметрів до локального оптимума.

2. Алгоритм Байєсівської регуляризації мінімізує лінійну комбінацію квадратів помилок і ваг. Крім того, він змінює лінійну комбінацію так, щоб отримана мережа мала високі якості узагальнення наприкінці навчання.

3. Алгоритм сполучених градієнтів, який є частиною методу сполучених градієнтів, використовує підхід Левенберга-Марквардта для масштабування розміру кроку, щоб уникнути пошуку рядків на одну ітерацію навчання. Цей метод дозволяє уникнути трудомісткого пошуку рядків на кожен ітерацію навчання завдяки механізму масштабування розміру кроку.

Deductor містить всі інструменти, які вам потрібні, щоб вирішити складне завдання прогнозування продажів. Збір даних, створення моделей і передача результатів у сторонні системи – усе це підтримується за допомогою використання даної платформи. Дедуктор включає майже всі сучасні алгоритми, які використовуються для побудови прогностичних моделей, починаючи з найпростіших і закінчуючи найскладнішими.

Існують різноманітні рішення для автоматизації бізнес-процесів, а не просто прогнозування продажів. Deductor Demand Planning вирішує проблему планування закупівель товарів з метою підвищення оборотності та мінімізації дефіциту [17].

Завдяки створенню штучних нейронних мереж і застосуванню їх до вхідних даних програмне забезпечення GMDH Shell вирішує проблеми інтелектуального аналізу даних і прогнозування часових рядів (РІС. 1.4). Вона дозволяє легко виконувати прогнозування та аналіз даних [18]. Програма може надавати найточніші прогнози часових рядів набагато швидше, ніж звичайні штучні нейронні мережі, завдяки методу прогнозування GMDH і найсучаснішій технології паралельної обробки.

Методи прогнозування GMDH є основою GMDH Shell. Оболонка GMDH є універсальним рішенням для прогнозування «під ключ» і аналізу часових рядів. З іншого боку, широкий спектр параметричних налаштувань і широкий вибір методів

прогнозування, а також велика кількість унікальних для галузі опцій дозволяють програмі відповідати будь-якому конкретному завданню, включаючи аналіз часових рядів, прогнозування попиту, прогнозування запасів, прогнозування продажів, прогнозування фондового ринку.

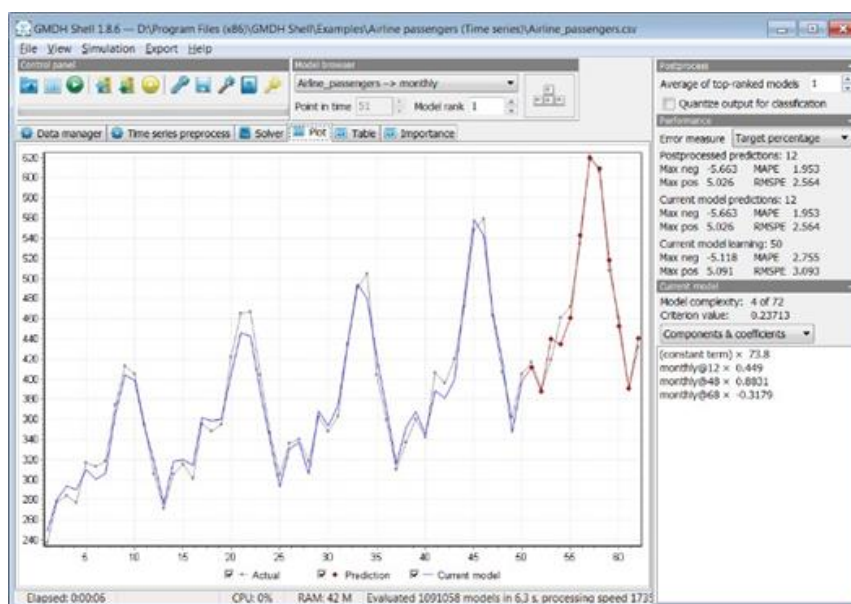


Рисунок 1.4 – Робоча область GMDH Shell

Нейромережеве прогнозування більш точне, оскільки воно більш гнучке, ніж традиційні лінійні або поліноміальні апроксимації. Експерт може використовувати нейронні мережі для пошуку та оцінки нелінійних зв'язків і відносин між даними, щоб побудувати модель-кандидат, яка має високу точність прогнозування. Крім того, GMDH Shell не дотримується абсолютної точності припасування та не вимагає попередньої нормалізації даних, що значно скорочує час обчислень. Отримання точних прогнозів на спорт, бізнес або біржу не вимагає багато часу чи зусиль, оскільки GMDH Shell тренує нейронні мережі та використовує їх для аналізу. GMDH Shell спрямовує всі вільні ресурси вашого ПК на нейромережевий аналіз завдяки унікальній системі балансування навантаження центрального процесора. Результати є більш точними та швидкими, ніж раніше.

Однією з найкращих і найефективніших нейромережевих програм є

автоматизована нейромережа STATISTICA (SANN). ВОНА пропонує широкий спектр можливостей і безліч унікальних переваг (рис. 1.5).

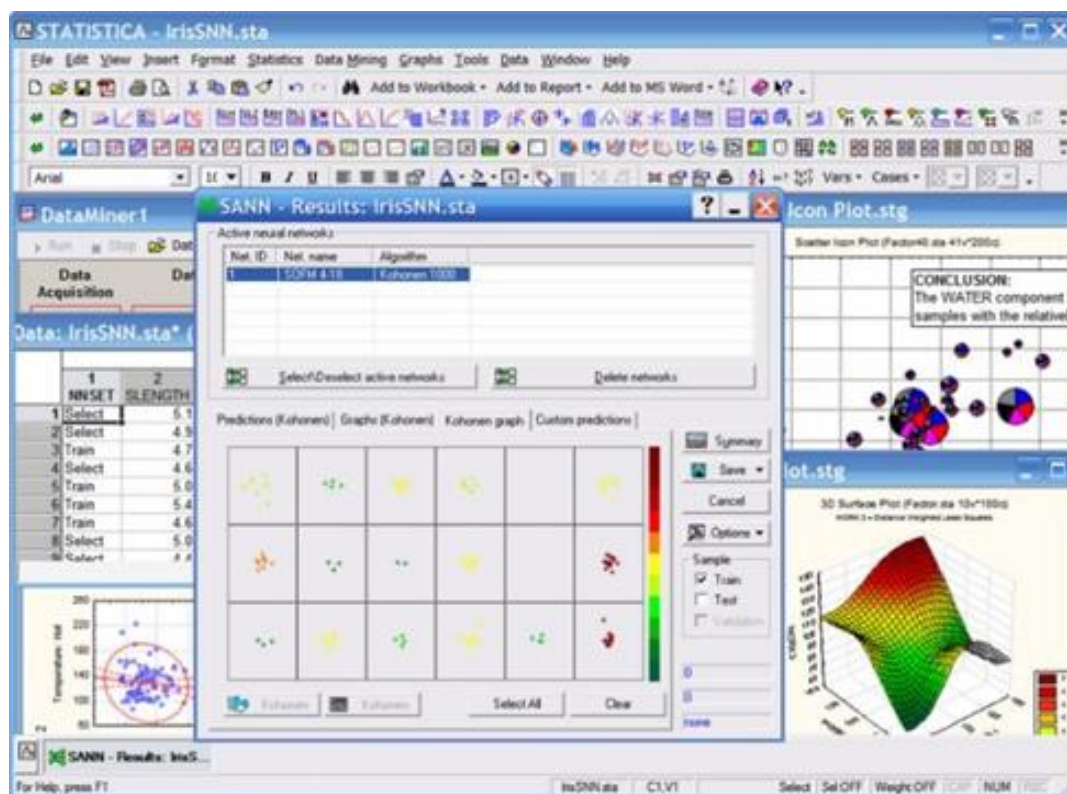


Рисунок 1.5 – Робоча область STATISTICA

Переваги використання автоматизованих нейромереж STATISTICA включають наступне:

- пре- та пост-процесування, яке включає вибір даних, кодування номінальних значень, шкалювання, нормалізацію, видалення пропущених даних з інтерпретацією для класифікації, регресії та завдань часових рядів;
- повний контроль над усіма параметрами, що впливають на якість мережі, такими як функції активації та помилок, складність мережі;
- повна інтеграція з потужними автоматичними інструментами STATISTICA;
- запис повноцінних макросів для будь-якого аналізу;
- використання STATISTICA Visual Basic для створення власних

нейромережових аналізів і додатків;

- виклик автоматичних нейромереж STATISTICA з будь-якої програми, що підтримує технологію CO;
- карти ознак, що самоорганізуються, багатошарові персептрони та радіальні базисні функції є одними з найбільш поширених мережових архітектур;
- підтримка різного роду статистичного аналізу та побудова прогнозуючих моделей, включаючи регресію, класифікацію, тимчасові ряди з безперервною та категоріальною залежною змінною, кластерний аналіз для зниження розмірності та візуалізації [19].

Автоматизований генератор коду нейромережових моделей STATISTICA може створювати системний програмний код для моделей нейромереж C, Java та PMML (Predictive Model Markup Language). Додатком до системи STATISTICA є Generator Code, який дозволяє користувачам створювати вихідний код моделей у форматі C або Java та інтегрувати його до незалежних зовнішніх додатків на основі результатів нейромережевого аналізу

Наведені системи дозволяють здійснювати прогнозування часових рядів різної природи. Однак у них є ряд недоліків. Пакети візуального моделювання мають обмежений набір елементів, а також вимагають знання програмування для можливості гнучкого налаштування моделей. Корпоративні системи BI та ERP є дорогими та потребують трудомісткого процесу розгортання та підтримки, а також не є гнучкими [20].

Статистичні пакети та табличні процесори не дозволяють використовувати системний підхід, а також потребують трудомістких процесів налаштування моделей та знань мов програмування. Тому існує потреба у розробці системи, яка має зручний та зрозумілий інтерфейс, проста у налаштуванні й автоматизує аналіз та прогнозування часових рядів обсягу подажу авозапчастин для інтернет-магазину: аналіз структури, вибір оптимальної моделі, оцінку її адекватності та точності і прогнозування нових значень.

1.5 Постановка задачі

Провівши аналіз сервісів онлайн-продаж було зроблено висновок про необхідність розробки системи прогнозування продажу товарів у сфері електронної комерції із вбудованими методами аналізу часових рядів та побудови прогнозних моделей. Створена система забезпечить доступ до відкритого програмного забезпечення, що дозволить використовувати та змінювати необхідні параметри для кожного завдання, враховуючи недоліки платформ. Прогнозування виконується за допомогою набору даних, який збирає дані про продажі за попередні роки. Тому потрібні дані протягом як мінімум трьох років, щоб зробити точне прогнозування результатів, зменшити помилки навчання та провести тестування системи.

Об'єктом дослідження є процес купівлі та продажу товарів через мережу Інтернету.

Предметом дослідження є програмні засоби та методи аналізу продажу і побудови прогнозних моделей у сфері електронної комерції.

Мета дослідження – підвищення ефективності системи прогнозування продажу товарів у сфері електронної комерції на основі методів та технологій прогнозування часових рядів.

Досягнення поставленої мети обумовлює необхідність вирішення наступних **завдань**:

- здійснити аналіз предметної сфери електронної комерції та дослідити теоретичні засади прогнозування продажу товарів;
- обґрунтувати вибір інструментальних засобів розробки системи прогнозування продажу;
- розробити та здійснити програмну реалізацію системи прогнозування продажу запчастин до автомобілів, дослідити якість прогнозної моделі та точність прогнозу.

Основною ціллю є розробка інтелектуальної системи, яка використовує

часові ряди для прогнозування продажу товарів на основі даних, зібраних за минулі роки. Її реалізація здійснюється у декілька етапів.

1. Збір та підготовка даних:

а) збір історичних даних про продажі: отримання часових рядів, які включають в себе інформацію про обсяги продажів протягом певного періоду;

б) обробка та очистка даних: виявлення та виправлення аномалій, відсутніх значень, перевірка на стабільність часового ряду.

2. Візуальний аналіз часового ряду:

а) побудова графіків часового ряду: визначення трендів, сезонностей та інших особливостей;

б) аналіз статистичних характеристик: оцінка рівня змінності, визначення автокореляційних функцій.

3. Вибір моделі для часового ряду та налаштування її параметрів.

4. Створення та валідація моделі: оцінка ефективності для визначення точності та стабільності прогнозів

5. Прогнозування та оцінка результатів: застосування моделі для прогнозування майбутніх значень.

Висновки до розділу 1

Установлено, що основним підходом до побудови прогнозу є моделювання часових рядів обсягів продажу за певний період часу, оскільки дозволяють враховувати тренди, сезонність та інші регулярні зміни у часі. Моделі часових рядів можуть ефективно адаптуватися до змін в споживчому попиті та інших факторах, що впливають на обсяги продаж.

Моделювання фінансових часових рядів є складною задачею, оскільки комерційна діяльність містить багато чинників, які можуть викликати ріст та падіння попиту на товари та їх вартість. На вибір методів аналізу та прогнозування

впливає стаціонарність та не стаціонарність часового ряду. До методів аналізу часових рядів, які базуються на припущенні про стаціонарність ряду, відносять авторегресійну модель (AR-модель) та її модифікації. Для підвищення ефективності планування онлайн продаж інтернет-магазину необхідна автоматизована система, яка дозволяє здійснювати аналіз динаміки обсягів продаж та будувати модель часового ряду з високою точністю прогнозу.

Здійснений аналіз дозволив виявити велику кількість методів та моделей, які використовують для побудови прогнозу продажу товарів. Це методи прогнозування із використанням глибинних нейронних мереж, авторегресійні AR-моделі, MA-модель ковзних середніх, EMA-модель експоненційних ковзних середніх, регресійні методи моделювання шляхом побудови лінійної та нелінійної аналітичної функції, адитивна AM-модель та узагальнена адитивна GAM-модель, мультиплікативна модель, ARIMA-модель інтегрованої авторегресії ковзного середнього та багато інших.

Важливе значення для побудови прогнозної моделі часового ряду має виявлення основних структурних складових часового ряду: тренду, сезонної, циклічної та випадкової компонент. Для прогнозування стаціонарного часового ряду будують авторегресійну AR-модель та модель ARMA. Для нестаціонарних часових рядів є два підходи до побудови прогностичної моделі: моделювання регулярних компонент у сукупності та розкладання часового ряду на компоненти та моделювання кожної компоненти окремо. Для часового ряду, у структурі якого виявлено тренд та відсутність сезонної компоненти, доцільною є модель ARIMA. Якщо у структурі часового ряду виявлено сезонну компоненту та відсутність тренду, можна передбачити побудову моделі SARIMA. У випадку наявності тренду та сезонності доцільно використовувати модель Хольта-Вінтерса.

Для перевірки адекватності прогнозної моделі часового ряду необхідно здійснити дослідження випадкової компоненти на стаціонарність і розрахунок метрик оцінки якості та точності побудованої моделі: MAE, MSE, RMSE, MAPE.

2 ІНСТРУМЕНТАЛЬНІ ЗАСОБИ РОЗРОБКИ СИСТЕМИ ПРОГНОЗУВАННЯ ПРОДАЖУ

2.1 Мова програмування Python

Python широко використовується для машинного навчання, аналізу даних і візуального представлення. Для науки про дані це одна з найбільш поширених мов. Він використовується для розробки різноманітних алгоритмів для аналітичних програм. Він використовується для обслуговування сховищ даних і хмарних сервісів, а також для аналізу довільних даних, отриманих через Інтернет. Наприклад, Google використовує Python, щоб індексувати сайти.

Python – це скриптова мова програмування, яка може вирішувати багато завдань для багатьох платформ, таких як IOS, Android і навіть серверних операційних систем. Веб-розробка, машинне навчання, аналіз даних, бізнес-додатки та системи ERP — усе це сфери, де використовується ця мова програмування.

Python має чудовий синтаксис, що робить його високорівневою мовою програмування, найчастіше використовується у веб-розробці через його розвинену екосистему швидкої розробки вебпроектів високої якості. Для розробки Python-застосунків існує ціла низка фреймворків: Django, Pyramid, Pylons, TurboGears, Flask і CherryPy.

Мова Python має такі переваги:

- розробники можуть легко читати та розуміти програми на Python, оскільки мова має базовий синтаксис, схожий на синтаксис англійської;
- Python допомагає розробникам бути більш продуктивними, оскільки вони можуть писати програми на Python, використовуючи менше рядків коду, ніж іншими мовами;
- Python має велику стандартну бібліотеку, що містить багаторазові коди практично для будь-якого завдання;

- розробники можуть легко поєднувати Python з іншими популярними мовами програмування: Java, C та C++;
- активна спільнота Python складається з мільйонів розробників з усього світу. У разі виникнення проблем співтовариство допоможе в їх вирішенні;
- крім того, в Інтернеті є безліч корисних ресурсів для вивчення Python. Наприклад, ви можете легко знайти відеоролики, навчальні посібники, документацію та посібники для розробників;
- Python можна переносити на різні операційні системи: Windows, MacOS, Linux і Unix.

Мова Python має кілька стандартних прикладів використання при розробці додатків, основними з яких є наступні.

1. *Веб-розробка на стороні сервера.* Веб-розробка на стороні сервера включає складні серверні функції, за допомогою яких веб-сайти відображають інформацію для користувача. Наприклад, веб-сайти повинні взаємодіяти з базами даних та іншими веб-сайтами, а також захищати дані під час їх надсилання по мережі.

Python корисний при написанні серверного коду, оскільки він пропонує безліч бібліотек, що складаються із попередньо написаного коду для складних серверних функцій. Також розробники використовують широкий спектр платформ Python, які надають всі необхідні інструменти для більш швидкого та простого створення інтернет-додатків. Наприклад, розробники можуть створити «скелет» інтернет-програми за лічені секунди, тому що їм не потрібно писати код з нуля. Потім можна протестувати за допомогою інструментів тестування платформи незалежно від зовнішніх інструментів тестування.

2. *Автоматизація за допомогою скриптів Python.* Мова скриптів – це мова програмування, яка автоматизує завдання, які зазвичай виконують люди. Програмісти широко використовують скрипти Python для автоматизації багатьох повсякденних завдань, серед яких:

- одночасне перейменування великої кількості файлів;
- перетворення файлу на інший тип файлу;

- видалення повторюваних слів у текстовому файлі;
- виконання базових математичних операцій;
- надсилання повідомлень електронної пошти;
- завантаження контенту;
- виконання базового аналізу журналів;
- пошук помилок у кількох файлах.

3. *Наука про дані та машинне навчання.* Наука про дані отримує цінну інформацію з даних, а машинне навчання (ML) дозволяє комп'ютерам автоматично вчитися на даних і робити точні прогнози. Фахівці з роботи з даними використовують Python для вирішення наступних завдань:

- виправлення та видалення неправильних даних (очищення даних);
- вилучення та вибір різних характеристик даних;
- розмітка даних додає даним значні імена;
- пошук статистичної інформації у даних;
- візуалізація даних за допомогою діаграм та графіків: лінійних діаграм, стовпчастих діаграм, гістограм та кругових діаграм.

Фахівці з роботи з даними використовують бібліотеки Python ML для створення моделей машинного навчання. Фахівці роботи з даними також використовують Python для глибокого навчання — передової техніки машинного навчання.

4. *Розробка програмного забезпечення.* Розробники програмного забезпечення часто використовують Python для різних завдань розробки та програмних додатків, серед яких:

- відстеження помилок у програмному коді;
- автоматичне складання програмного забезпечення;
- управління програмними проектами;
- розробка прототипів програмного забезпечення;
- розробка настільних додатків з використанням бібліотек графічного інтерфейсу користувача (ДПІ);

– розробка ігор: від простих текстових ігор до складних ігор.

5. *Автоматизація тестування програмного забезпечення.* Тестування програмного забезпечення – це процес перевірки відповідності фактичних результатів програмного забезпечення очікуваним результатам, який дозволяє переконатися, що програмне забезпечення не містить помилок. Розробники використовують середовища модульного тестування Python (Unittest, Robot та PyUnit) для тестування написаних функцій. Тестувальники програмного забезпечення використовують Python для написання прикладів тестів для різних сценаріїв. Наприклад, мова застосовується для тестування інтерфейсу інтернет-додатку, декількох програмних компонентів і нових функцій. Розробники можуть використовувати кілька інструментів автоматичного запуску тестових скриптів.

Python має багато бібліотек, які можна використовувати для обчислень і наукових досліджень: SciPy – бібліотека, яка містить кілька наукових інструментів, NumPy — це розширення, яке додає математичні функції для роботи з матрицями та багатовимірними масивами, Matplotlib — це бібліотека, яка дозволяє працювати з графікою в двох і трьох вимірах.

2.2 IDE для розробки на Python PyCharm

Для розробки системи прогнозування було обрано PyCharm – комплексну IDE для професійної розробки на Python, яка пропонує всі необхідні інструменти, а також чіткий, інтуїтивно зрозумілий інтерфейс з можливостями налаштування користувача.

PyCharm надає великий набір інструментів: вбудований відладчик та інструмент запуску тестів, профілювальник Python, повнофункціональний вбудований термінал, інструменти для роботи з базами даних. IDE інтегрована з популярними системами контролю версій, містить вбудований SSH-термінал, підтримує можливості віддаленої розробки та віддалені інтерпретатори, а також інтеграцію з Docker та Vagrant.

PyCharm підтримує всі основні реалізації мови Python: Python 2.x та 3.x, Jython, IronPython, PyPy та Cython, забезпечуючи: підсвічування синтаксису, перевірку коду на помилки, можливості форматування; автодоповнення коду, що враховує контекст; навігацію за кодом та перегляд структури; швидкий пошук входжень та інструменти рефакторингу; інспекції коду та багато іншого; юніт-тестування.

PyCharm дозволяє легко виконувати модульне тестування завдяки інтеграції з популярними тестовими фреймворками: doctests, nose та attest. У PyCharm можна запустити тестовий файл, тест для окремого класу чи методу або всі тести з папки одночасно. Результати тестів зручно переглядати в графічному інтерфейсі інструменту запуску тестів, який відображає статистику їх виконання і дозволяє швидко навігуватися за кодом, що тестується.

PyCharm інтегровано з інструментом аналізу покриття коду Coverage.py. Він відстежує, які частини програми були виконані, та аналізує вихідний код, щоб визначити, які фрагменти коду могли бути виконані, але цього не сталося. Результати відображаються у наочному форматі для подальшого аналізу та швидкої навігації за кодом.

PyCharm підтримує два популярні профільники: yappi і cProfile. Можна робити снєпшоти та збирати статистику роботи програми прямо в PyCharm – результати відображаються у вигляді кольорового графа викликів функцій. Можна переглянути зведений звіт, також передбачена навігація з графа до коду.

PyCharm підтримує два найбільш популярні інструменти для розробки з використанням BDD - Behave і Lettuce. Можна писати зрозумілий код тестів, що описує поведінку вашої програми. Підтримка включає: допомогу в інсталяції та конфігурації, Run Configuration та Intellisense.

Візуалізація паралельних потоків допомагає контролювати багатопотокові програми. Це дозволяє запустити програму в режимі Concurrency Diagram і перевіряти стан потоків всередині запущеного процесу на діаграмі в реальному часі.

Підтримка Git, SVN, Mercurial, Perforce та інших систем контролю версій допомагає керувати локальними змінами та здійснювати складні операції з гілками. Всі одноманітні завдання (додавання та видалення файлів) виконуються автоматично. Наступні функції доступні незалежно від того, яка система контролю версій використовується:

- інтегровані списки змін — групуйте зміни за різними списками для зручності організації;
- відкладені зміни — Збережіть окремо зміни, які потрібно внести пізніше;
- перегляд змін репозиторію - відстежуйте, які зміни внесли ваші колеги;
- перегляд вхідних змін — Переглядайте зміни, які ще не були внесені до локальної копії;
- повідомлення про неактуальні зміни — отримайте попередження про те, що файл, з яким ви працюєте, було змінено після останньої синхронізації.

Доступ до Oracle, SQL Server, PostgreSQL, MySQL та інших баз даних здійснюється безпосередньо з IDE. PyCharm допомагає редагувати SQL-код, переглядати дані, змінювати схеми та таблиці, виконувати запити та аналізувати схеми за допомогою UML-діаграм. Код SQL можна вставити у файли іншими мовами і працювати з ним у SQL-редакторі, який надає підсвічування синтаксису, автодоповнення, аналіз коду на льоту, форматування та зручну навігацію.

PyCharm здійснює аналіз коду бібліотеки SQLAlchemy, на основі якого пропонує варіанти автодоповнення. IDE також підтримує візуалізацію даних за допомогою діаграм SQLAlchemy.

2.3 Бібліотеки Python для аналізу часових рядів та побудови прогнозних моделей

Для створення моделі прогнозу продажів товарів із використанням інструментів машинного навчання розглянемо Python бібліотеки Pandas, NumPy, Matplotlib, Scikit-Learn, Statsmodels та проаналізуємо їх.

Бібліотека Pandas – це бібліотека, яка використовує мову програмування Python і призначена для обробки даних на високому рівні. Цей інструмент має відкритий код, швидкий, потужний, гнучкий та досить простий у використанні. Бібліотека, побудована на пакеті Numpy, має структуру даних DataFrame. Python є потужним інструментом для аналізу даних завдяки даному пакету. Будувannya зведених таблиць, виконання угруповань, зручний доступ до табличних даних і візуалізація графіків на основі отриманих наборів даних — усе це можна зробити за допомогою пакета Matplotlib [21].

Бібліотеки Pandas використовують такі структури: DataFrame і Series. DataFrame — це проіндексований масив значень, у якому кожен стовпець DataFrame має структуру серії. Можна зберігати та обробляти табличні дані у вигляді рядків спостережень і стовпців змінних за допомогою DataFrames.

Серія складається з об'єкта, який нагадує проіндексований одновимірний масив значень, але з додаванням індексів і міток, які пов'язані з кожним елементом списку. У Python ця функція перетворює його на асоціативний масив або словник, де ім'я елемента відповідає індексу, а значення запису відповідає значенням запису. При роботі з великими кількостями даних бібліотека Pandas є хорошою альтернативою Microsoft Excel.

NumPy (Numerical Python) – це бібліотека Python з відкритим вихідним кодом, яка використовується практично у всіх галузях науки та техніки. Це універсальний стандарт для роботи з числовими даними Python, що лежить в основі наукових екосистем Python і PyData. Користувачі NumPy включають усіх, від програмістів-початківців до досвідчених дослідників, які займаються сучасними науковими та промисловими дослідженнями та розробками.

Бібліотека NumPy містить багатовимірні масиви та матричні структури даних. NumPy можна використовувати для виконання множини математичних операцій з масивами. Він додає до Python потужні структури даних, що гарантують ефективні обчислення з масивами та матрицями, та надає величезну бібліотеку

високорівневих математичних функцій, які працюють із цими масивами та матрицями.

Ось кілька областей, де NumPy активно використовується: наукові та інженерні обчислення; машинне навчання та штучний інтелект; обробка даних; візуалізація даних; високопродуктивні обчислення.

NumPy надає вам величезний вибір швидких та ефективних способів створення масивів та маніпулювання числовими даними всередині них. Хоча список Python може містити різні типи даних в одному списку, всі елементи в масиві NumPy мають бути однорідними. Математичні операції, призначені до виконання над масивами, були б вкрай неефективні, якби масиви були однорідними.

Масиви NumPy швидше та компактніше, ніж списки Python. Масив споживає менше пам'яті (що дуже важливо для оптимізації швидкості обчислень або потужності, що використовується) і зручний у використанні. NumPy використовує набагато менше пам'яті для зберігання даних та надає механізм вказівки типів даних. Це дозволяє ще більше оптимізувати код.

Бібліотека Matplotlib — це бібліотека, яка працює з мовою програмування Python [22]. Бібліотека розроблена для візуалізації дво- та тривимірних графіків. Ілюстрації, створені Matplotlib, можна використовувати в інтерактивній графіці, наукових публікаціях, графічному інтерфейсі користувача, веб-додатках і в інших місцях, де потрібна побудова діаграм. Бібліотека Matplotlib базується на принципах ООП і є простим у використанні пакетом, який використовується разом із NumPy, SciPy та IPython, щоб надати команди, подібні до MATLAB. Наразі пакет працює з багатьма графічними бібліотеками, такими як wxWindows і PyGTK. У пакеті доступні такі графіки та діаграми: лінійні та контурні графіки, стовпчасті, кругові, спектральні та у вигляді стовбур-листа діаграми, гістограми, поля градієнтів. Користувач може вказати осі координат, додати надписи та пояснення до осей графіку, використавши полярні координати або логарифмічну шкалу.

Бібліотека Scikit-Learn є найпоширенішою бібліотекою для вирішення

різноманітних завдань, пов'язаних із класичними процесами машинного навчання. Бібліотека пропонує широкий спектр алгоритмів машинного навчання, які включають вчителів і без них. Навчання з вчителем передбачає наявність попереднього датасету з визначеними значеннями цільового показника. Незважаючи на те, що навчання без вчителя не вимагає наявності датасету, довільні дані пропонують корисні знання. Бібліотека має велику перевагу в тому, що вона працює на основі багатьох поширених математичних бібліотек і легко інтегрується з ними. Scikit-learn має докладну документацію та широку спільноту.

Промислові системи проводять різноманітні дослідження з використанням бібліотеки. Крім того, для новачків у машинному навчанні дана бібліотека ідеальна. Для виконання своїх завдань Scikit-Learn використовує такі поширені бібліотеки: NumPy для математичних операцій і операції над тензорами; SciPy для науково – технічних обчислень; Matplotlib для візуалізації даних; IPython у якості інтерактивної консолі для Python; SymPy для символічної математики; Pandas для обробки, маніпуляції та аналізу даних.

Для вирішення завдань машинного навчання з вчителем, таких як класифікація та регресія, Scikit-Learn спеціалізується на алгоритмах машинного навчання, які також використовуються для вирішення завдань машинного навчання без вчителя, таких як кластеризація (розбиття даних по класах), зниження розмірності (подання даних у просторі меншої розмірності з мінімальною втратою корисної інформації) і детектування аномалій. Бібліотека використовує такі основні методи [23]:

- лінійні – моделі, завдання яких побудувати роздільну або аппроксимируючу гіперплоскість;
- метричні – моделі, які обчислюють відстань між об'єктами вибірки по одній із доступних метрик;
- дерева рішень – методи, які використовуються для навчання моделей, що базуються на безлічі умов, оптимально обраних для вирішення завдання;
- ансамблеві – методи, засновані на основі дерев рішень, що комбінують

міць безлічі дерев і таким чином підвищують їх якість роботи, дозволяють проводити відбір ознак;

– нейронній мережі – комплексний нелінійний метод для задач регресії і класифікації даних.

Крім цього, Scikit-Learn містить функції для розрахунку значень метрик, вибору моделей, препроцесінга даних та інші.

Бібліотека statsmodels – це бібліотека Python, яка надає можливість оцінити безліч різних моделей, а також провести статистичні тести та дослідження даних. Існує ряд статистичних тестів та функцій, до яких бібліотека надає доступ, включаючи звичайні регресії найменших квадратів (OLS), узагальнені лінійні моделі, логіт-моделі, Аналіз головних компонентів (PCA) та Авторегресійні інтегровані ковзаючі середні (ARIMA) моделі. Результати цих моделей постійно перевіряються на відповідність іншим статистичним пакетам, щоб переконатися у їх точності.

У даному дослідженні бібліотека statsmodels використовується для оцінки якості та точності прогнозних моделей, побудованих на основі часових рядів.

2.4 Система керування пакетами PIP

PIP – система керування пакетами, яка використовується для встановлення пакетів із Python Package Index (PyPI). Pip є пакетний менеджер для мови програмування Python. Він дозволяє встановлювати, оновлювати та видаляти сторонні пакети, розширення та бібліотеки, необхідні для розробки Python-проектів. Без Pip було набагато складніше керувати залежностями проекту. З його допомогою можна легко встановити необхідні бібліотеки, вказати їх версії та забезпечити сумісність між різними проектами [24].

Pip — інструмент командного рядка для встановлення, оновлення та видалення пакетів Python. Він є одним з основних інструментів управління залежностями при розробці мовою Python. Pip дозволяє легко встановлювати

сторонні пакети та бібліотеки, потрібні для роботи з кодом.

Основна перевага Pip полягає в тому, що він автоматично встановлює всі необхідні залежності обраного пакету. Це дозволяє уникнути ручної установки кожної бібліотеки окремо та заощадити час розробника. Крім того, Pip дозволяє оновлювати пакети до останніх версій, що особливо корисно при роботі з бібліотеками, що активно розробляються.

Pip є стандартним інструментом встановлення пакетів у Python і поставляється разом із встановленням мови. Він заснований на системі пакетування PEP (Python Enhancement Proposals), яка визначає стандарти розробки пакетів для мови Python. Якщо розробник на платформі Python зіштовхується із зовнішніми залежностями або просто хоче максимально спростити встановлення пакетів – Pip є надійним помічником у цьому.

Важливою особливістю pip є те, що він може встановлювати пакети з офіційного репозиторію Python Python Package Index (PyPI). Це означає, що більшість популярних бібліотек, таких як NumPy, Django, requests та багато інших, можуть бути встановлені за допомогою pip однією командою. Крім того, pip також підтримує встановлення пакетів з інших джерел, таких як Git-репозиторії або локальні файли, що надає ще більшу гнучкість при установці сторонніх пакетів.

За допомогою pip можна керувати залежностями проекту автоматично. Проект може містити файл залежностей (наприклад, requirements.txt), у якому перелічені всі необхідні пакети із зазначенням їх версій. Під час встановлення проекту pip автоматично встановить усі вказані пакети відповідних версій.

Основні можливості та функції, що надаються Pip, включають:

- установка пакетів: Pip дозволяє легко встановити пакети, написані на Python, просто запустивши команду `pip install`;
- оновлення пакетів: Pip також дозволяє оновити вже встановлені пакети до найновішої доступної версії;
- видалення пакетів: Pip дозволяє видалити встановлені пакети, які не потрібні, за допомогою команди `pip uninstall`;

- список встановлених пакетів: Pip надає можливість одержати список всіх встановлених пакетів з їх версіями, використовуючи команду `pip list`.
- Крім того, Pip також підтримує:
- встановлення пакетів із файлу залежностей: Pip дозволяє встановити всі пакети, перелічені у файлі залежностей, за допомогою команди `pip install -r requirements.txt`;
- встановлення пакетів з вихідного коду: Pip підтримує інсталяцію пакетів безпосередньо з вихідного коду, що знаходиться на віддалених репозиторіях, або з локальної директорії.

Завдяки своїй простоті використання та широкому функціоналу, Pip є невід'ємним інструментом для розробників Python. Він значно спрощує процес встановлення та управління пакетами, роблячи його більш ефективним та продуктивним.

2.5 Засоби веброзробки системи прогнозування

Одним із найпопулярніших середовищ веброзробки є IDE PhpStorm. Це потужне інтегроване середовище розробки, яке пропонує широкий спектр можливостей та функцій, що допомагають розробникам створювати високоякісні вебсайти та програми. Основні можливості PhpStorm:

- автоформатування коду;
- підтримка Docker, Composer, вбудованим клієнтом REST;
- крос-платформність;
- рефакторинг;
- автоматичне виправлення помилок у кодифікації;
- розумне автогенерування коду.

IDE PhpStorm – це зручне середовище програмування, яке вирізняється своєю функціональністю. Вміє працювати з фреймворками та спеціалізованими плагінами для провідних фреймворків PHP. PhpStorm інтегрований з найпопулярнішими

системами контролю версій, включаючи Git, Subversion, Mercurial, Perforce, CVS та TFS. Всі одноманітні завдання (наприклад, додавання та видалення файлів) виконуються автоматично.

Однією з головних переваг використання PhpStorm є його зручний інтерфейс. Він розроблений так, щоб бути інтуїтивно зрозумілим та простим у використанні, тому навіть новачки можуть швидко навчитися користуватися програмою без особливих зусиль. Крім того, він має вбудовані можливості налагодження, що полегшує розробникам швидкий пошук та виправлення помилок.

Ще однією перевагою є широкий набір функцій, які включають завершення коду, підсвічування синтаксису, інтеграцію із системами контролю версій, такими як GIT або SVN, підтримку багатьох мов, включаючи HTML5, CSS та ін.

Для розробки вебінтерфейсу вебзастосунку системи було використано мову гіпертекстової розмітки тексту HTML та мову опису стилів CSS. HTML по праву вважається базовою мовою веб-розробки, оскільки його використовують для структурування та відображення сторінок сайтів і всього розміщеного на них контенту. Таким чином, HTML-код виступає як інструкція для браузера, допомагаючи йому зрозуміти, як правильно показати вміст веб-сторінки.

До переваг HTML відносять [25]:

- відходить майже всім браузерів, а код можна написати в будь-якому текстовому редакторі;
- HTML можна використовувати безкоштовно;
- щоб освоїти ази мови розмітки гіпертексту, достатньо кількох годин. У нього простий та зрозумілий синтаксис;
- підходить для професійних розробників сайтів та для любителів;
- статичні сайти на HTML, яким не потрібно звертатися до баз даних, швидко завантажуються, тому що менше важать і в них відсутній сміттєвий код. Вони дешевші у розробці, і для таких сайтів підійде найдешевший хостинг;
- HTML інтегрується з іншими програмними мовами;
- HTML-елементи містять у собі семантику і працюють з урахуванням

цифрової доступності — ними можуть користуватися незрячі люди;

– HTML здатний сам закривати теги та пропускати синтаксичні помилки. На відміну від помилки в JS, сторінка не впаде, а все одно завантажиться, хай і з непередбачуваними спотвореннями.

Якщо потрібно, щоб сайт виглядав красиво та реагував на дії користувача, доведеться застосувати ще CSS та JS. За допомогою HTML розробник збирає "скелет" сторінок вебзастосунку. Після чого працює з CSS-кодом для стилізації сторінок. У фіналі застосовують JavaScript для створення інтерактивних вебсторінок.

CSS (Cascade Style Sheets) - це формальна мова опису зовнішнього вигляду сторінки; каскадні таблиці стилів [26]. Технології CSS дозволили розділити зміст документа та його візуальне оформлення. Завдяки їм стало легко застосувати єдиний стиль для кожного документа, просто підключаючи HTML файл з CSS. Достатньо змінити значення будь-якої стилізаційної властивості тільки в одному файлі, і воно застосовувалося на всіх сторінках, де цей файл було підключено.

JavaScript належить до мультипарадигменним мов завдяки одночасній підтримці кількох стилів: функціонального, об'єктно-орієнтованого та імперативного, що в практичному застосуванні означає його вбудовуваність, що забезпечує інтерактивність веб-браузерів. Завдяки цьому інструменту, можна не перезавантажуючи сторінки здійснювати запуск анімації, підсвічувати елементи керування та отримувати доступ до керування іншими функціями та ефектами. Наприклад, поява меню, що випадає, при наведенні курсору на відповідний значок у вигляді трьох точок – робота відповідного скрипта, головним у роботі з якими якраз і є JavaScript.

Основою JavaScript є скрипти – документи, в яких описані дії, що дозволяють обробляти та виконувати команди. Кросплатформні мови програмування, до яких належить JS, працюють із будь-якими браузерами.

PHP — широко використовувана серверна сценарна мова загального призначення з відкритим кодом, яка ідеально підходить для веб-розробки та може

бути вбудована в HTML. Данна мова програмування є потужним інструментом для створення динамічних, інтерактивних веб-сторінок. Це безкоштовна та ефективна альтернатива конкурентам, таким як ASP Microsoft. Головною перевагою PHP є те, що, хоча він пропонує багато вдосконалених функцій для професійних програмістів, він надзвичайно простий для новачків. Програміст може розпочати написання простих серверних сценаріїв за допомогою великої кількості вбудованих функцій мови програмування.

PHP може працювати з різними службами за допомогою протоколів, таких як LDAP, IMAP, SNMP, NNTP, POP3, HTTP, COM (у Windows) та багатьох інших.

Користувач може відкрити необроблені мережеві розетки та використовувати будь-який інший протокол для взаємодії. Майже всі мови веб-програмування підтримують складний обмін даними WDDX через PHP. Що стосується взаємодії, PHP підтримує інстанціювання Java-об'єктів і дозволяє їх використовувати як PHP-об'єкти. Підтримка різноманітних баз даних є однією з найважливіших характеристик PHP. Написати веб-сторінку з підтримкою бази даних дуже просто, використовуючи одне з розширень бази даних (наприклад, mysql) або рівень абстракції (наприклад, PDO), або використовуючи розширення ODBC, щоб підключитися до будь-якої бази даних, яка підтримує стандарт Open Data Connection. ODBC або сокети, як CouchDB, можуть бути використані іншими базами даних. З вищезазначеного можна впевнено сказати, що перерахувати всі функції та переваги, які може запропонувати мова програмування PHP, досить просто.

Висновки до розділу 2

У даному розділі було обгрунтовано вибір для розробки мови програмування Python, середовища розробки PyCharm та менеджера управління Python-пакетами PIP для аналізу часових рядів та створення, навчання та оцінки якості прогнозних моделей. Серед Python бібліотек з цією метою було вирішено застосовувати Pandas,

NumPy, Matplotlib, Scikit-Learn, Statsmodels. Для управління даними та виведення результатів було використано бібліотеку Matplotlib. Для аналізу часових рядів, побудови прогнозної моделі та оцінки її якості використано бібліотеки Pandas, Scikit-learn і Statsmodels.

Для веброзробки було вирішено використовувати середовище IDE PhpStorm, мову HTML, CSS для стилізації сторінок та JavaScript для створення інтерактивних вебсторінок. PHP було застосовано для вбудови серверного коду в HTML та доступу до серверу Apache і роботи з базою даних СУБД MySQL із інформацією про онлайн-транзакції інтернет-магазину продажу авозапчастин. Для асинхронного програмування при запиті до серверу вирішено використовувати JavaScript.

3 РОЗРОБКА ТА ПРОГРАМНА РЕАЛІЗАЦІЯ СИСТЕМИ ПРОГНОЗУВАННЯ ПРОДАЖУ ЗАПЧАСТИН ДО АВТОМОБІЛІВ

3.1 Основні етапи побудови прогнозної моделі

На етапі проектування розроблено структурну схему побудови прогнозних моделей та отримання прогнозу (рис. 3.1) [27]. Основна частина функціональності системи полягає в аналізі часових рядів, побудові за його результатами прогнозної моделі та оцінці її якості.

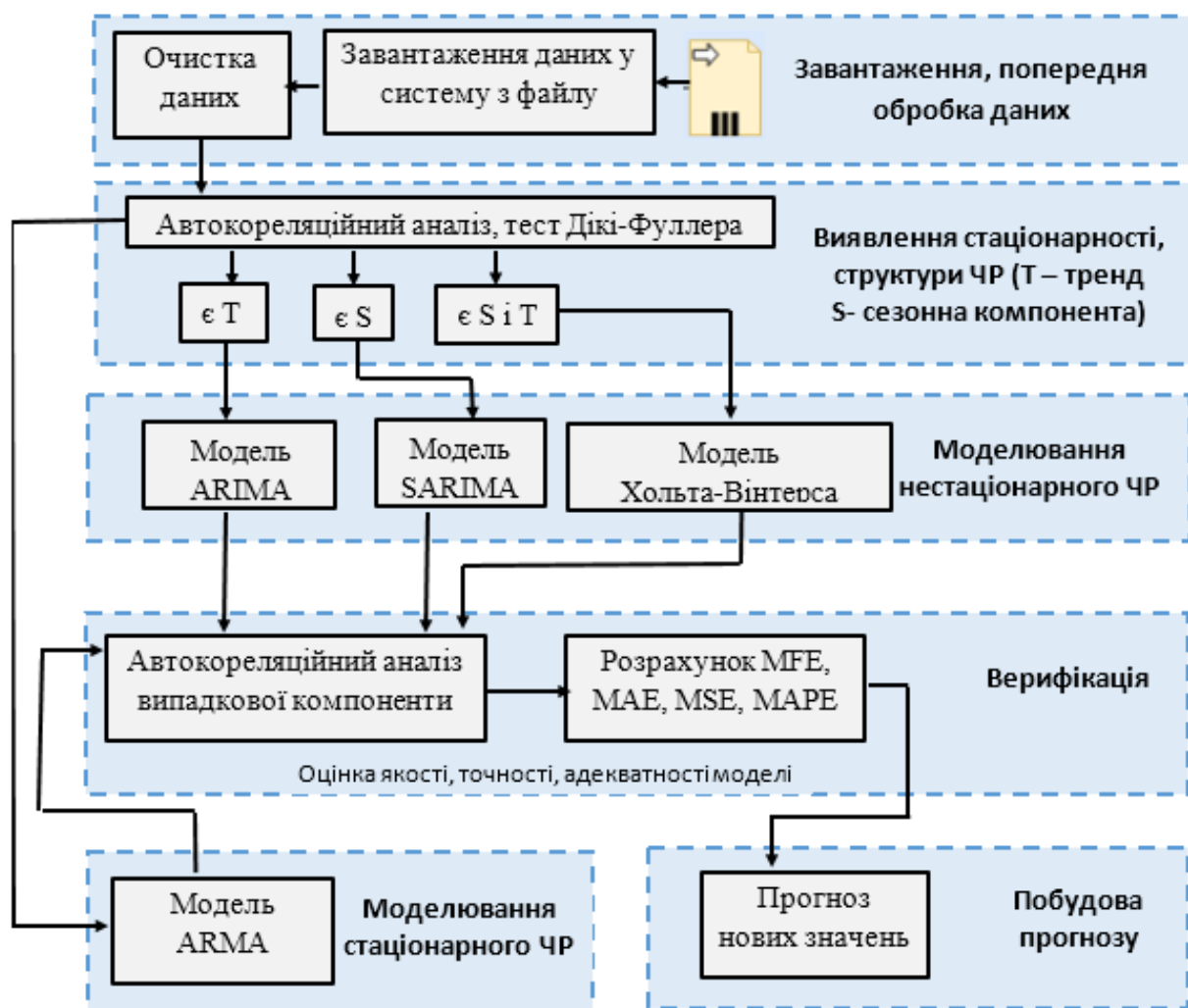


Рисунок 3.1 – Структурна схема побудови прогнозу

На початку роботи із системою завантажують набір даних зі значеннями часового ряду курсу стосовно продажу товарів. У роботі було використані дані стосовно продажу деталей автомобілів комерційної фірми, яка займається їх продажем. На етапі попередньої обробки даних здійснюється перевірка наявності пропусків, дублікатів та аномальних значень і у разі потреби – очистка даних.

Далі відбувається перевірка часового ряду на стаціонарність і виявлення його структури шляхом проведення автокореляційного аналізу та з використанням тесту Дікі-Фуллера. Для стаціонарного часового ряду у системі передбачено побудову ARMA-моделі. Для часового ряду, у структурі якого виявлено тренд та відсутність сезонної компоненти, передбачено побудову моделі ARIMA. Якщо у структурі часового ряду виявлено сезонну компоненту та відсутність тренду, передбачено побудову моделі SARIMA. У випадку наявності тренду та сезонності передбачено використання моделі Хольта-Вінтерса – потрійного експоненційного сгладжування.

Для перевірки адекватності прогнозної моделі часового ряду здійснюється дослідження випадкової компоненти на стаціонарність і розрахунок метрик оцінки якості та точності побудованої моделі. Під час навчання моделей для оцінки параметрів якості та точності для підбору параметрів було застосовано кросс-валідацію, яка враховує хронологічну структуру часових рядів.

Після чого побудована прогнозна модель готова до її застосування з метою здійснення короткострокового прогнозу продажу, який можна враховувати при плануванні закупок та продажу у майбутньому.

3.2 Попередня підготовка та обробка даних

Програмна реалізація системи прогнозування здійснювалася для онлайн-магазину продажу автозапчастин, онлайн-платформа якого функціонує на базі системи CMS OpenCart.

Для зберігання даних ця система використовує СКБД MySQL, у якій відсутня єдина таблиця з записами усіх потрібних параметрів транзакцій: дата, кількість, ціна, необхідних для формування часових рядів продажу товарів. Тому було завантажено дві таблиці в форматі .csv, які потребували відбору необхідних атрибутів (рис. 3.2).

order_product_id	order_id	product_id	name	model	quantity	price	total	tax	reward
7	3	61	"РЎС,РµРµР»Pѕ P·P°PrPSPµPNё P»PµPIPsPNё PrPIPµCЪPё"	"1052717-00"	1	200.0000	200.0000	0.0000	0
8	3	60	"РЎС,РµРµР»Pѕ PµPµCЪPµPrPSPµPNё P»PµPIPsPNё PrPIPµCЪPё	1028822-00"	1	170.0000	170.0000	0.0000	0
6	3	62	"PµPѕCЪC,PѕC¶PёP° C¶C,РµРµP°P° P·P°PrPSPµPNё P»PµPIPsPNё PrPIPµCЪPё	1055275-00-F "	1	100.0000	100.0000	0.0000	0
5	3	63	"PµPѕCЪC,PѕC¶PёP° PёCЪC¶P»P° P·P°PrPSPµPIPs P»PµPIPsPIPs	1034906-80-F"	1	125.0000	125.0000	0.0000	0
9	3	58	"PµPµC,Р»C¶ PSPёP¶PSC¶C¶ PµCЪP°PIPsPNё PµPµCЪPµPrPSPµPNё PrPIPµCЪPё"	6009569-00-E"	1	50.0000	50.0000	0.0000	0
10	3	57	"PµC¶PёP»PёC,РµP»CЪ P±P°PµPµCЪP° PµPµCЪPµPrPSPµPIPs"	6005884-00-K"	1	350.0000	350.0000	0.0000	0
353	277	2895	"P·P»PµPёC,CЪPѕPµCЪPёPµPIPsPr P»CЪPµPµCЪC¶C¶	1104854-01"	2	20.0000	40.0000	0.0000	0
354	278	3140	"PµPѕP»PёP° P±P°PµPµP¶P¶PSPѕPIPs PѕC,PrPµP»PµPSPёC¶ PµPµCЪPµPrPSC¶C¶	1492951-00-PЎ"	1	75.0000	75.0000	0.0000	0
12	5	527	"PµP°P±PµP»CЪ P¶PµPµP° 1032274-00-P"	1032274-00-P"	1	70.0000	70.0000	0.0000	0
374	291	323	"25 P P°PµP° PµCЪP°PIPsPNё P·P°PrPSPµPNё PrPIPµCЪPё	1028783-00"	1	90.0000	90.0000	0.0000	0
17	7	340	"PµPIPs»PѕC,PSPёC,РµP»CЪ PrPIPµCЪPё P·P°PrPSPµPNё P»PµPIPsPNё	6009585-00-I "	2	100.0000	200.0000	0.0000	0
372	289	3073	"4 P¶PSC,РµPSP° bluetooth 1508179-00-E"	1508179-00-E"	1	27.0000	27.0000	0.0000	0
373	290	2892	"3 PµPѕP»PёP° P±P°PµPµP¶P¶PSPѕPIPs PѕC,PrPµP»PµPSPёC¶ PSPёP¶PSC¶C¶	1492949-00-B"	1	75.0000	75.0000	0.0000	0

Рисунок 3.2 – Вихідний набір даних у форматі .csv

На етапі попередньої обробки отриманих даних передбачено відбір даних стосовно обраних типів деталей з характеристиками, необхідними для отримання хронологічної послідовності їх продажу (рис. 3.3, рис. 3.4). Для відібраних даних передбачено очищення, яке включає виявлення та обробку пропусків і дублікатів.

```
date,quantity
2019-04,6
2019-09,3
2019-10,2
2019-11,12
2019-12,2
2020-01,4
2020-02,5
2020-03,4
2020-04,1
2020-05,3
```

Рисунок 3.3 – Дані часового ряду

```

columns_to_drop = ["model", "order_product_id", "tax", "reward"]
columns_to_drop1 = ["invoice_no", "invoice_prefix", "store_id", "store_name", "store_url",
                    "customer_group_id", "firstname", "lastname", "email", "telephone", "f
                    "payment_firstname", "payment_lastname", "payment_company", "payment_a
                    "payment_address_2", "payment_city", "payment_postcode", "payment_coun
                    "payment_country_id", "payment_zone", "payment_zone_id", "payment_addr
                    "payment_custom_field", "payment_method", "payment_code", "shipping_fi
                    "shipping_lastname", "shipping_company", "shipping_address_1", "shippi
                    "shipping_city", "shipping_postcode", "shipping_country", "shipping_co
                    "shipping_zone", "shipping_zone_id", "shipping_address_format", "shipp
                    "shipping_method", "shipping_code", "comment", "order_status_id", "aff
                    "marketing_id", "tracking", "language_id", "currency_id", "currency_co
                    "ip", "forwarded_ip", "user_agent", "accept_language", "date_modified"]

df = df.drop(columns=columns_to_drop)
df1 = df1.drop(columns=columns_to_drop1)
merged_df = pd.merge(df, df1, on='order_id')
merged_df['date_added'] = pd.to_datetime(merged_df['date_added'])
merged_df['date_added'] = merged_df['date_added'].dt.strftime('%Y-%m')
ordered_by_date = merged_df.sort_values(by='date_added', ascending=True)
ordered_by_date = ordered_by_date.drop(columns=["total_x", "total_y"])
ordered_by_date.rename(columns={'date_added': 'date'}, inplace=True)
ordered_by_date = ordered_by_date.groupby('date')['quantity'].sum().reset_index()
ordered_by_date.to_csv('oc_order_clean_1.csv', index=False)
return ordered_by_date

```

Рисунок 3.4 – Очищення даних та відбір необхідних характеристик

Структура набору даних після очищення і попередньої обробки містить атрибути: date (тип даних datetime) та quantity (тип даних int). Стовець «date» має тип даних «datetime» для обрізання дати до потрібного формату, коли дата перетворена на тип «datetime», вона стає універсальною і може бути використана у різних контекстах, як усередині коду Python, так і в інших застосунках та бібліотеках. Стовець «quantity» має цілочисловий тип даних, тому що дані набору представлені у кількості проданих запчастин.

3.3 Програмна реалізація побудови моделей часового ряду

Розробка система здійснювалася у середовищі PyCharm на мові програмування Python. Для здійснення аналізу часових рядів та візуального відображення було завантажено бібліотеки Scikit-Learn, statsmodels, Pandas, NumPy

и Matplotlib (рис. 3.65). Для розробки системи було використано PIP – система керування пакунками, що використовується для встановлення та управління програмними пакетами, які написані на Python.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.tsa.holtwinters import ExponentialSmoothing
from statsmodels.tsa.stattools import acf
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.tsa.stattools import adfuller
from statsmodels.tsa.seasonal import seasonal_decompose
```

Рисунок 3.5 – Завантаження бібліотек

Перед початком розробки моделі отриманні дані потрібно проаналізувати. Спочатку ряд перевіряється на стаціонарність. Для перевірки ряду на стаціонарність використано тест Дікі-Фуллера. У статистиці тест Дікі-Фуллера використовують для перевірки нульової гіпотези про наявність одиничного кореня в авторегресійній моделі. Альтернативна гіпотеза варіюється в залежності від версії тесту, але зазвичай означає наявність стаціонарності або тренд-стаціонарності. Фрагмент коду з тестом Дікі-Фуллера наведено на рисунку 3.6.

```
result_adf = adfuller(df['quantity'])
print(f'ADF Statistic: {result_adf[0]}')
print(f'p-value: {result_adf[1]}')
print('Critical Values:')
for key, value in result_adf[4].items():
    print(f'    {key}: {value}')
```

Рисунок 3.6 – Код тесту Дікі-Фуллера

Для здійснення автокореляційного аналізу та візуального відображення графіку корелограми було реалізовано функцію, зображену на рисунку 3.7.

```
def autocorelation(df):  
    plt.figure(figsize=(12, 6))  
    plot_acf(df['quantity'], lags=15, alpha=0.05)  
    plt.title('Корелограма часового ряду')  
    plt.xlabel('Лаги')  
    plt.ylabel('Кореляція')  
    plt.grid(True)  
    plt.show()
```

Рисунок 3.7 – Побудова корелограми

У разі нестационарності часового ряду робиться його декомпозиція та візуальне відображення складових (рис. 3.8).

```
result = seasonal_decompose(df['quantity'], model='additive', period=9)  
plt.figure(figsize=(12, 8))  
plt.subplot(*args: 3, 1, 1)  
plt.plot(*args: df['quantity'], label='Визідні дані')  
plt.legend()  
plt.subplot(*args: 3, 1, 2)  
plt.plot(*args: result.trend, label='Тренд')  
plt.legend()  
plt.subplot(*args: 3, 1, 3)  
plt.plot(*args: result.resid, label='Залишок')  
plt.legend()  
plt.tight_layout()  
plt.show()
```

Рисунок 3.8 – Код декомпозиції часового ряду

Корелограма дозволяє візуально проаналізувати структуру часового ряду. Для обґрунтованого підтвердження наявності тренду та сезонності було написано код, представлений на рисунках 3.9 та 3.10.

```
def has_trend(dataset, threshold=0.05):
    X = np.arange(len(dataset)).reshape(-1, 1)
    y = dataset['quantity']
    model = LinearRegression()
    model.fit(X, y)
    p_value = model.score(X, y)
    if p_value > 0.05:
        return True
    return False
```

Рисунок 3.9 – Виявлення тренду у структурі часового ряду

```
def has_seasonal(dataset, max_lag=20):
    acf_values = acf(dataset['quantity'], nlags=max_lag, fft=False)
    for lag in range(1, max_lag + 1):
        if abs(acf_values[lag]) > acf_values[1]:
            return True
    return False
```

Рисунок 3.10 – Виявлення сезонної компоненти у структурі часового ряду

Після аналізу структури часового ряду у системі передбачено побудову прогнозних моделей відповідно до виявленої структури та їх навчання:

- ARMA: для стаціонарного часового ряду (рис. 3.11);
- ARIMA: для нестаціонарного часового ряду, який містить тренд (рис. 3.12);
- SARIMA: для нестаціонарного часового ряду, який містить сезонну компоненту (рис. 3.13);
- модель Хольта-Вінтерса: для нестаціонарного часового ряду, який містить сезонну компоненту і тренд (рис. 3.14).

```
p = 1 # Порядок авторегресії (AR)
q = 1 # Порядок ковзного середнього (MA)
model = ARIMA(train['quantity'], order=(p, 0, q)) # d=0, ряд є стаціонарним
result_arma = model.fit()
forecast = result_arma.forecast(len(test))
```

Рисунок 3.11 – Побудова та навчання моделі ARMA


```
model = ARIMA(train['quantity'], order=(1, 2, 5))
results = model.fit()
forecast = results.get_forecast(steps=len(test))
forecast_mean = forecast.predicted_mean
```

Рисунок 3.12 – Побудова та навчання моделі ARIMA

```
order = (1, 2, 5) # (p, d, q)
# Порядок авторегресії (AR) p
# Порядок ковзного середнього (MA) q
seasonal_order = (1, 1, 1, k) # (P, D, Q, S), S - період сезонності

model = SARIMAX(train['quantity'], order=order, seasonal_order=seasonal_order)
results = model.fit()
forecast = results.get_forecast(steps=len(test))
forecast_mean = forecast.predicted_mean
```

Рисунок 3.13 – Побудова та навчання моделі SARIMA

```
seasonal_periods = k
trend='additive'
seasonal='additive'
model = ExponentialSmoothing(train['quantity'], trend=trend, seasonal=seasonal,
                             seasonal_periods=seasonal_periods)
model_fit = model.fit()
forecast = model_fit.forecast(len(test))
```

Рисунок 3.14 – Побудова та навчання моделі Хольта-Вінтерса

Перед побудовою обраної моделі здійснюється підбір її параметрів. Модель ARIMA(p , q , d) має три параметри: d – порядок інтегрування (різниць вихідного часового ряду), рівний 0 для сатціонарного часового ряду; p – порядок авторегресії AR; q – порядок ковзного середнього.

У моделі SARIMA до параметрів p , q і d моделі ARIMA додаються параметри сезонної компоненти: P – порядок сезонної складової SAR; D – порядок

інтегрування сезонної складової; Q – порядок сезонної складової SMA; S – період сезонної компоненти.

В бібліотеці `statsmodels` немає можливості автоматизованого підбору параметрів. Підбір параметрів здійснюється шляхом перебору різних наборів параметрів із урахуванням зробленого аналізу корелограми часового ряду і зупинці та тих параметрах, які дають кращу точність прогнозу.

При прогнозуванні економічних показників прогноз є короткочасним. Прогнозування часових рядів на більшу кількість тимчасових кроків уперед може бути складним завданням і має обмеження [28]. Ось кілька причин, чому робити прогнози на більш ніж 3 тимчасові кроки вперед може бути важко:

- збільшення невизначеності;
- погіршення якості прогнозу;
- збільшення обчислювальної складності;
- збільшення помилки прогнозу;
- нестійкість моделі.

Загалом прогнозування на більш ніж 3 тимчасових кроки вперед може бути корисним для загального орієнтування, але для більш точних і надійних прогнозів часто рекомендується використовувати коротші горизонти прогнозування. Тому було вирішено прогнозувати на три кроки вперед для отримання прогнозу продаж деталей [29].

Для визначення адекватності моделі здійснюється побудова корелограми залишкової компоненти – якщо вона є стаціонарним рядком – модель адекватна, якщо ні – необхідно здійснити побудову моделі з іншими параметрами або іншої моделі.

Для оцінки показників точності моделей було написано код, представлений на рисунку 3.15. у випадку задовільної та високої точності прогнозу модель використовується для прогнозування продажу товарів на три періоди уперед. Якщо точність є незадовільною, здійснюється побудова моделі з іншими параметрами або підбір іншої моделі.


```
mse = mean_squared_error(actual_values, forecast_values)
mae = mean_absolute_error(actual_values, forecast_values)
mfe = np.mean(actual_values - forecast_values)
rmse = np.sqrt(mse)
1 usage
def mean_absolute_percentage_error(y_true, y_pred):
    return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
mape = mean_absolute_percentage_error(actual_values, forecast_values)

print(f'MSE: {mse}')
print(f'MAE: {mae}')
print(f'MFE: {mfe}')
print(f'RMSE: {rmse}')
print(f'MAPE: {mape}')
```

Рисунок 3.15 – Код для визначення показників точності моделі

3.4 Аналіз та прогнозування продажу

Роботу розробленої системи було виокрістано для аналізу та прогнозу продажу онлайн магазину автозапчастин. З цією метою меню адміністративної панелі платформи CMS OpenCart, на якій функціонує магазин, було розширено – додано новий пункт Прогнозування (рис. 3.16).

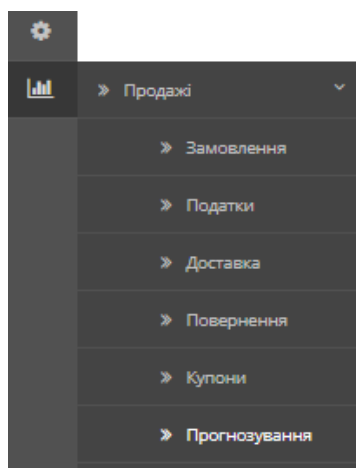


Рисунок 3.16 – Пункт меню Прогнозування в Адміністративній панелі

При виборі у адміністративній панелі пункту меню Прогнозування буде відкрита сторінка Прогнозування (рис. 3.17). При натисненні кнопки «Завантаження даних» буде відправлено асинхронний запит за допомогою AJAX у метод контроллера, який в свою чергу звертається до бази даних із запитом на завантаження даних двох необхідних для роботи таблиць.

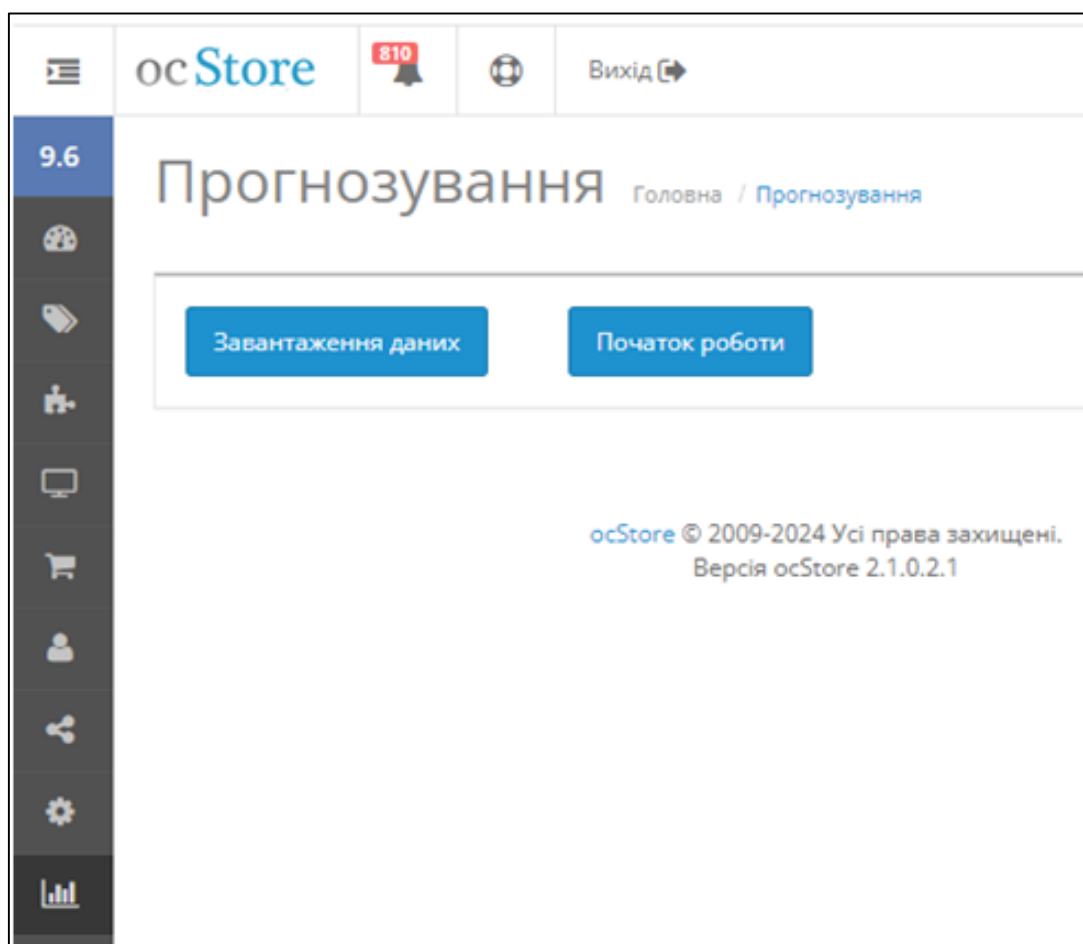


Рисунок 3.17 – Сторінка Прогнозування

Після успішного запиту буде виведено відповідне повідомлення, якщо виникла помилка теж буде відображено повідомлення про це (рис. 3.18). Якщо все пройшло успішно для початку аналізу відібраних даних необхідно натиснути кнопку «Початок роботи».

Прогнозування Головна / Прогнозу

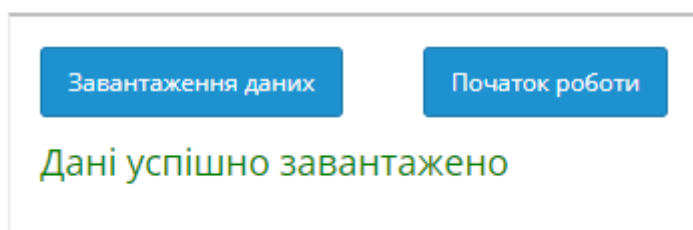


Рисунок 3.18 – Результат успішного запиту до БД

Для підготовки до аналізу передбачено відбір продажу комплектуючих певного типу. На початку буде виведено графік продажу, візуальний аналіз якого дає можливість стверджувати, що ряд може мати тренд і можливо сезонність (рис. 3.19).

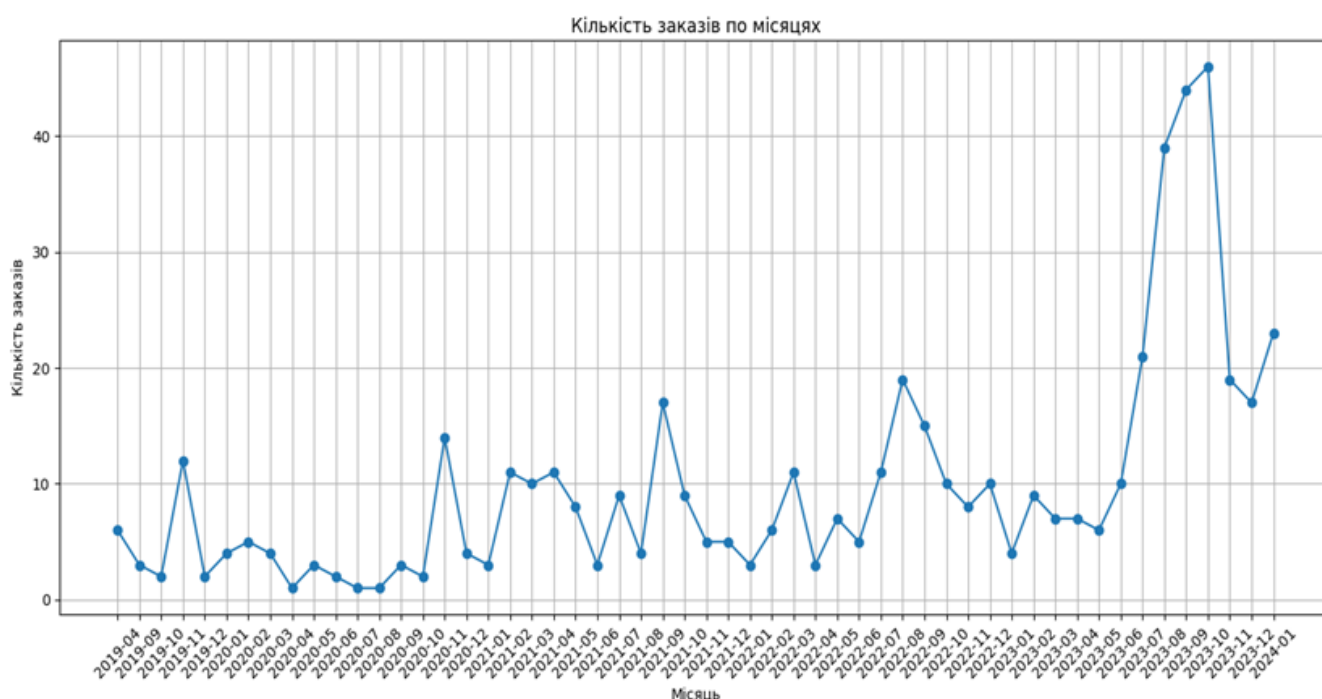


Рисунок 3.19 – Продаж деталей з 04.2019 року до 01.2024 року

Перевірки стаціонарності ряду здійснювалася з допомогою тесту Дікі-Фуллера (рис. 3.20). У статистиці, тест Дікі-Фуллера використовують для

перевірки нульової гіпотези про наявність одиничного кореня в авторегресійній моделі. Альтернативна гіпотеза варіюється в залежності від версії тесту, але зазвичай означає наявність стаціонарності або тренд-стаціонарності.

Як бачимо, ймовірність (p-value) даних значно вище 0,05. Ми не можемо відкинути нульову гіпотезу. Процес нестаніонарний.

```
ADF Statistic: 1.268698018515006
p-value: 0.9964277919555198
Critical Values:
  1%: -3.596635636000432
  5%: -2.933297331821618
 10%: -2.6049909750566895
```

Рисунок 3.20 – Результати перевірки на часового ряду на стаціонарність

Оскільки ряд нестаніонарний, виявляємо його структуру – наявність тренду та сезонної компоненти. Для цього будується автокореляційна функція і робиться її аналіз (рис. 3.21).

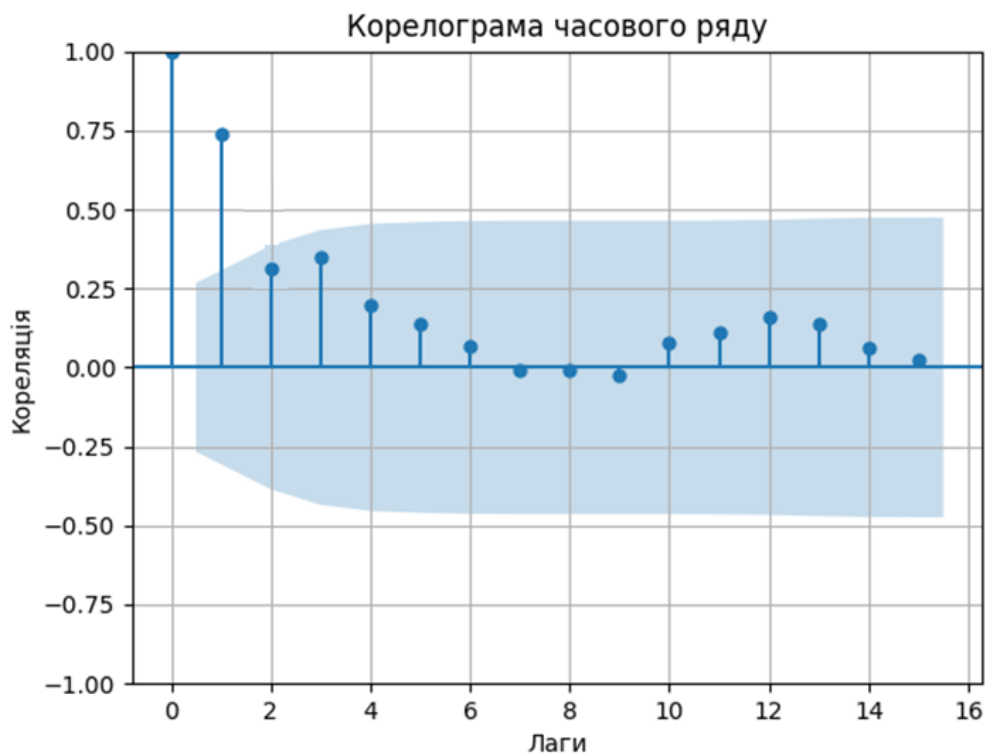


Рисунок 3.21 – Корелограма часового ряду

Якщо самим більшим значущим є лаг порядку 1, то ЧР має тренд. Якщо корелограма має вигляд коливань – то є сезонні коливання, а їх період – це лаг, при якому значущий мінімум або максимум є найвищим. Якщо коливань корелограма не має – сезонності не має. Якщо усі значення корелограми не значущі (нижче синьої лінії) – ряд стаціонарний [30].

Аналіз корелограми часового ряду свідчить про те, що ряд має тренд і не має сезонної компоненти. Декомпозицію часового ряду зображено на рисунку 3.22.

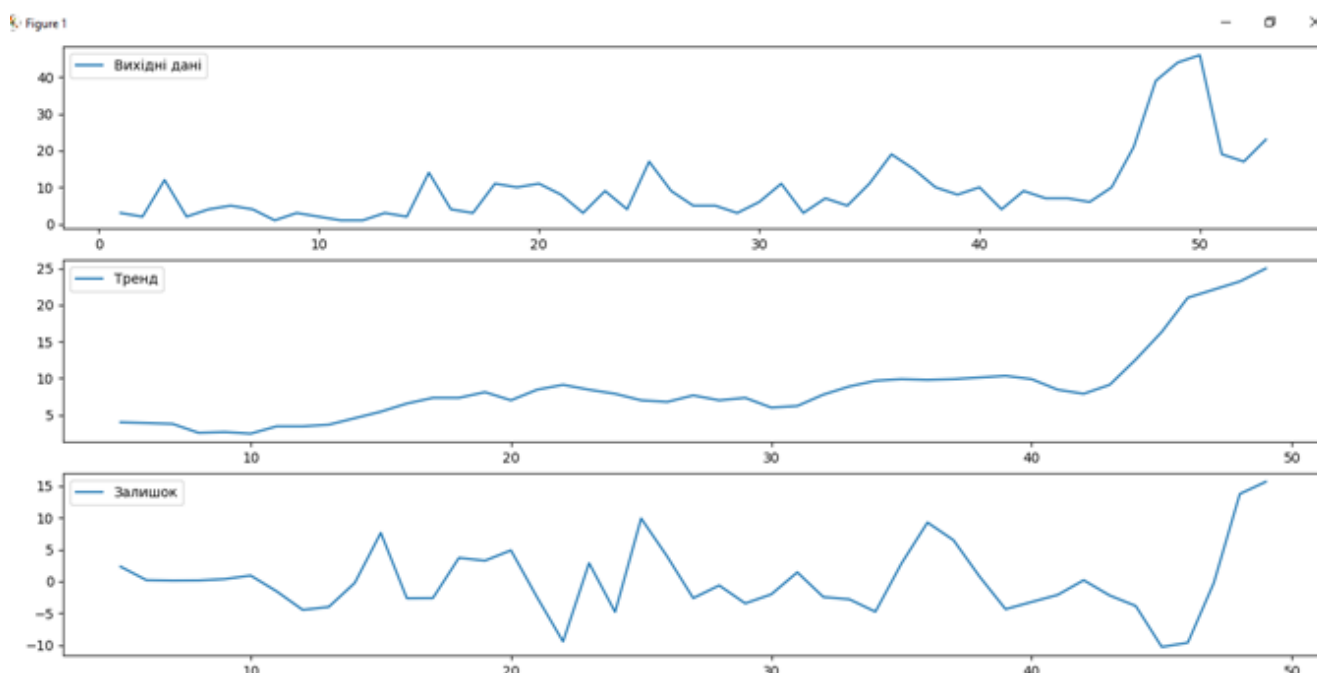


Рисунок 3.22 – Декомпозиція часового ряду

Якщо в часовому часовому ряду є тренд, то диференціювання може бути корисним інструментом для зменшення або видалення тренду перед прогнозуванням. Диференціювання є процес віднімання кожного елемента часового ряду з його попереднього значення. Це дозволяє прибрати тренд з даних і зробити ряд стаціонарним. Стаціонарний часовий ряд має постійне середнє значення та дисперсію у часі. Це спрощує моделювання та прогнозування ряду, оскільки стаціонарні ряди можуть бути проаналізовані з використанням статистичних методів, що передбачають стаціонарність.

Оскільки сезонної компоненти не було виявлено, для побудови прогнозної моделі було обрано модель ARIMA. Її застосування дало можливість отримати прогноз на 3 періоди вперед (рис. 3.23). На графіку відображено вихідні дані, тестові та прогнозні.

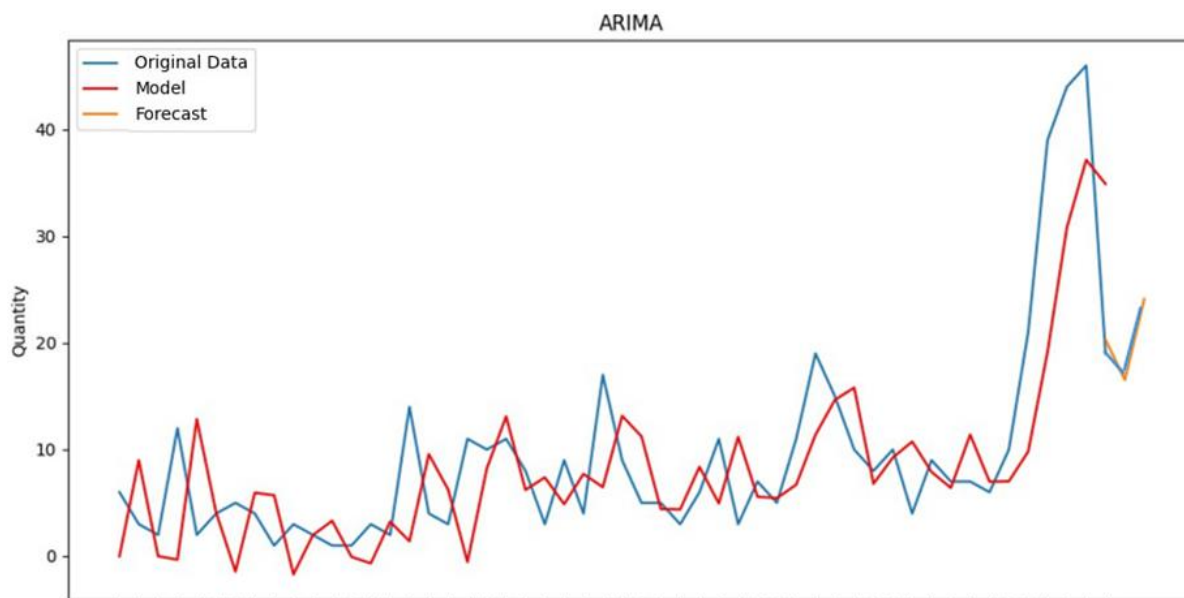


Рисунок 3.23 – Прогноз продажу деталей

Автокореляційний аналіз залишкової компоненти показав, що вона є стаціонарним рядом (рис. 3.24). Отже, прогнозна модель є адекватною.

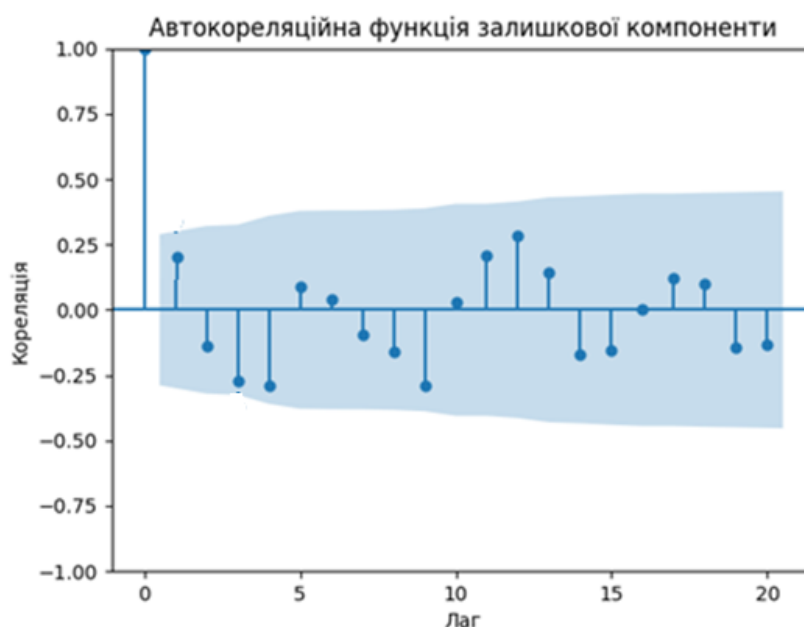


Рисунок 3.24 – Автокореляційна функція залишкової компоненти

Розраховані показники якості моделі показали, що точність отриманий прогноз має високу точність (рис. 3.25).

MSE: 1.0388879742490766
MAE: 0.9562130815050421
MFE: -0.6426051585977378
RMSE: 1.0192585414158062
MAPE: 4.794157379790894

Рисунок 3.25 – Розраховані показники якості моделі та точності прогнозу

Як можна побачити, згідно з графіком та показниками – модель робить досить успішний прогноз, похибки малі а точність складає 4,79%.

Висновки до розділу 3

У даному розділі було описано розробку та програмну реалізацію системи аналізу і прогнозування подажу товарів. Розроблена система дозволяє здійснювати очистку та попередню обробку набору даних зі значеннями продажу деталей автомашин у хронологічному порядку, виявляти структурні компоненти часового ряду, наявність тренду, стаціонарність. На основі здійсненого аналізу є можливість побудови ARMA моделі для стаціонарного часового ряду. Для нестационарного часового ряду у системі передбачено побудову моделей ARIMA та SARIMA з розрахунком показників, які визначають якість прогнозної моделі та точність прогнозу. На основі здійсненого аналізу є можливість обрати більш якісну модель, що має високу точність прогнозу, яку можна використовувати для прогнозування продажу товарів при плануванні закупки товару та формування маркетингової політики.

ВИСНОВКИ

Проведене дослідження дозволяє зробити наступні висновки. Ринок онлайн продаж є важливою складовою ринкової економіки,. Застосування аналізу та прогнозування часових рядів дозволяє отримувати важливу інформацію про майбутні показники продажу та допомагає в прийнятті обґрунтованих рішень щодо планування та діяльності комерційних фірм.

Установлено, що основним підходом до побудови прогнозу є моделювання часових рядів обсягів продажу за певний період часу, оскільки дозволяють враховувати тренди, сезонність та інші регулярні зміни у часі. Моделі часових рядів можуть ефективно адаптуватися до змін в споживчому попиті та інших факторах, що впливають на обсяги продаж.

Моделювання фінансових часових рядів є складною задачею, оскільки комерційна діяльність містить багато чинників, які можуть викликати ріст та падіння попиту на товари та їх вартість. На вибір методів аналізу та прогнозування впливає стаціонарність та не стаціонарність часового ряду. До методів аналізу часових рядів, які базуються на припущенні про стаціонарність ряду, відносять авторегресійну модель (AR-модель) та її модифікації. Для підвищення ефективності планування онлайн продаж інтернет-магазину необхідна автоматизована система, яка дозволяє здійснювати аналіз динаміки обсягів продаж та будувати модель часового ряду з високою точністю прогнозу.

Здійснений аналіз дозволив виявити велику кількість методів та моделей, які використовують для побудови прогнозу продажу товарів. Це методи прогнозування із використанням глибинних нейронних мереж, авторегресійні AR-моделі, MA-модель ковзних середніх, ЕМА-модель експоненційних ковзних середніх, регресійні методи моделювання шляхом побудови лінійної та нелінійної аналітичної функції, адитивна АМ-модель та узагальнена адитивна GAM-модель, мультиплікативна модель, ARIMA-модель інтегрованої авторегресії ковзного середнього та багато інших.

Важливе значення для побудови прогнозної моделі часового ряду має виявлення основних структурних складових часового ряду: тренду, сезонної, циклічної та випадкової компонент. Для прогнозування стаціонарного часового ряду будують авторегресійну AR-модель та модель ARMA. Для нестаціонарних часових рядів є два підходи до побудови прогностичної моделі: моделювання регулярних компонент у сукупності та розкладання часового ряду на компоненти та моделювання кожної компоненти окремо. Для часового ряду, у структурі якого виявлено тренд та відсутність сезонної компоненти, доцільною є модель ARIMA. Якщо у структурі часового ряду виявлено сезонну компоненту та відсутність тренду, можна передбачити побудову моделі SARIMA. У випадку наявності тренду та сезонності доцільно використовувати модель Хольта-Вінтерса.

Установлено, що для перевірки адекватності прогнозної моделі часового ряду необхідно здійснити дослідження випадкової компоненти на стаціонарність і розрахунок метрик оцінки якості та точності побудованої моделі: MAE, MSE, RMSE, MAPE.

Для розробки системи аналізу та прогнозування продажу було вирішено використати мову програмування Python, середовище розробки PyCharm та менеджера управління Python-пакетами PIP. Для аналізу часових рядів та створення, навчання й оцінки якості прогнозних моделей серед Python бібліотек було вирішено застосовувати Pandas, NumPy, Matplotlib, Scikit-Learn, Statsmodels. Для управління даними та виведення результатів було використано бібліотеку Matplotlib. Для аналізу часових рядів та побудови прогнозної моделі використано бібліотеки Pandas, Scikit-learn, Statsmodels.

Основною мовою для створення вебсторінок та структурування контенту обрано мову HTML та PHP – сценарну мову загального призначення з відкритим кодом, яка ідеально підходить для веб-розробки та може бути вбудована в HTML. Для веброзробки використано також середовище IDE PhpStorm, CSS для стилізації сторінок та JavaScript для створення інтерактивних вебсторінок. PHP було застосовано для вбудови серверного коду в HTML та доступу до серверу Apache і

роботи з базою даних СУБД MySQL із інформацією про онлайн-транзакції інтернет-магазину продажу авозапчастин. Для асинхронного програмування при запиті до серверу вирішено використовувати JavaScript.

Поставлені завдання виконано повністю, однак функціонал розробленого застосунку може бути вдосконалений у подальшому шляхом розширення можливостей для вибору прогнозних моделей.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. What is eCommerce. [Електронний ресурс] – Режим доступу: – <https://www.shopify.com/encyclopedia/what-is-ecommerce>.
2. Jan Manthey. E-Commerce: Innovations, Technologies, Platforms, and Applications. Springer, 2015. – 466 с.
3. Андрусенко Ю.О. Аналіз основних моделей прогнозування часових рядів. Збірник наукових праць Харківського національного університету Повітряних сил. 2020. № 3(65). С. 91-96.
4. Електронний ресурс] / John Edwards // СІО. – 2019. – Режим доступу до ресурсу: <https://www.cio.com/article/3273114/what-is-predictive-analyticstransforming-data-into-future-insights.html>.
5. Predictive Analytics History & Current Advances [Електронний ресурс] // SAS – Режим доступу до ресурсу: https://www.sas.com/en_us/insights/analytics/predictive-analytics.html.
6. Часові ряди : підручник. Прикарпаття, 2018. 82 с. URL: <https://kstat.pnu.edu.ua/wp-content/uploads/sites/63/2018/04/Часові-ряди.pdf> (дата звернення: 02.05.2023).
7. Bradlow, E.T., Gangwar, M., Kopalle, P., Voleti, S.: The role of big data and predictive analytics in retailing. *Journal of Retailing* 93(1) (2017) 79–95
8. Abbott D. (2014) *Applied predictive analytics: Principles and techniques for the professional data analyst*. Indianapolis, IN: John Wiley & Sons. 456 p.
9. The Importance Predictive Analytics for E-commerce Stores [Електронний ресурс] // The Startup. – 2018. – Режим доступу до ресурсу: <https://medium.com/swlh/the-importance-of-predictive-analytics-for-e-commercestores-d7ef0ce2d32e>
10. Мармоза А.Т. Теорія статистики : підручник. К.: Центр учбової літератури, 2013. 592 с.
11. Introduction to the Fundamentals of Time Series Data and Analysis. APTECH

: website. URL: <https://www.aptech.com/blog/introduction-to-the-fundamentals-of-time-series-data-and-analysis/> (дата звернення: 12.05.2023).

12. Долгіх А.О., Байбуз О.Г. Аналіз методів, моделей та програмних засобів прогнозування часових рядів. Відкриті інформаційні та комп'ютерні інтегровані технології. 2018. № 79. С. 74-85.

13. Стаціонарні та нестаціонарні часові ряди. Основні характеристики часових рядів : вебсайт. URL: <https://studfile.net/preview/2398204/page:22/> (дата звернення: 08.05.2023).

14. Montgomery D. C., Jennings C. L., Kulahci M. Introduction to Time Series Analysis and Forecasting. 2nd Edition. New Jersey : Wiley-Interscience, 2015. 515 p.

15. Андрусенко Ю. О. Аналіз основних моделей прогнозування часових рядів. Збірник наукових праць Харківського національного університету Повітряних сил. 2020. № 3(65). С. 91-96

16. Прогнозування продажу з використанням програмного засобу MatLab: вебсайт. URL: <https://naukarus.com/prognozirovanie-vyruchki-ot-prodazh-na-osnovesovremennyh-instrumentov-adaptivno-neyro-nechetkoy-seti-sistemy-matlab>

17. Прогнозування продажу з використанням програмного засобу Deductor: вебсайт. URL: <https://basegroup.com/solutions/tasks/demand-forecast>

18. GMDH Shell: вебсайт. URL: <https://gmdhsoftware.com/ru/neural-networksoftware/> (дата звернення: 09.11.2022).

19. STATISTICA Automated Neural Networks (SANN): вебсайт. URL: http://statsoft.ru/products/STATISTICA_Neural_Networks/

20. Моделювання і прогнозування по часових рядах : вебсайт. URL: https://stud.com.ua/75017/statistika/modelyuvannya_prognozuvannya_chasovih_ryadah (дата звернення: 07.05.2023).

21. Python Pandas DataFrame [Електронний ресурс]. – 2018. - Режим доступу: <https://www.geeksforgeeks.org/python-pandas-dataframe/>

22. Документація бібліотеки Matplotlib [Електронний ресурс]. – 2012. - Режим доступу: <https://matplotlib.org/contents.html> – Дата доступу: квітень 2020.

23. Огляд методів машинного навчання в задачі прогнозування фінансових часових рядів / Гурєєва К.М., Кудін О.В. Лісняк Ф.О. Вісник Запорізького національного університету. Фізико-математичні науки. 2018. №2. С. 18-28. URL: http://nbuv.gov.ua/UJRN/Vznu_mat_2018_2_5 (дата звернення: 05.05.2023).

24. Ляшенко О.І., Кравець Т.В., Хрущ Л.З. Застосування пакетів прикладних програм в економетричному моделюванні фінансових часових рядів. Економіко-математичне моделювання соціально-економічних систем. 2017. Вип. 22. С. 33-59. URL: <http://dspace.nbuv.gov.ua/bitstream/handle/123456789/132550/02-Liashenko.pdf?sequence=1> (дата звернення: 04.05.2023).

25. HTML5 specification. W3C Working Draft : вебсайт. URL: <https://www.w3.org/TR/2011/WD-html5-20110405/> (дата звернення: 13.05.2023).

26. CSS3 specification. W3C Group Draft Note : вебсайт. URL: <https://www.w3.org/TR/css3-roadmap/#intro> (дата звернення: 14.05.2023).

27. Котова О., Савинова В. Структура и информационно-логическая схема работы системы прогнозирования экономических временных рядов «СГМ Горизонт». Modern Economy Success. 2020. №5. С. 190-197.

28. Фінансове прогнозування : електронний посібник. Луцьк, 2018. URL: https://elib.lntu.edu.ua/sites/default/files/elib_upload/Електронний%20посібник%20ФІНАНСОВЕ%20ПРОГНОЗУВАННЯ/index.html (дата звернення: 06.05.2023).

29. Фінансове моделювання : електронний посібник. Луцьк, 2018. URL: https://elib.lntu.edu.ua/sites/default/files/elib_upload/Електронний%20посібник%20ФМ/index.html (дата звернення: 09.05.2023).

30. Прогнозування та аналіз часових рядів : методичні вказівки / укл. Юрченко М. Є. Чернігів : ЧНТУ, 2018. 88 с.