

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА»
ФАКУЛЬТЕТ ІНФОРМАЦІЙНО-КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНЖЕНЕРІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

ПОЯСНЮВАЛЬНА ЗАПИСКА

до кваліфікаційної роботи освітнього ступеня «магістр»
за спеціальністю 121 «Інженерія програмного забезпечення»
(освітня програма «Інженерія програмного забезпечення»)

на тему:

**«Аналіз та прогнозування продажів у роздрібній
торгівлі»**

Виконав студент групи ІПЗм-23-1
ПІДГАЄЦЬКИЙ Владислав Валентинович

Керівник роботи:
БАКАЛЮК Тетяна Анатоліївна

Рецензент:
ЛЕВКІВСЬКИЙ Віталій Леонідович

Житомир – 2024

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА»
ФАКУЛЬТЕТ ІНФОРМАЦІЙНО-КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНЖЕНЕРІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

«ЗАТВЕРДЖУЮ»

зав. кафедри інженерії
програмного забезпечення
Тетяна ВАКАЛЮК
«01» жовтня 2024 р.

ЗАВДАННЯ
на кваліфікаційну роботу

Здобувач вищої освіти: ПІДГАСЦЬКИЙ Владислав Валентинович

Керівник роботи: д.пед.н., професор ВАКАЛЮК Тетяна Анатоліївна

Тема роботи: **«Аналіз та прогнозування продажів у роздрібній торгівлі»**,
затверджена Наказом закладу вищої освіти від «01» жовтня 2024 р. № 520/с

Вихідні дані для роботи: **процеси аналізу та прогнозування продажів в роздрібній торгівлі; розробка програмного забезпечення для оптимізації процесів аналізу та прогнозування продажів в роздрібній торгівлі.**

Керівники з магістерської кваліфікаційної роботи із зазначенням розділів,
що їх стосуються:

Розділ	Керівник	Завдання видав	Завдання прийняв
1	Вакалюк Т. А.	01.10.2024р.	01.10.2024р.
2	Вакалюк Т. А.	01.10.2024р.	01.10.2024р.
3	Вакалюк Т. А.	01.10.2024р.	01.10.2024р.

РЕФЕРАТ

Кваліфікаційна робота освітнього ступеня «магістр» присвячена автоматизації процесу аналізу та прогнозування продажів у роздрібній торгівлі. Пояснювальна записка складається з вступу, трьох розділів та переліку використаної літератури. Пояснювальна записка до кваліфікаційної магістерської роботи містить 62 сторінки, 10 ілюстрацій, 3 таблиці, 22 формули, 3 додатки та 20 посилань на зовнішні джерела.

Метою роботи є створення системи для аналізу та прогнозування продажів у роздрібній торгівлі, розробка зручного графічного інтерфейсу для користування. В роботі описано процес створення програми, яка аналізує історичні дані та робить прогноз на наступний місяць. Досліджено способи машинного навчання. Проведено тестування програми.

КЛЮЧОВІ СЛОВА: АНАЛІЗ ДАНИХ, ПРОГНОЗ, МОДЕЛІ, LIGHTGBM, ARIMA, ABC-XYZ.

ABSTRACT

The qualification work of the educational degree "Master" is devoted to the automation of the process of analysis and forecasting of sales in retail. The explanatory note consists of an introduction, three sections and a list of references. The explanatory note to the qualification master's thesis contains 62 pages, 10 illustrations, 3 tables, 22 formulas, 3 appendices and 20 links to external sources.

The purpose of the work is to create a system for analyzing and forecasting sales in retail, developing a user-friendly graphical interface for use. The paper describes the process of creating a program that analyzes historical data and makes a forecast for the next month. Methods of machine learning have been investigated. The program was tested.

KEY WORDS: DATA ANALYSIS, FORECAST, MODELS, LIGHTGBM, ARIMA, ABC-XYZ.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	7
ВСТУП	8
РОЗДІЛ 1. ТЕОРЕТИЧНИЙ ОГЛЯД АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ПРОДАЖІВ В РОЗДРІБНІЙ ТОРГІВЛІ	10
1.1. Визначення та основні аспекти аналізу продажів в роздрібній торгівлі	10
1.1.1. Поняття роздрібною торгівлі	10
1.1.2. Важливість аналізу продажів для роздрібною торгівлі	11
1.1.3. Огляд аналогів.....	12
1.2. Формалізовані моделі для прогнозування	13
1.2.1. Часові ряди	14
1.2.2. Модель LightGBM.....	17
1.2.3. Моделі прогнозування попиту	19
1.3. Методи та алгоритми в контексті роздрібною торгівлі.....	20
1.3.1. Часові ряди в контексті роздрібною торгівлі	20
1.3.2. LightGBM в контексті роздрібною торгівлі	22
1.3.3. Управління продуктовим портфелем в роздрібній торгівлі	22
1.3.4. Постановка задачі	25
Висновки до розділу 1	26
РОЗДІЛ 2. РОЗРОБКА СИСТЕМИ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ПРОДАЖІВ В РОЗДРІБНІЙ ТОРГІВЛІ	27
2.1. Огляд бази даних	27
2.2. Методологія та інструменти розробки	30
2.2.1. Python як інструмент для аналізу та прогнозування	30
2.2.2. C# для створення інтерфейсу користувача	31
2.3. Розробка модулів аналізу продажів	33
2.3.1. Аналіз сезонності	33
2.3.2. Аналіз впливу рекламних акцій	33
2.3.3. ABC-XYZ аналіз продажів	34

2.4. Розробка модуля прогнозування продажів	36
2.5. Розробка інтерфейсу користувача	37
2.5.1. Етапи розробки інтерфейсу	38
2.5.2. Переваги інтерфейсу	39
2.6. Системні вимоги	40
Висновки до розділу 2	41
РОЗДІЛ 3. ОГЛЯД ТА АНАЛІЗ РЕЗУЛЬТАТІВ	43
3.1. Аналіз моделей прогнозування	43
3.2. Огляд результатів аналізу продажів	45
3.3. Аналіз результатів прогнозування продажів	49
Висновки до розділу 3	52
ВИСНОВКИ.....	54
ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ.....	55
ДОДАТКИ.....	57

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

БД – база даних

ACF – (англ. Autocorrelation Function) – функція автокореляції

ARIMA – (англ. AutoRegressive Integrated Moving Average) – авторегресивна інтегрована модель ковзного середнього

ETS – (англ. Error, Trend, Seasonality) – похибка, тенденція, сезонність

IBM – (англ. International Business Machines Corporation) – міжнародна корпорація бізнес-машин

LightGBM – (англ. Light Gradient Boosting Machine) – модель градієнтного бустингу

PACF – (англ. Partial Autocorrelation Function) – функція часткової автокореляції

ВСТУП

Актуальність теми. У сучасному динамічному світі роздрібні торговці стикаються з низкою викликів, які підкреслюють важливість аналізу та прогнозування продажів. Зміна споживчих уподобань, швидкі технологічні зміни та конкурентний тиск зробили аналіз і прогнозування ключовим елементом ефективного стратегічного управління в роздрібній торгівлі.

Актуальність теми «Аналіз та прогнозування продажів в роздрібній торгівлі» відображається в необхідності адаптації до швидко мінливих ринкових умов і своєчасного реагування на зміни споживчого попиту, конкуренції та інших впливових факторів.

Розвиток онлайн-торгівлі, глобалізація ринку та виклики економічної нестабільності роблять прогнозування ще більш складним. Використання сучасних технологій, таких як штучний інтелект та аналітика, дозволяє компаніям оптимізувати свої бізнес-стратегії та забезпечити свою конкурентоспроможність.

Аналіз та прогнози враховують не лише історичні дані та сезонні тенденції, але й глобальні та регіональні фактори, такі як зміна споживчого попиту, технологічні інновації та стратегії конкурентів. Різні бізнес-середовища та потреби клієнтів вимагають індивідуальних підходів до прогнозування, щоб забезпечити ефективне управління запасами та відповідати унікальним вимогам ринку.

Мета і завдання дослідження. Метою даного дослідження є вивчення та аналіз методів прогнозування продажів у секторі роздрібної торгівлі з метою розробки програмного забезпечення для оптимізації управління запасами та прогнозування продажів.

Встановлена мета обумовлює наступні завдання:

1. Провести огляд наукових статей, книг, та публікацій, що стосуються прогнозування та аналізу продажів у роздрібній торгівлі.
2. Обрати методи прогнозування, що задовільняють поставлену мету.
3. Обрати методи для оптимізації запасів.

Об'єктом дослідження є процеси аналізу та прогнозування продажів в роздрібній торгівлі.

Предметом дослідження є розробка програмного забезпечення для оптимізації процесів аналізу та прогнозування продажів в роздрібній торгівлі.

Практичне значення. Результати цього дослідження допоможуть магазинам і компаніям краще розраховувати, скільки товарів їм потрібно замовляти та продавати. Розроблене програмне забезпечення автоматизує ці процеси, допомагаючи уникати ситуацій, коли товарів або занадто багато, або не вистачає. Це дозволяє зменшити витрати, швидше реагувати на зміни ринку та бути конкурентоспроможними.

Методи дослідження. У ході роботи будуть використовуватись різноманітні методи дослідження, такі як статистичний аналіз ринкових даних та математичне моделювання. Інтеграція цих методів дозволяє отримувати комплексні та точні прогнози, необхідні для ефективного управління роздрібною торгівлею в умовах постійної динаміки ринку.

РОЗДІЛ 1. ТЕОРЕТИЧНИЙ ОГЛЯД АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ПРОДАЖІВ В РОЗДРІБНІЙ ТОРГІВЛІ

1.1. Визначення та основні аспекти аналізу продажів в роздрібній торгівлі

1.1.1. Поняття роздрібної торгівлі

Роздрібна торгівля є однією з форм господарсько-комерційної діяльності суб'єктів господарювання, а також однією з форм організації товарних ринків. Вона є невід'ємною частиною внутрішньої торгівлі.

Функції роздрібної торгівлі - це багатогранні сервісні обов'язки продавця перед кінцевим споживачем. Вони різняться залежно від форми бізнесу. Наприклад, у великих супермаркетах з самообслуговуванням деякі завдання вирішують самі покупці, а до фахівців звертаються за допомогою лише в разі потреби.

До функціональних завдань роздрібної торгівлі можна віднести наступні:

- Аналіз потреб потенційних клієнтів і формування асортименту відповідно до них.
- Надання інформації про товари, які купують клієнти.
- Заохочення виробників до виробництва конкретних товарів на основі даних досліджень.
- Послуги та консультування.

Важливою є також різноманітність продукції. Особливу увагу продавці повинні приділяти продажу складного технічного обладнання, оскільки тільки вони досить добре розбираються в тому, що продають.

Роздрібна торгівля характеризується наступними особливостями:

- Це самостійна, ініціативна, пов'язана з власним ризиком діяльність юридичних та фізичних осіб з метою отримання прибутку;
- Основним видом діяльності є продаж товарів, але можуть також здійснюватися супутні операції та надаватися різні послуги. Допоміжні послуги включають пакування, обгортання та складання. Роздрібні послуги включають пакування подарунків, доставку придбаних товарів та багато іншого;

- Товари продаються безпосередньо фізичним особам або іншим кінцевим користувачам для особистих, некомерційних цілей;
- Товари продаються за готівку або за допомогою методів оплати, еквівалентних готівці (наприклад, банківська кредитна картка, переказ на банківський рахунок).
- Роздрібні торговці можуть продавати лише певні види товарів (продовольчі та непродовольчі товари, алкогольні напої та тютюнові вироби, лікарські засоби та медичне обладнання, певні хімічні продукти, транспортні засоби та їх номери, ювелірні вироби, певні види зброї, боєприпаси та спеціальне обладнання, а також деякі інші товари). Ця сфера є вужчою, ніж оптова торгівля, але значно ширшою, ніж у сфері громадського харчування.[1-2]

1.1.2. Важливість аналізу продажів для роздрібної торгівлі

Аналіз продажів є важливою складовою успішного управління роздрібною торгівлею, оскільки він допомагає вирішувати ключові бізнес-завдання та впливає на різні аспекти роботи компанії. Завдяки цьому процесу можна ефективно планувати запаси, орієнтуючись на попит, що дозволяє уникати надлишків або дефіциту товарів. Це сприяє оптимізації витрат і забезпечує клієнтів потрібною продукцією, що підвищує їхню лояльність.

Крім того, аналіз продажів дозволяє виявляти найуспішніші рекламні та маркетингові стратегії, допомагаючи бізнесу адаптувати їх до сезону або ринкових умов. Завдяки цьому компанії можуть ефективніше залучати нових клієнтів і збільшувати продажі. Розуміння того, які товари є найбільш прибутковими, дозволяє бізнесу зосередитися на цих сегментах, що не лише підвищує дохід, але й зміцнює позиції на ринку, дозволяючи залишатися конкурентоспроможним.

Аналіз даних про продажі також є основою для покращення обслуговування клієнтів. Він дає змогу глибше розуміти їхні вподобання та потреби, персоналізувати пропозиції й створювати програми лояльності, що сприяє довготривалим відносинам із клієнтами. Окрім того, аналітика дозволяє прогнозувати

майбутні тенденції, що допомагає своєчасно реагувати на зміни попиту, розробляти нові продукти чи коригувати асортимент.

Ще однією перевагою аналізу є можливість підвищення ефективності операційного управління. Збираючи й аналізуючи дані, компанії можуть краще планувати поставки, оптимізувати виробничі процеси та знижувати витрати. Усе це забезпечує стабільний розвиток бізнесу та сприяє його довгостроковому зростанню. Таким чином, аналітика продажів стає незамінним інструментом для ухвалення обґрунтованих рішень і забезпечення ефективної роботи роздрібною торгівлі.[3]

1.1.3. Огляд аналогів

Існує ряд застосувань аналогів програмного забезпечення для аналізу та прогнозування продажів у роздрібній торгівлі. Ці інструменти дозволяють підприємствам різних галузей проводити глибокий аналіз ринку та вдосконалювати стратегії продажів. Це вкрай важливо в умовах постійно зростаючої конкуренції і швидкої зміни споживчих переваг. Ключовими користувачами такого програмного забезпечення є маркетологи, аналітики, менеджери з розвитку бізнесу та дослідники ринку в роздрібній торгівлі. Вони мають доступ до різноманітних інструментів, таких як аналіз покупок, вивчення попиту на конкретні товари чи послуги, прогнозування та відстеження ринкових тенденцій.

В результаті використання такого програмного забезпечення компанії можуть більш ефективно конкурувати на ринку, адаптувати свою стратегію до змін споживчого попиту, вдосконалювати свої продукти або послуги для задоволення потреб клієнтів.

Деякі існуючі платформи, такі як SAS Retail Forecasting, IBM Planning Analytics, Oracle Retail Predictive Application Server, надають можливості візуалізації та аналізу даних. З їх допомогою можна прогнозувати продажі в роздрібній торгівлі за допомогою вбудованих аналітичних інструментів і статистичних методів.

Наприклад, SAS Retail Forecasting спеціалізується на аналітиці та бізнес-інтелекті. Програма використовує статистичні методи, такі як ETS та ARIMA,

для детального аналізу часових рядів. Її основна особливість полягає в високій точності прогнозування, інтеграції з різними джерелами даних, а також оптимізації управління запасами.[4]

Існує також IBM Planning Analytics, що відома своєю експертизою в галузі інформаційних технологій. Це програмне забезпечення використовує прогностичні моделі та аналітичні інструменти для точного прогнозування споживчого попиту. Його ключовою особливістю є інтеграція з іншими системами управління роздрібною торгівлею та здатність оптимізації різних бізнес-процесів.[5]

Oracle Retail Predictive Application Server - це програмне забезпечення, яке базується на алгоритмах машинного навчання та надає точний аналіз та прогнозування великого обсягу даних. Його особливість полягає в інтеграції з іншими рішеннями Oracle Retail, що дозволяє створювати комплексні стратегії управління продажами.[6]

Підводячи підсумок, можна сказати, що існує ряд ключових пріоритетів, які слід враховувати при розробці програмного забезпечення для прогнозування продажів в роздрібній торгівлі. Система повинна бути гнучкою і легко адаптуватися до змін умов ведення бізнесу і даних. Це включає в себе здатність адаптуватися до нових технологій, методів аналізу, змін споживчого попиту.

Не менш важливо створити зручне середовище для взаємодії. Користувачам має бути легко користуватися системою без особливого навчання. Інтуїтивно зрозумілий інтерфейс і доступ до ключових функцій дозволяють ефективно взаємодіяти з програмою.

1.2. Формалізовані моделі для прогнозування

Формалізовані моделі прогнозування є важливим інструментом управління попитом та оптимізації бізнес-процесів. Ці моделі використовуються для математичного опису різних аспектів та допомагають приймати обґрунтовані стратегічні та тактичні рішення.

1.2.1. Часові ряди

Часовий ряд – це сукупність вимірювань деякої змінної величини, які проводяться у часі. Характерною особливістю часового ряду є те, що спостереження за об'єктом здійснюються безперервно в часі. Наприклад, щоденні продажі продукту, вартість акцій компанії, рівень інфляції, обмінний курс – все це часові ряди.

Для аналізу часових рядів вирішальним є порядок у рядах, тобто час є одним із визначальних факторів. Це відрізняє часові ряди від звичайних випадкових вибірок, де нижні індекси вводяться просто для легшої ідентифікації. Фундаментальна відмінність між часовими рядами та простим статистичним агрегуванням полягає в наступному:

- По-перше, рівні часового ряду не є незалежними. Іншими словами, якщо майбутні значення змінної можуть бути визначені, то вони є функціями минулих значень цієї змінної;
- По-друге, рівні часових рядів розподілені нерівномірно. Розподіл ймовірностей цих випадкових величин, особливо їх математичне сподівання та дисперсії, може залежати від часу.

Незалежно від характеру кожного часового ряду, можна виділити наступні основні типи проблем, які зазвичай розглядаються під час початкового аналізу даних. На першому етапі вони намагаються побудувати просту математичну систему або модель, яка описує поведінку часових рядів у короткій формі. Потім пробують пояснити його поведінку за допомогою інших змінних і з'ясувати ступінь зв'язку між спостереженнями однієї послідовності та між різними послідовностями.

До основних методів аналізу часових рядів можна віднести метод ковзного середнього, експоненціального згладжування і проектування тренду.

Метод рухомого (ковзного) середнього:

Цей метод полягає в тому, що розрахунок показника на прогнозований момент часу будується шляхом усереднювання значень цього показника за декілька попередніх моментів.

$$f_k = \frac{1}{N} \sum_{i=1}^N x_{k-i}, \quad (1.1)$$

де x_{k-i} – реальне значення показника у момент часу t_{k-i} ;

N – число попередніх моментів часу;

f_k – прогноз на момент часу t_k .

Метод зваженого (ковзного) середнього:

При прогнозуванні методом усереднювання часто доводиться спостерігати, що ступінь впливу використаних при розрахунку реальних показників виявляється неоднаковим, при цьому звичайно більш «свіжі» дані мають більшу вагу. Математично метод зваженого рухомого середнього можна записати як:

$$f_k = \frac{\sum_{i=1}^N w_{k-i} x_{k-i}}{\sum_{i=1}^N w_{k-i}}, \quad (1.2)$$

де x_{k-i} – реальне значення показника у момент часу;

N – кількість попередніх моментів часу, що використовуються при розрахунку;

f_k – прогноз на момент часу t_k ;

w_i – вага, з якою використовується показник x_{k-i} при розрахунку.

Вага – це завжди додатне число. У разі, коли вся вага однакова, ми одержуємо формулу (1.1).

Метод експоненціального згладжування:

Метод експоненціального згладжування є одним із методів прогнозування часових рядів. Цей метод часто використовується для прогнозування майбутніх значень на основі минулих спостережень.

Цей метод працює за принципом, що останнім даним призначається вищий ваговий коефіцієнт, тим самим надаючи меншу вагу старішим спостереженням. Цей метод використовує експоненціально зважене середнє для обчислення прогнозованого значення. Формула:

$$F_{t+1} = a \cdot X_t + (1 - a) \cdot F_t, \quad (1.3)$$

де F_{t+1} – прогноз для наступного періоду;

X_t – значення часового ряду в момент часу t ;

F_t – прогноз для поточного періоду;

α – параметр згладжування ($0 < \alpha < 1$).

Метод проектування тренду:

Основна ідея методу проекції тренду полягає в побудові прямої лінії з найменшим можливим «середнім» відхиленням від масиву точок (t, x) заданого часового ряду який описується рівнянням :

$$x = at + b, \quad (1.4)$$

де a, b – сталі коефіцієнти.

Коефіцієнти a і b розраховуються методом найменших квадратів. Щоб знайти коефіцієнти a і b , виконують такі дії: для кожного значення t і змінної x , користуючись формулою (1.4), обчислюють відповідне значення змінної x_i : $x_i = at_i + b, i = 1, 2, \dots, n$. Потім знаходять різницю $at_i + b - x_i, i = 1, 2, \dots, n$, яку потім підносять до квадрату: $(at_i + b - x_i)^2, i = 1, 2, \dots, n$.

Функція приймає мінімальне значення у тому випадку, коли величини a і b задовольняють таку лінійну систему:

$$\begin{aligned} \varphi(a, b) &= \sum_{i=1}^n (at_i + b - x_i)^2 \rightarrow \min, \\ \begin{cases} a \sum_{i=1}^n t_i + b_n = \sum_{i=1}^n x_i \\ a \sum_{i=1}^n t_i^2 + b \sum_{i=1}^n t_i = \sum_{i=1}^n t_i x_i \end{cases} \end{aligned} \quad (1.5)$$

Ця система завжди має єдине рішення.

Модель ARIMA:

ARIMA – це клас моделей часових рядів, який комбінує в собі компоненти авторегресії (AR), інтеграції (I) та рухомого середнього (MA). Ці моделі широко використовуються для аналізу та прогнозування часових рядів.

AR: ця частина ARIMA поєднує минулі значення залежної змінної для прогнозування майбутніх значень. Математично це виглядає так:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t, \quad (1.6)$$

де Y_t – поточне значення часового ряду;

c – константа; $\phi_1, \phi_2, \dots, \phi_p$ – параметри авторегресії;

$Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ – попередні значення часового ряду;

ε_t – біла шумова компонента.

I: Ця частина включає різницю між послідовними значеннями для стабілізації часового ряду. Щоб перетворити часовий ряд у стаціонарну форму:

$$Y'_t = Y_t - Y_{t-1}, \quad (1.7)$$

МА: ця частина враховує минулі значення білого шуму, щоб передбачити майбутні значення. Математично це виглядає так:

$$Y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad (1.8)$$

де μ – середнє значення часового ряду;

$\theta_1, \theta_2, \dots, \theta_q$ – параметри рухомого середнього;

$\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ – попередні значення білого шуму.

Загальна формула ARIMA(p, d, q) виглядає так:

$$Y'_t = c + \phi_1 Y'_{t-1} + \phi_2 Y'_{t-2} + \dots + \phi_p Y'_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad (1.9)$$

де p - порядок авторегресії;

d - ступінь інтеграції;

q - порядок рухомого середнього.[7]

1.2.2. Модель LightGBM

LightGBM — це сучасний і швидкий алгоритм машинного навчання, створений для задач прогнозування та класифікації. Його основна ідея — побудова ансамблю дерев рішень, кожне з яких поступово зменшує помилки попередніх дерев. LightGBM є популярним вибором через високу точність і здатність працювати з великими наборами даних.

Замість створення дерев рівномірної глибини, LightGBM додає вузли до гілки з найбільшою помилкою. Цей підхід дозволяє краще мінімізувати втрати, але може призвести до нерівномірно глибоких дерев.

Формула для обчислення приросту в листі:

$$\Delta = \frac{(\sum_{i \in G} g_i)^2}{\sum_{i \in G} h_i + \lambda}, \quad (1.10)$$

де G – набір спостережень у листі.

g_i – градієнти для кожного спостереження.

h_i – гессіани (друга похідна функції втрат).

λ – параметр регуляризації, який запобігає перенавчанню.

LightGBM підтримує паралельну та розподілену обробку даних, що дозволяє навчати моделі на великих наборах даних. LightGBM обробляє категоріальні ознаки без необхідності кодування one-hot encoding. Це зменшує обсяг пам'яті та покращує швидкість роботи. Алгоритм використовує метод гістограм для обчислення значень розбиття. Діапазон значень кожної ознаки розбивається на бінарні сегменти, що значно прискорює процес навчання.

Формула для побудови гістограм:

$$H[j] = \sum_{i \in B_j} g_i, \quad (1.11)$$

$$S[j] = \sum_{i \in B_j} h_i, \quad (1.12)$$

де $H[j]$ – градієнти для бінарного сегмента j ,

а $S[j]$ – гессіани.

Алгоритм належить до класу моделей градієнтного бустингу. Це означає, що він мінімізує спеціальну математичну функцію, яка описує помилку прогнозу. Наприклад, для задач регресії (прогнозування числових значень) функція втрат виглядає так:

$$L(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (1.13)$$

де y_i – реальне значення,

$\hat{y}_i$ – передбачене значення,

n – кількість прикладів у наборі даних.

Кожне дерево в LightGBM додається до моделі з урахуванням похідних цієї функції, які вказують, як сильно алгоритм помилився, і як потрібно скорегувати прогноз. У цьому процесі використовуються градієнти (перші похідні) і гессіани.

Особливістю LightGBM є спосіб побудови дерев. Замість того, щоб рівномірно розгалужувати дерево, як це роблять інші алгоритми, LightGBM працює з окремими гілками, що дозволяє знаходити складні залежності в даних і зменшувати помилки. Формула для обчислення приросту інформації при розділенні вуз-

ла виглядає так:

$$Gain = \frac{1}{2} \left(\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right), \quad (1.14)$$

де G_L, G_R – суми градієнтів для лівого та правого вузлів,

H_L, H_R – суми гессіанів,

Щоб уникнути перенавчання, LightGBM має вбудовані механізми, такі як обмеження глибини дерева та L1-/L2-регуляризація, які допомагають моделі залишатися узагальненою.[8]

1.2.3. Моделі прогнозування попиту

Модель дифузії Басса

Модель дифузії Басса — це теоретична модель, яка використовується для пояснення процесу впровадження нових продуктів або інновацій на споживчому ринку. Розроблена Френком Бассом у 1969 році, ця модель відображає розподіл нових продуктів серед споживачів у часі. Модель Басса особливо корисна для прогнозування запуску нових продуктів, оцінки можливостей ринку та розуміння впливу маркетингових зусиль на впровадження. По-перше, це засновано на концепції, що адаптери бувають різних форм:

1. Інноватори (p): вони перші користувачі повністю нових продуктів або інновацій. Вони схильні ловити випадкові речі та починати рано.
2. Імітатори (q): Це люди, які приймають абсолютно новий продукт після того, як побачили, що його використовує друг або новатор. Вони надихаються джерелами та підказками ідей інших людей.

Формулювання моделі:

$$\frac{f(t)}{1-F(t)} = p + qF(t), \quad (1.15)$$

де $f(t)$ – функція зміни встановленої базової частини;

$F(t)$ – функція встановленої базової частини;

p – коефіцієнт інновацій;

q – коефіцієнт імітації.

Продажі $S(t)$ – швидкість зміни встановленої бази $f(t)$ помножена на граничний потенціал ринку m :

$$S(t) = mf(t),$$

$$S(t) = m \cdot \frac{(p+q)^2}{p} \frac{e^{-(p+q)t}}{\left(1 + \frac{q}{p} e^{-(p+q)t}\right)^2} \quad (1.16)$$

Час пікових продажів t^* :

$$t^* = \frac{\ln q - \ln p}{p+q}, \quad (1.17)$$

Коефіцієнт p називають коефіцієнтом інноваційності, коефіцієнтом зовнішнього впливу або коефіцієнтом рекламного ефекту. Коефіцієнт q називають коефіцієнтом імітації, внутрішнім впливом або ефектом «з вуст в уста».[9]

1.3. Методи та алгоритми в контексті роздрібної торгівлі

1.3.1. Часові ряди в контексті роздрібної торгівлі

Коефіцієнт згладжування (α) у методі експоненціального згладжування визначає важливість останнього спостереження при формуванні прогнозу. Цей коефіцієнт приймає значення від 0 до 1, де 0 означає, що прогноз повністю залежить від попередніх значень, а 1 означає, що прогноз базується лише на останньому спостереженні.

Зменшення значення α призводить до більш плавного згладжування, яке враховує попередні значення, тоді як збільшення α надає більшій ваги останньому спостереженню.

Також, для прогнозування продажів в роздрібній торгівлі краще використовувати формулу трійного експоненціального згладжування (покращена формула 1.3).

Потрійне експоненціальне згладжування – це метод прогнозування часових рядів, який враховує тенденції та сезонні коливання, що дає змогу точніше моделювати та прогнозувати дані, коли відбуваються динамічні зміни в часі. Формула потрійного експоненціального згладжування:

$$F_t = \alpha(Y_t - S_{t-L}) + (1 - \alpha) \cdot (F_{t-1} + T_{t-1}), \quad (1.18)$$

де F_t – прогноз для періоду t ;

Y_t – фактичне спостереження для періоду t ;

S_{t-L} – сезонний ефект для попереднього періоду;

F_{t-1} – прогноз для попереднього періоду;

T_{t-1} – тренд для попереднього періоду;

α – коефіцієнт згладжування ($0 < \alpha < 1$); L - довжина сезону.

Щоб додати тренд і сезонний ефект, потрібно використовувати наступні формули:

$$T_t = \beta \cdot (F_t - F_{t-1}) + (1 - \beta) \cdot T_{t-1}, \quad (1.19)$$

$$S_t = \gamma \cdot (Y_t - F_t) + (1 - \gamma) \cdot s_{t-L}, \quad (1.20)$$

де β і γ - коефіцієнти згладжування для тренду і сезонності.

Цей метод дозволяє врахувати не тільки основну тенденцію часових рядів, а й сезонні коливання. Тривалість сезону (L) визначає кількість періодів (наприклад, місяців або кварталів), протягом яких повторюються сезонні зміни.[10]

- Модель ARIMA:

ARIMA позначається як $ARIMA(p, d, q)$. Щоб застосувати ARIMA для прогнозування роздрібної торгівлі потрібно визначити стаціонарність та параметри(p, d, q).

Стаціонарність є важливою концепцією в аналізі часових рядів і моделюванні часових процесів. Стаціонарні часові ряди мають постійну середню та незалежну від часу мінливість. Це полегшує моделювання та прогнозування ряду. Щоб визначити стаціонарність часового ряду, можна використовувати метод розрахунку ACF і PACF: якщо ACF і PACF зменшуються дуже повільно і зберігають статистичну значущість, це може вказувати на не стаціонарність ряду. Ряд повинен бути стаціонарним, у іншому випадку потрібно робити диференціювання до тих пір, поки ряд не стане стаціонарним.

ACF і PACF є інструментами для визначення авто-кореляційних лагів у часових рядах і відіграють ключову роль у визначенні параметрів моделі

ARIMA. Функція автокореляції визначає кореляцію між часовим рядом і його версіями з відставанням. Формула ACF для лага k має вигляд:

$$ACF(k) = \frac{\sum_{t=1}^{T-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^T (Y_t - \bar{Y})^2}, \quad (1.21)$$

де Y_t – значення часового ряду в момент часу t ;

$\bar{Y}$ – середнє значення часового ряду;

T – кількість спостережень у ряді.

Функція часткової автокореляції визначає кореляцію між часовим рядом та його лагованими версіями, враховуючи вплив інших лагів. Формула PACF для лага k виглядає наступним чином:

$$PACF(k) = \frac{cov(Y_t, Y_{t+k} - \hat{Y}_{t+k} | Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1})}{var(Y_t | Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1})}, \quad (1.22)$$

де $\hat{Y}_{t+k}$ – прогнозоване значення часового ряду в момент часу $t + k$ на основі значень від $t + 1$ до $t + k - 1$;

cov – коваріація;

var – дисперсія.

Обидві функції використовуються для аналізу автокореляції часових рядів і визначення параметрів моделі ARIMA. Зокрема, PACF дозволяє визначити параметри p (порядок авторегресії) і q (порядок ковзного середнього), тоді як ACF можна використовувати для визначення параметра d . [11]

- *Застосування:*

- Прогнозування тенденцій та сезонності у споживацькому попиті.
- Оптимізація управління запасами на основі історичних даних продажів.

1.3.2. LightGBM в контексті роздрібної торгівлі

Алгоритм LightGBM дозволяє значно покращити процеси прогнозування та аналізу, які є важливими для роздрібної торгівлі. Одне з основних завдань у цій сфері — це передбачення попиту на товари, що допомагає уникати втрат че-

рез надлишок чи дефіцит запасів. LightGBM аналізує дані, враховуючи сезонність, цінову політику та вплив рекламних кампаній. Наприклад, якщо попит на новорічні прикраси минулого року зріс у грудні, алгоритм прогнозує схожий сценарій і цього року, враховуючи можливі зміни в ціні чи маркетинговій стратегії.

LightGBM також допомагає оптимізувати управління запасами, автоматизуючи розподіл товарів між складами та магазинами. Він прогнозує попит для кожного регіону, враховуючи місцеві особливості, та допомагає мінімізувати логістичні витрати, прогнозуючи потрібні обсяги поставок. Наприклад, якщо алгоритм прогнозує високий попит на певний товар влітку, компанія може заздалегідь збільшити його запаси на складах.

Ще одним важливим напрямом є крос-продажі, які спрямовані на збільшення середнього чека. LightGBM аналізує поведінку покупців і визначає товари, які вони ймовірно придбають разом. Наприклад, якщо клієнт купує кавоварку, алгоритм може запропонувати кавові капсули чи аксесуари. Це базується на аналізі історії покупок і схожої поведінки інших клієнтів. Завдяки такій аналітиці компанії можуть ефективніше планувати свої стратегії, покращуючи обслуговування та збільшуючи продажі.

1.3.3. Управління продуктовим портфелем в роздрібній торгівлі

ABC-аналіз – це метод, який класифікує ресурси за ступенем їхнього впливу на прибуток компанії. У більшості випадків метод ABC застосовується до товарного асортименту для визначення прибутковості групи продуктів або окремих позицій і для коригування товарного асортименту.

В основі методу лежить принцип Парето: 20 відсотків зусиль дають 80 відсотків результатів. Стосовно товарного асортименту його можна сформулювати наступним чином: 20% продуктів генерують 80% продажів.

Суть цього методу полягає в ранжуванні товарів з точки зору їхнього внеску в загальну прибутковість. В результаті всі товари поділяються на три групи:

- Група А: найцінніші товари. На ці товари припадає 20% всього асортименту і 80% продажів.

- Група В: Проміжні товари. На них припадає 30% асортименту і 15% продажів.
- Група С: товари з найнижчою вартістю. На них припадає 50-60% асортименту і 5% продажів.

Таким чином, продаж товарів групи А є прибутковим і на ньому варто сконцентруватися, тоді як товари групи С приносять лише мінімальний прибуток – їх кількість можна сміливо зменшувати або виводити з асортименту.

АВС-аналіз не має сенсу, якщо він проводиться лише один раз. Необхідно аналізувати ресурси щонайменше раз на рік, бажано щоквартально, і оцінювати результати за кілька періодів. Таким чином можна відстежувати прибутковість товару в часі.[12]

Щоб доповнити аналіз запасів, компанії використовують метод XYZ, основна мета якого - дослідити стабільність показників: АВС-аналіз допомагає визначити внесок окремих елементів у кінцевий результат, тоді як XYZ-аналіз виявляє відхилення та нестабільність у продажах.

За результатами XYZ-аналізу можна виділити наступні групи:

- Група Х - високий рівень стабільності та точного прогнозування (коефіцієнт варіації до 10%);
- Група Y - із середнім рівнем стабільності та середньої передбачуваності (коефіцієнт варіації від 10 до 25 відсотків);
- Група Z - нестабільні цілі з низькою передбачуваністю (коефіцієнт варіації від 25%).

Алгоритм XYZ-аналізу складається з декількох етапів. Спочатку розраховується коефіцієнт варіації – це відносний показник, який характеризує варіабельність, тобто мінливість ознаки. Його визначають за певний період часу. Потім продукти групуються в напрямку більшого коефіцієнта варіації. Проаналізовані показники поділяються на категорії Х, Y та Z відповідно.

Щоб виконати комбінований аналіз, спочатку необхідно провести АВС-аналіз, а потім XYZ-аналіз цих продуктів за той самий період часу, наприклад, за кількістю щомісячної реалізації за рік. Потім результати об'єднуються.

При об'єднанні визначаються дев'ять товарних груп:

- Товари в групах AX і BX добре продаються, і прогнозування попиту на них є простим;
- Товари в групах AY і BY продаються добре, але прогнозування попиту є складним;
- Товари груп AZ і BZ продаються добре, але попит непередбачуваний;
- Товари групи CX приносять певний прибуток, але попит регулярний;
- Товари групи CY приносять певний прибуток, але попит нерегулярний;
- Усі товари групи CZ є новими продуктами і є продуктами спонтанного попиту, які постачаються на замовлення.[13]

Розглядання можливих ризиків та невизначеності у попиті та постачанні допомагає розробити резервні плани та стратегії для управління несподіваними ситуаціями.

Таблиця 1.1

Матриця характеристик груп об'єктів за результатами ABC-XYZ-аналізу

	A	B	C
X	- Висока прибутковість - Регулярність продажів - Високий рівень прогнозованості	- Середня прибутковість - Регулярність продажів - Високий рівень прогнозованості	- Низька прибутковість - Регулярність продажів - Високий рівень прогнозованості
Y	- Висока прибутковість - Періодичність продажів - Середній рівень прогнозованості	- Середня прибутковість - Періодичність продажів - Середній рівень прогнозованості	- Низька прибутковість - Періодичність продажів - Середній рівень прогнозованості
Z	- Висока прибутковість - Нерегулярність продажів - Низький рівень прогнозованості	- Середня прибутковість - Нерегулярність продажів - Низький рівень прогнозованості	- Низька прибутковість - Нерегулярність продажів - Низький рівень прогнозованості

1.3.4. Постановка задачі

Проведений аналіз дозволив виявити і систематизувати методи для покращення управління запасами та прогнозування продажів. Можна перерахувати коло задач, які необхідно вирішити на теоретичному й експериментальному рівні

в рамках аналізу та прогнозування продажів в роздрібній торгівлі:

1. Зібрати та проаналізувати історичні дані про продажі з різних точок продажу та онлайн-платформ.
2. Дослідити вплив факторів, таких як сезонність, маркетингові кампанії на зміну обсягів продажів.
3. Розробити модель аналізу та прогнозування продажів, використовуючи машинне навчання.
4. Провести тестування розробленої моделі на історичних даних та оцінити її точність та ефективність.

Висновки до розділу 1

У першому розділі був проведений огляд теоретичних аспектів аналізу та прогнозування продажів у роздрібній торгівлі. також, були висвітлені поняття роздрібною торгівлі, її особливості, функції та значення аналізу продажів. Важливість аналізу виконується в оптимізації запасів, формуванні ефективних маркетингових стратегій, підвищенні прибутковості та утриманні клієнтів.

Розглянуті основні моделі та методи прогнозування, включаючи часові ряди, моделі машинного навчання, такі як LightGBM, і методи управління продуктовим портфелем, наприклад аналіз ABC-XYZ. Ці підходи довели свою ефективність у прогнозуванні попиту та оптимізації запасів у роздрібній торгівлі.

Також було підкреслено значення інтеграції сучасних технологій та алгоритмів у процеси розробки рішень, що дає змогу підвищити конкурентоспроможність та адаптивність бізнесу до змінених умов ринку. Огляд аналогів продемонстрував широкий спектр інструментів, які вже застосовуються в цій сфері, та їх вплив на розвиток роздрібною торгівлі.

Загалом, розділ заклав теоретичну основу для розробки практичної системи аналізу та прогнозування продажів, яка буде висвітлена в наступних розділах роботи

РОЗДІЛ 2. РОЗРОБКА СИСТЕМИ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ПРОДАЖІВ В РОЗДРІБНІЙ ТОРГІВЛІ

2.1. Огляд бази даних

Дата	Товар	Кількість_продано	Ціна_закупівлі	Ціна_продажі	Сезон	Рекламна_акція
2023-01-01	Товар А	2144	152,07	182,8	Зима	Ні
2023-01-01	Товар Б	1255	219,3	273,6	Зима	Так
2023-01-01	Товар В	818	300,14	389,84	Зима	Ні
2023-01-01	Товар Г	2645	404,54	545,02	Зима	Ні
2023-01-02	Товар А	953	152,07	182,8	Зима	Ні
2023-01-02	Товар Б	1657	219,3	273,6	Зима	Так
2023-01-02	Товар В	1974	300,14	389,84	Зима	Ні
2023-01-02	Товар Г	1315	404,54	545,02	Зима	Ні
2023-01-03	Товар А	2257	152,07	182,8	Зима	Ні
2023-01-03	Товар Б	2677	219,3	273,6	Зима	Так
2023-01-03	Товар В	2000	300,14	389,84	Зима	Ні
2023-01-03	Товар Г	463	404,54	545,02	Зима	Ні

Рисунок 2.1. База даних

Для прогнозування продажів у роздрібній торгівлі використовували набір даних, що містить інформацію про щоденні продажі товарів. Дані організовані у вигляді таблиці, в якій кожен стовпець відповідає за певну характеристику, важливу для аналізу та побудови моделей прогнозування.

Стовпчик «Дата» вказує на конкретний день, коли було зафіксовано продажі. Ця інформація є основою для аналізу часових рядів, що дозволяє виявляти зміни попиту з плином часу. Крім того, дати дозволяють визначати сезонні закономірності, такі як зростання продажів у певні періоди року, наприклад, перед святами або в літній сезон.

Стовпчик «Товар» містить назву або унікальний ідентифікатор кожного товару. Завдяки цій змінній можна аналізувати продажі на рівні окремих продуктів і будувати індивідуальні моделі прогнозування для кожного з них. Це важливо, оскільки товари можуть мати різні тренди та залежати від різних факторів.

Стовпчик «Кількість_продано» показує кількість одиниць товару, проданих у конкретний день. Це ключовий показник, який є основною метрикою для прогнозування. Моделі навчаються на основі історичних даних про кількість

проданих одиниць і використовують цю інформацію для передбачення майбутніх продажів.

Стовпчик «Ціна_закупівлі» відображає вартість товару на момент його закупівлі. Ця інформація дозволяє обчислювати різницю між ціною закупівлі та продажу, що є важливим показником для аналізу рентабельності. Крім того, різниця між цими цінами може впливати на попит, оскільки зміни закупівельної ціни часто пов'язані з ціноутворенням.

Стовпчик «Ціна_продажі» вказує на вартість, за якою товар був реалізований. Ціна продажу може змінюватися залежно від багатьох факторів, включаючи проведення акцій, зміну попиту або сезонні знижки. Аналіз цього показника допомагає зрозуміти, як ціна впливає на кількість проданих одиниць.

Стовпчик «Сезон» є категоріальною змінною, яка вказує на пору року, наприклад, зиму, весну, літо чи осінь. Ця змінна має велике значення для прогнозування, оскільки попит на багато товарів залежить від сезонності. Наприклад, зимові товари матимуть більший попит у холодну пору року, тоді як літні товари популярні в теплі місяці.

Стовпчик «Рекламна_акція» є бінарною змінною, яка вказує на те, чи проводилася рекламна акція у день продажу. Значення цього стовпця (так або ні) допомагають враховувати вплив маркетингових заходів на продажі. Рекламні акції можуть суттєво збільшити попит у короткостроковій перспективі, тому їх врахування підвищує точність прогнозів.

Для ефективного прогнозування продажів базових даних недостатньо, тому було створено додаткові ознаки, що дозволяють враховувати різноманітні фактори, які впливають на попит. Кожна з цих ознак допомагає моделі краще адаптуватися до специфіки продажів.

Ознака «День_тижня» відображає, який саме день тижня відповідає кожній даті. Наприклад, будні дні зазвичай характеризуються меншим попитом, ніж вихідні. Завдяки цьому фактору модель може враховувати короткострокові зміни у попиті.

«Місяць» показує календарний місяць, у якому відбулися продажі. Ця ознака дозволяє враховувати середньострокові закономірності, такі як зростання продажів у грудні через новорічні свята або їх зниження влітку.

«Тиждень_року» вказує на номер тижня в межах року. Ця змінна допомагає враховувати періодичні зміни, які можуть повторюватися щотижня, наприклад, у різні тижні торгових періодів.

Ознака «Квартал» відображає, до якого з чотирьох кварталів року належить дата. Це важливий фактор для аналізу, адже попит у багатьох компаніях корелює із квартальними фінансовими звітами.

«Святковий_день» є бінарною змінною, яка вказує, чи відповідає дата офіційному святу. Передсвяткові дні часто характеризуються зростанням попиту, тоді як після свят продажі зазвичай знижуються.

Лагові ознаки (наприклад, «Лаг_1», «Лаг_7», «Лаг_14», «Лаг_30») вказують на кількість товару, проданого за певний період до поточного дня, тижня, двох тижнів або місяця. Вони допомагають моделі враховувати зв'язок між минулими продажами і поточним попитом.

Рухоме середнє (наприклад, «3-добове середнє», «7-добове середнє») є середнім значенням продажів за останні кілька днів. Ця ознака згладжує короткострокові коливання і допомагає краще виявляти тренди у даних.

Синус і косинус дня тижня перетворюють день тижня у синусоїдальні значення, що дозволяє моделі враховувати циклічні закономірності, характерні для тижневого циклу.

Ознака «Зміна_ціни» розраховується як різниця між поточною та попередньою ціною продажу. Вона дає змогу моделі визначати, як коливання ціни впливають на попит.

«Ціна_закупівлі_на_сезон» є результатом множення ціни закупівлі на значення сезону. Ця змінна враховує сезонний вплив на витрати компанії.

Ознака «Тренд» відображає кількість днів, що минули від початку періоду спостереження. Вона дає змогу враховувати довгострокові зміни, наприклад, поступове зростання популярності товару.

Усі ці додаткові ознаки створені для того, щоб допомогти моделям краще зрозуміти взаємозв'язки між різними факторами та точніше прогнозувати майбутні продажі. Завдяки цим ознакам моделі отримують змогу враховувати як короткострокові коливання, так і довгострокові тренди.

2.2. Методологія та інструменти розробки

Розробка системи аналізу та прогнозування продажів потребувала комплексного підходу, що охоплює вибір мов програмування, інструментів для роботи з даними, створення моделей прогнозування та розробку зручного графічного інтерфейсу. Основною метою цього підходу було створення ефективної, швидкої та зручної для кінцевого користувача системи. Для цього було використано сучасні технології та мови програмування, які відповідають потребам різних етапів роботи: аналізу даних, побудови моделей і взаємодії з користувачем.

2.2.1. Python як інструмент для аналізу та прогнозування

Мова програмування Python була обрана для виконання основних завдань, пов'язаних із обробкою даних, аналізом і прогнозуванням. Python є одним із найкращих інструментів для роботи з даними завдяки його популярності у сфері Data Science, великій кількості готових бібліотек і можливостям автоматизувати обробку великих обсягів інформації. Python дозволяє виконувати всі етапи роботи з даними: від їх завантаження та обробки до створення складних моделей прогнозування.

Однією з переваг Python є його простий і зрозумілий синтаксис, який дозволяє швидко писати код і адаптувати його для різних завдань. Завдяки цьому можна легко створювати нові модулі, вносити зміни або розширювати функціонал системи.

У системі використовувалися кілька ключових бібліотек Python, кожна з яких виконувала свою важливу функцію.

Pandas слугує основним інструментом для роботи з таблицями даних. Вона дозволяє завантажувати інформацію з Excel-файлів, обробляти її, виконувати фільтрацію, групування та інші операції. *Pandas* також застосовується для створення додаткових ознак, що сприяють покращенню точності прогнозів.

Бібліотека *Matplotlib* використовується для побудови графіків, які допомагають візуалізувати дані. За її допомогою можна створювати стовпчасті діаграми для відображення сезонності продажів або порівнювати обсяги продажів під час акцій і у звичайні дні.

LightGBM є сучасною бібліотекою машинного навчання, що спеціалізується на створенні швидких і ефективних моделей. Вона особливо підходить для роботи з великими обсягами даних і складними завданнями. У цьому проєкті *LightGBM* використовується для прогнозування майбутніх продажів на основі історичних даних.

Інструмент *Optuna* допомагає автоматизувати процес налаштування параметрів моделей машинного навчання. Завдяки цій бібліотеці можна швидко знаходити оптимальні значення параметрів, таких як кількість дерев у моделі чи швидкість навчання. Це дозволяє значно підвищити точність прогнозів і заощаджує час.

Бібліотека *OpenPyXL* надає можливість працювати з Excel-файлами безпосередньо у Python. Вона використовується для збереження результатів аналізу і прогнозів у зручному форматі, який легко переглядати або використовувати для подальшої роботи.[8]

2.2.2. C# для створення інтерфейсу користувача

Для створення інтерактивного інтерфейсу користувача було використано кілька бібліотек, які допомогли реалізувати функціонал програми, зробити її зручною у використанні та забезпечити високу продуктивність. Кожна бібліотека виконувала свою роль, доповнюючи загальну систему.

EPPlus використовується для роботи з Excel-файлами, включаючи зчитування даних, створення нових таблиць та запис результатів аналізу чи прогнозу-

вання. Ця бібліотека забезпечує просту інтеграцію програми з Excel, дозволяючи користувачам отримувати готові таблиці з результатами. Завдяки її зручності та потужним функціям EPPlus є оптимальним вибором для роботи з табличними даними.

Бібліотека *System.Diagnostics* дозволяє запускати зовнішні процеси, наприклад, виконання Python-скриптів із програми на C#. Вона забезпечує передачу параметрів, таких як шляхи до файлів, та отримання результатів їх виконання. Це дозволяє програмі ефективно взаємодіяти з обчислювальними модулями, створеними на Python.

System.Windows.Forms застосовується для створення графічного інтерфейсу користувача. Вона дозволяє розробляти елементи, як-от кнопки, текстові поля, таблиці чи панелі. Завдяки цій бібліотеці інтерфейс програми є інтуїтивно зрозумілим і зручним для роботи.

Для налаштування зовнішнього вигляду інтерфейсу використовується *System.Drawing*. З її допомогою додаються графічні елементи, як-от градієнти, круглі кнопки чи інші візуальні ефекти. Це робить дизайн програми сучасним і привабливим для користувача.

System.Data надає можливість роботи з таблицями даних у програмі. Ця бібліотека використовується для обробки результатів аналізу чи прогнозування перед їхнім відображенням у таблицях. Вона спрощує маніпуляцію даними, роблячи їх представлення зрозумілим і доступним.

Фонова обробка задач реалізується за допомогою *System.Threading.Tasks*. Ця бібліотека дозволяє виконувати тривалі операції, такі як аналіз чи прогнозування, без блокування основного потоку програми. Це забезпечує безперервну взаємодію користувача з інтерфейсом під час виконання складних обчислень.

Для роботи з файлами та директоріями застосовується *System.IO*. Вона дозволяє зчитувати файли, створювати нові папки для результатів і видаляти застарілі дані. Завдяки їй організація файлів у програмі стає впорядкованою та ефективною.

Бібліотека *System.ComponentModel* використовується для налаштування властивостей елементів інтерфейсу. Вона допомагає змінювати стилі й поведінку таких компонентів, як текстові поля чи таблиці, роблячи інтерфейс функціональним і зручним.

Обробка даних у колекціях і масивах виконується за допомогою *System.Linq*. Вона дозволяє здійснювати фільтрацію, сортування та інші операції з даними, які потім відображаються в програмі. Завдяки цій бібліотеці робота з даними є швидкою та ефективною.[14]

2.3. Розробка модулів аналізу продажів

Для того щоб система аналізу продажів могла давати корисну інформацію, було розроблено кілька модулів. Кожен із них виконує конкретну задачу: аналізує сезонність, оцінює ефективність рекламних акцій або виконує ABC-XYZ аналіз. Ці модулі допомагають краще зрозуміти поведінку продажів і приймати більш обґрунтовані бізнес-рішення. Нижче детально описано, як створювалися ці модулі.

2.3.1. Аналіз сезонності

Сезонність суттєво впливає на продажі багатьох товарів. Наприклад, зимовий одяг купують переважно взимку, а морозиво користується попитом улітку. Для врахування цих особливостей розробили модуль аналізу сезонності.

Спочатку дані про продажі розділяють за сезонами, використовуючи інформацію зі стовпця «Сезон» у базі даних. Це дозволяє побачити, як змінюється кількість проданого товару протягом року. Потім для кожного сезону підраховують загальну кількість продажів, що допомагає зрозуміти, коли товар купують найчастіше. Наприклад, якщо товар продавався як влітку, так і взимку, можна легко побачити, який сезон був більш успішним.

Результати аналізу зображають у вигляді графіка, щоб візуально показати, який сезон приносить найбільше продажів. Ця інформація є корисною для

планування закупівель і маркетингових активностей. Наприклад, заздалегідь підготувати склад до періоду, коли попит на товар найбільший.

2.3.2. Аналіз впливу рекламних акцій

Рекламні акції часто застосовують для збільшення продажів, але їхній вплив може суттєво відрізнятись залежно від товару. Щоб зрозуміти, наскільки ефективними є акції, створили модуль для аналізу їхнього впливу.

Спочатку дані про продажі розділяють на дві групи: періоди, коли діяли акції, і коли їх не було. Для цього використовують стовпець «Рекламна_акція», де вказано, чи була акція активною. Потім для кожного товару підраховують, скільки одиниць продали під час акцій і в звичайні дні. Це допомагає оцінити, наскільки акція вплинула на попит.

Результати аналізу зображають на графіку, який показує, як змінювалися продажі. Наприклад, можна побачити, що під час акції продажі подвоїлися, а без неї залишалися стабільними. Такий аналіз дозволяє визначити, які акції були успішними, а які не дали очікуваного результату. Якщо акція не вплинула на продажі, можливо, для цього товару слід спробувати інший маркетинговий підхід. Це допомагає ефективніше використовувати рекламний бюджет.

2.3.3. ABC-XYZ аналіз продажів

ABC аналіз ділить товари на три групи залежно від того, яку частку прибутку вони приносять. Група А включає найприбутковіші товари, які забезпечують близько 80% доходу компанії, навіть якщо їх кількість порівняно невелика. Група В охоплює товари середньої прибутковості, які приносять близько 15% доходу. Група С складається з товарів із найменшим прибутком – вони формують лише 5% загального доходу, але їхня кількість може бути значною. У нашому модулі це реалізується шляхом розрахунку прибутку для кожного товару на основі його ціни закупівлі, ціни продажу та кількості

проданих одиниць. Після цього товари сортуються за прибутковістю, а накопичувальний дохід використовується для визначення груп А, В і С.

XYZ аналіз, своєю чергою, оцінює стабільність продажів. Товари класифікуються на три категорії: Х, Y і Z залежно від варіабельності попиту. Категорія Х включає товари зі стабільним і передбачуваним попитом, які характеризуються низькою варіабельністю. Товари категорії Y мають помірну варіабельність, їхній попит може змінюватися через сезонні фактори чи акції. Категорія Z охоплює товари з нерегулярним попитом, що робить їх важкими для прогнозування. У нашому модулі стабільність визначається за допомогою коефіцієнта варіації, який розраховується як відношення стандартного відхилення до середнього рівня продажів.

Після проведення окремих аналізів результати ABC і XYZ об'єднуються. Наприклад, товар може отримати категорію А-Х, якщо він приносить значний дохід і має стабільний попит, або категорію С-Z, якщо його дохідність низька, а попит нестабільний. Це дозволяє побачити повну картину кожного товару і визначити, як найкраще з ним працювати.

У нашій системі ABC-XYZ аналіз допомагає вирішити кілька важливих завдань. Наприклад, товари категорії А-Х є найбільш цінними і повинні бути в пріоритеті при плануванні маркетингових кампаній та закупівель. Товари В-Y можуть вимагати додаткових акцій, щоб знизити варіабельність попиту. Товари категорії С-Z, які є найменш вигідними та найбільш нестабільними, можуть бути замінені або продані зі знижками, щоб оптимізувати складські запаси.

Реалізація модуля включає кілька основних етапів. Спочатку завантажуються дані про продажі, включаючи дати, назви товарів, кількість проданих одиниць, ціни закупівлі та продажу. Потім для кожного товару розраховуються прибутковість і стабільність продажів. Отримані результати зберігаються в Excel-файлі, де для кожного товару вказується його категорія (наприклад, А-Х), загальний обсяг продажів, прибуток і коефіцієнт варіації. Це дає змогу швидко і зручно аналізувати товари.

Загалом ABC-XYZ аналіз у нашій системі виконує важливу роль, дозволяючи зрозуміти, які товари є найприбутковішими та найстабільнішими, а які можуть вимагати перегляду. Такий підхід допомагає оптимізувати асортимент і зосередитися на товарах, які мають найбільший вплив на успіх бізнесу.[15]

2.4. Розробка модуля прогнозування продажів

Прогнозування продажів – це ключовий етап нашої системи, який допомагає визначити, скільки одиниць товару потрібно підготувати на майбутній період. Це дозволяє уникнути дефіциту популярних товарів чи зайвих витрат на товари, які можуть не продаватися. Для цього ми використовуємо сучасні методи машинного навчання, які дозволяють аналізувати історичні дані та передбачати майбутній попит.

Спочатку дані проходять ретельну підготовку. Ми завантажуюмо інформацію про продажі: дати, назви товарів, кількість проданих одиниць, ціни закупівлі й продажу, сезонність і рекламні акції. Ці дані доповнюємо новими ознаками, які допомагають моделі краще розуміти закономірності. Наприклад, додаємо день тижня, місяць, квартал і святкові дні, оскільки продажі можуть змінюватися залежно від цих факторів. Також ми створюємо так звані лагові ознаки – це дані про кількість продажів за попередні дні. Вони дають моделі уявлення про динаміку попиту в минулому, що дуже важливо для прогнозування.

Після підготовки даних ми обираємо модель машинного навчання. У нашому випадку це *LightGBM* – потужний і швидкий алгоритм, який добре працює із задачами прогнозування. Ця модель будує послідовність рішень, які дозволяють знайти найкращі відповіді на основі наявних даних. Її особливість у тому, що вона може працювати навіть з великими обсягами інформації, зберігаючи при цьому високу швидкість.

Щоб налаштувати модель максимально ефективно, ми використовуємо інструмент *Optuna*. Це спеціальна бібліотека, яка автоматично знаходить найкращі параметри для моделі, наприклад, кількість дерев рішень або

швидкість навчання. Завдяки цьому модель може робити точніші прогнози, не вимагаючи тривалого ручного налаштування.

Коли модель налаштована, вона використовується для прогнозування продажів на наступний місяць. Для кожного товару модель враховує всі доступні ознаки – історичні дані, сезонність, вплив акцій, зміну цін тощо. У результаті ми отримуємо детальні прогнози, які показують, скільки одиниць кожного товару буде продано щодня протягом наступних 30 днів. Ці прогнози зберігаються в Excel, щоб користувач міг зручно їх переглядати та використовувати для планування.

Для оцінки точності моделі ми порівнюємо прогнозовані значення з реальними даними, використовуючи метрики, такі як середня абсолютна помилка (MAE) і середній квадрат помилки (MSE). У нашому випадку точність прогнозів становить трохи більше 77%. Це означає, що наша модель справляється з передбаченням загальних тенденцій продажів, хоча в окремих випадках може виникати похибка.

Прогнози використовуються для багатьох бізнес-задач. Вони допомагають заздалегідь закупити необхідну кількість товарів, щоб уникнути дефіциту чи надлишків. Вони також дозволяють ефективніше планувати складські запаси, зменшуючи витрати на їхнє зберігання. Крім того, знаючи, які товари будуть популярними, можна краще планувати маркетингові кампанії.

Таким чином, модуль прогнозування продажів є важливим елементом нашої системи. Він використовує сучасні методи машинного навчання, щоб створювати точні прогнози, які допомагають бізнесу приймати обґрунтовані рішення. Хоча точність прогнозів поки що становить близько 77%, це вже дозволяє отримувати корисну інформацію для планування ресурсів і стратегії розвитку.[16-20]

2.5. Розробка інтерфейсу користувача

Інтерфейс користувача — це частина програми, яка забезпечує взаємодію між користувачем і функціональними модулями системи. Основна мета

розробки інтерфейсу полягала в тому, щоб зробити роботу з системою простою та зрозумілою навіть для тих, хто не має технічних знань. Для реалізації інтерфейсу було обрано мову програмування C#, яка добре підходить для створення сучасних десктопних додатків.

2.5.1. Етапи розробки інтерфейсу

Розробка інтерфейсу складалася з кількох етапів. Спочатку визначили, як користувачі будуть працювати з програмою. Інтерфейс розділили на головне меню і три основні панелі: завантаження бази даних, аналіз даних і прогнозування.

Головне меню стало центральною частиною для навігації. У ньому розмістили кнопки для переходу між панелями. Наприклад, кнопка «Завантажити базу даних» відкриває панель для завантаження файлу, кнопка «Аналіз даних» запускає аналіз, а «Прогноз на наступний місяць» — активує функцію прогнозування.

Кожна панель виконує свою роль. Панель завантаження дозволяє вибрати файл Excel, переглянути його вміст і передати дані для обробки. Тут є текстове поле для введення шляху до файлу, кнопка «Обрати» для вибору файлу через провідник і таблиця для відображення даних. Панель аналізу показує результати аналізу у вигляді графіків, наприклад, сезонність продажів, і таблиці з ABC-XYZ аналізом. Візуалізації створюються Python-скриптом. Панель прогнозування дозволяє переглянути прогноз продажів на наступний місяць у таблиці, яку можна також зберегти у форматі Excel.

Інтерфейс зробили інтерактивним, з можливістю зручно переходити між панелями, завантажувати дані та переглядати результати. Усі дії супроводжуються повідомленнями про успіх чи помилки, щоб користувачі завжди розуміли, що відбувається.

На рисунку 2.2 зображено діаграму компонентів системи, яка ілюструє взаємодію між інтерфейсом користувача, функціональними модулями та файлами.

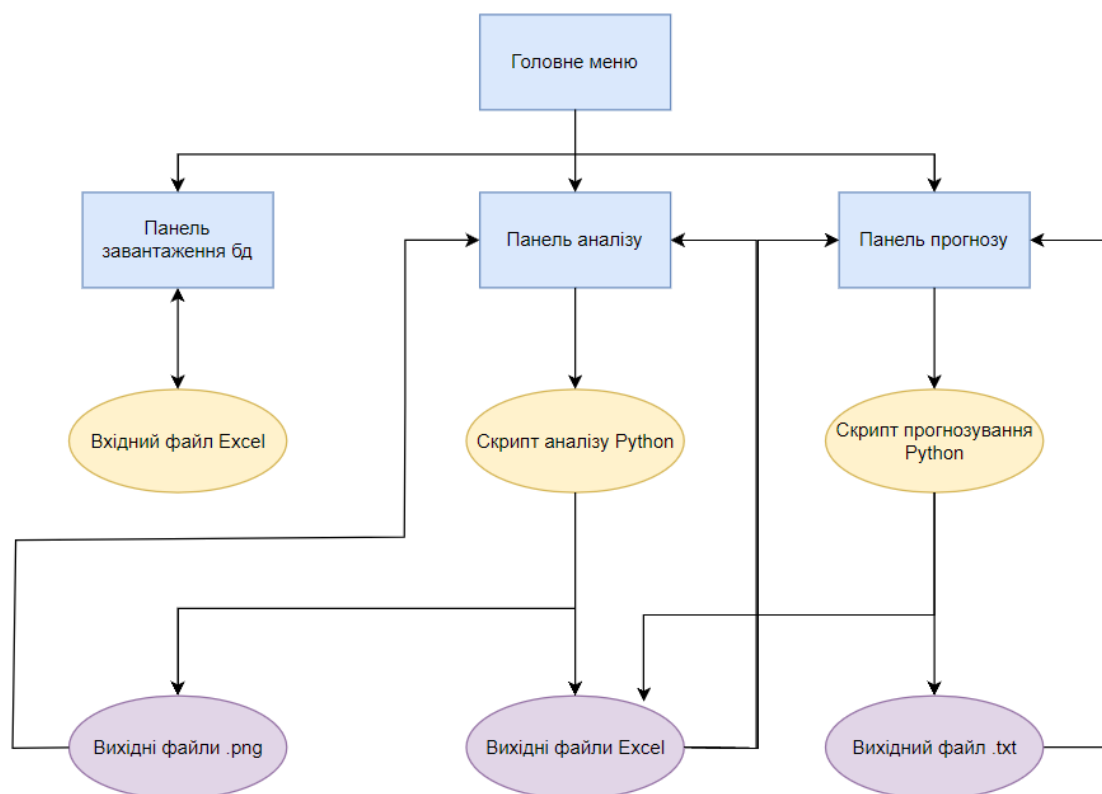


Рисунок 2.2. Діаграма компонентів

«Головне меню» є центральним елементом інтерфейсу програми. Воно дозволяє користувачеві зручно перемикатися між основними розділами: завантаженням бази даних, аналізом даних і прогнозуванням.

«Панель завантаження бази даних» відповідає за отримання вхідних даних. Користувач вибирає файл у форматі Excel, і програма автоматично передає його для обробки Python-скриптам через файлову систему.

«Панель аналізу» працює з Python-скриптом, який аналізує вхідні дані. На основі файлу Excel генеруються результати, які включають графіки, текстові файли з показниками та таблиці. Усі ці матеріали створюються у форматах, зручних для перегляду та подальшого використання.

«Панель прогнозування» запускає Python-скрипт, що створює модель машинного навчання. Цей модуль працює з тим самим Excel-файлом, використовуючи його для побудови прогнозу продажів. Результати прогнозування зберігаються у вигляді Excel-файлів, які можна переглянути безпосередньо через інтерфейс.

«Вхідні та вихідні файли» відіграють важливу роль у роботі програми. Вхідні файли містять початкові дані про продажі, які аналізуються та використовуються для прогнозування. Вихідні файли, створені Python-скриптами, повертаються до інтерфейсу для зручного перегляду результатів користувачами.

2.5.2. Переваги інтерфейсу

Програма вирізняється своєю простотою у використанні. Завдяки інтуїтивно зрозумілому дизайну навіть нові користувачі можуть швидко освоїти основні функції без додаткових зусиль.

Інтерфейс є інтерактивним, що дозволяє користувачам миттєво отримувати результати своїх дій. Наприклад, графіки, таблиці прогнозів та інші результати з'являються одразу після виконання відповідних операцій.

Система інтегрована з Python, що забезпечує виконання складних розрахунків у фоновому режимі. Це дозволяє зберігати простоту взаємодії для користувачів, приховуючи технічну складність процесів.

Інтерфейс програми має високу гнучкість. Його структура легко адаптується до додавання нових функцій, що робить систему готовою до майбутнього розширення або вдосконалення.

2.6. Системні вимоги

Для забезпечення ефективної роботи системи прогнозування продажів у роздрібній торгівлі необхідно відповідати певним мінімальним системним вимогам(див. табл. 2.1). Це гарантує, що система зможе обробляти великі обсяги даних, виконувати навчання моделей машинного навчання та працювати з достатньою продуктивністю. Наведені нижче вимоги є мінімальними, і дотримання їх дозволить забезпечити стабільну роботу системи навіть з великими наборами даних.

Таблиця 2.1

Системні вимоги

Тип	Мінімальні	Рекомендовані
Операційна система	Windows 10	Windows 11
Процесор (CPU)	4-ядерний CPU 2.5 ГГц (Intel i5 або AMD Ryzen 5)	6-8 ядерний CPU (Intel i7/AMD Ryzen 7)
Оперативна пам'ять (RAM)	8 ГБ	16 ГБ і більше
Накопичувач (Storage)	256 ГБ SSD	512 ГБ і більше
Відеокарта (GPU)	Інтегрована графіка (Intel HD)	NVIDIA GeForce GTX 1050+ з підтримкою CUDA
Дисплей	1280x720 (HD)	1920x1080 (Full HD)

Висновки до розділу 2

В даному розділі був розроблений комплекс системного аналізу та прогнозування продажів, що складається з кількох ключових компонентів, кожен із яких спрямований на вирішення певних завдань у сфері роздрібної торгівлі.

Першим було здійснено огляд бази даних, яка містить інформацію про продаж товарів, включаючи дату, обсяги продажів, ціну закупівлі та продажу, сезонність і рекламні акції. Особлива увага приділялася створенню додаткових ознак, таких як день тижня, святковий день, лагові ознаки та тренди. Ці ознаки створені з метою підвищення точності прогнозів і розширення можливостей аналізу.

Методологія розробки базувалася на використанні сучасних мов програмування Python і C#. Python був обраний для виконання аналітичних завдань та завдань прогнозування за допомогою своїх бібліотек Pandas, Matplotlib, LightGBM і Optuna. C# застосовувався для створення зручного інтерфейсу користувача, що забезпечує інтерактивну взаємодію із системою.

Було реалізовано декілька аналітичних модулів, зокрема для аналізу сезонності, впливу рекламних дій та проведення аналізу ABC-XYZ. Ці модулі допомагають оцінити поведінку продажів, розмістити найбільш прибуткові товари й оцінити ефективність маркетингових стратегій.

Модуль прогнозування продажів використовує алгоритм LightGBM, який забезпечує точність прогнозування понад 77%. Оптимізація параметрів роботи моделі проводилася за допомогою бібліотеки Optuna, що дозволило підвищити результати без значного збільшення витрат часу.

Завершальним етапом стала розробка інтерфейсу користувача, яка дозволяє інтегрувати всі функції системи та забезпечити доступ до даних у зручній і зрозумілій формі. Інтерфейс дозволяє завантажувати дані, виконувати аналіз, прогнозувати продажі та отримувати результати у вигляді таблиць і графіків.

Таким чином, у другому розділі була створена система, яка об'єднує методи аналізу та прогнозування для підвищення ефективності управління продажами в роздрібній торгівлі. Вона забезпечує не тільки високу точність прогнозів, але й зручність використання, що є місцем фактором для впровадження в реальних бізнес-процесах.

РОЗДІЛ 3. ОГЛЯД ТА АНАЛІЗ РЕЗУЛЬТАТІВ

3.1. Аналіз моделей прогнозування

У процесі розробки системи прогнозування продажів було проведено тестування різних моделей, щоб знайти оптимальне рішення, яке забезпечить високу точність і швидкість роботи. На першому етапі було вирішено використовувати модель ARIMA. ARIMA є класичною моделлю, яка добре зарекомендувала себе в роботі з часовими рядами, тобто даними, що мають залежність від часу, наприклад, щоденними продажами товарів.

Модель ARIMA базується на трьох параметрах: p , d і q . Ці параметри визначають, як модель враховує минулі значення даних та їхні зміни для прогнозування майбутніх значень.

На початковому етапі ARIMA було налаштовано для роботи з історичними даними про продажі товарів. Проте виявилось, що цей підхід має суттєві обмеження. Перш за все, ARIMA потребувала окремого налаштування для кожного товару. Це означало, що для великої бази даних з кількома десятками чи сотнями товарів потрібно будувати стільки ж окремих моделей. Через це час виконання суттєво зростає, що робило модель малопридатною для практичного використання.

Крім того, результати точності були недостатньо високими. Середня точність прогнозів ARIMA становила близько 60%. Це означало, що майже половина передбачень суттєво відрізнялася від реальних значень. Було проведено низку експериментів із параметрами моделі, такими як p , d , q , і попередньою обробкою даних: видаленням тренду, сезонних компонентів тощо. Проте це дало лише незначне покращення результатів.

Ще одним суттєвим обмеженням ARIMA стало те, що вона ефективно працює лише з часовою залежністю. Вплив інших важливих факторів, таких як акції, ціна чи святкові дні, модель врахувати не могла. Для цього потрібно було б додатково ускладнювати моделювання, що ще сильніше збільшило б час і ресурси для виконання даного процесу.

Розуміючи обмеження ARIMA, було вирішено спробувати сучасні методи машинного навчання. LightGBM було обрано через його гнучкість, високу точність і здатність обробляти великі набори даних із багатьма факторами.

Перед тим, як розпочати навчання моделі LightGBM, був підготовлений новий набір ознак, що враховують додаткові фактори, які можуть впливати на продажі.

Однією з таких ознак є лагові значення, які показують обсяг продажів за попередні дні. Наприклад, дані про продажі за останні 7 днів можуть вказувати на існуючий тренд або зниження попиту.

Також були додані ознаки, пов'язані з датою, такі як день тижня, місяць, квартал та рік. Це дозволило моделі враховувати сезонні коливання та специфічні особливості окремих днів, що можуть впливати на обсяг продажів.

Святкові дні також стали важливим фактором, оскільки вони можуть значно змінювати поведінку споживачів і впливати на попит.

Була введена й ознака, що відображає проведення рекламних акцій, яка показує, чи мала місце акція в конкретний день. Це допомагає моделі оцінювати вплив рекламних кампаній на продажі.

Ці ознаки розширили можливості моделі та дозволили враховувати складні залежності, які були недоступні для ARIMA.

Для налаштування LightGBM було використано автоматизований пошук гіперпараметрів за допомогою бібліотеки Optuna. Це дозволило знайти оптимальні значення таких параметрів, як:

- кількість листків у дереві;
- максимальна глибина дерева;
- швидкість навчання;
- мінімальна кількість даних у вузлі.

Оптимізація допомогла досягти балансу між точністю моделі та швидкістю її навчання.

Результати після застосування LightGBM продемонстрували значне покращення. Точність прогнозів перевищила 77%, а час виконання суттєво

скоротився. На відміну від ARIMA, LightGBM дозволила обробляти всі товари одночасно, що зробило процес більш ефективним.

Модель також надала інформацію про важливість ознак. Наприклад, було встановлено, що найбільший вплив на точність прогнозу мають лагові значення, сезонність і рекламні акції. Це дало змогу краще зрозуміти, які фактори найбільше впливають на продажі.

Таблиця 3.1

Порівняння моделей

Параметр	ARIMA	LightGBM
Точність прогнозів	~60%	>77%
Час виконання	Декілька годин	~13хв
Врахування додаткових факторів	Обмежено	Не обмежено

Перехід від ARIMA до LightGBM показав важливість використання сучасних методів для складних даних. Якщо ARIMA була обмежена часовою залежністю та вимагала тривалого часу для обробки, LightGBM дозволила швидко й ефективно враховувати багатфакторні впливи, такі як акції, сезонність і зміни цін. Це зробило LightGBM більш придатним інструментом для задач у роздрібній торгівлі, де важливі точність прогнозу, швидкість роботи та можливість масштабування.

3.2. Огляд результатів аналізу продажів

Для кожного продукту окремо було створено графіки, що відображають ключові аспекти аналізу даних продажів за останні два роки. У даному випадку представлені результати для Товару А, який демонструє характерний вплив рекламних акцій на кількість проданих одиниць.

На графіку, що ілюструє вплив рекламних акцій (див. рис. 3.1), видно суттєву різницю між обсягами продажів залежно від проведення акцій. Протягом акцій продажі перевищують 700 тисяч одиниць, тоді як без них вони становлять близько 500 тисяч одиниць. Це чітко вказує на важливість рекламних заходів для стимулювання активності купівлі.

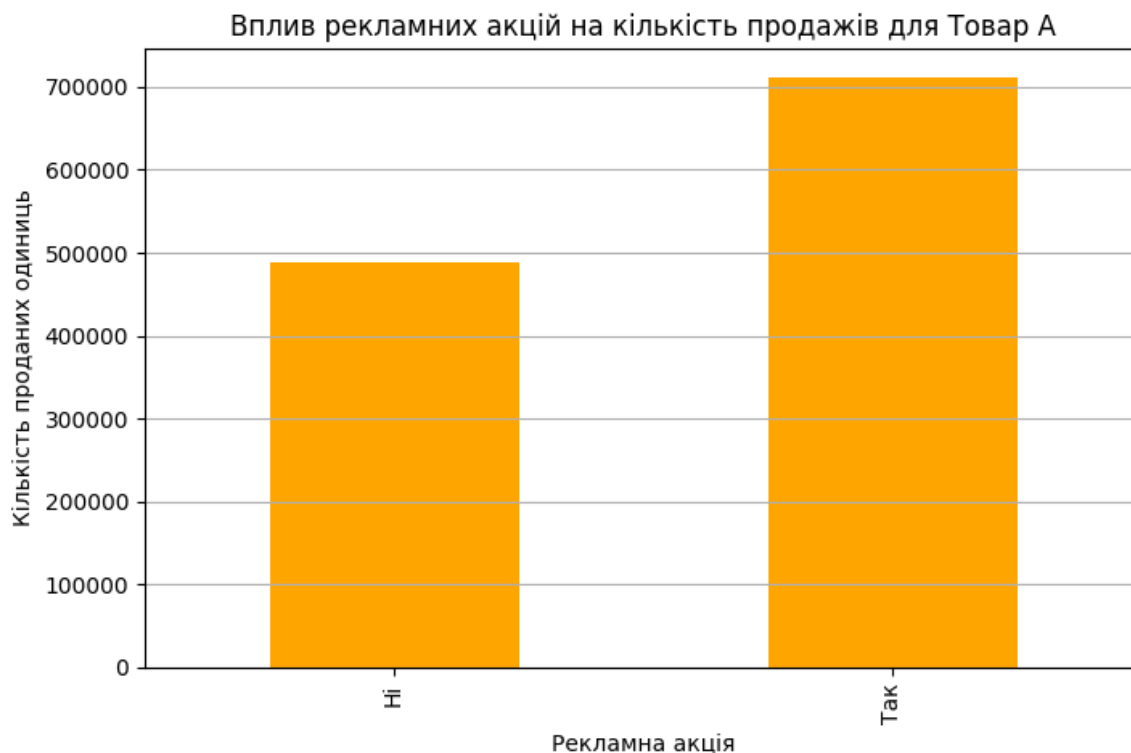


Рисунок 3.1. Вплив рекламних акцій на продажі

Ефективність рекламних акцій залежить від кількох факторів. По-перше, зниження цін на товар часто привертає більше покупців, оскільки вони вважають пропозицію вигідною. По-друге, проведення рекламних акцій сприяє покращенню обізнаності про товар, оскільки додаткові маркетингові заходи дозволяють більшій кількості людей дізнатися про продукт. Нарешті, створення відчуття терміновості у клієнтів — коли акція має обмежений час або кількість товару — стимулює їх швидше здійснити покупку.

Водночас підприємству важливо не зловживати частотою акцій, щоб уникнути зниження загальної прибутковості та формування у клієнтів звички купувати товар лише за зниженою ціною. Оптимальним підходом може бути планування акцій у періоди зниженої активності попиту або в рамках маркетингових стратегій для підвищення лояльності клієнтів.

Аналіз сезонності продажів (див. рис. 3.2) також проводився для кожного товару на основі даних за два останні роки. На представленому графіку для Товару А показано, що продажі суттєво залежать від сезону. Найвищий рівень продажів припадає на літо — понад 350 тисяч одиниць. Водночас найменший

попит спостерігається восени. Весна і зима мають приблизно однакові показники продажів, які коливаються в межах 250 тисяч одиниць.

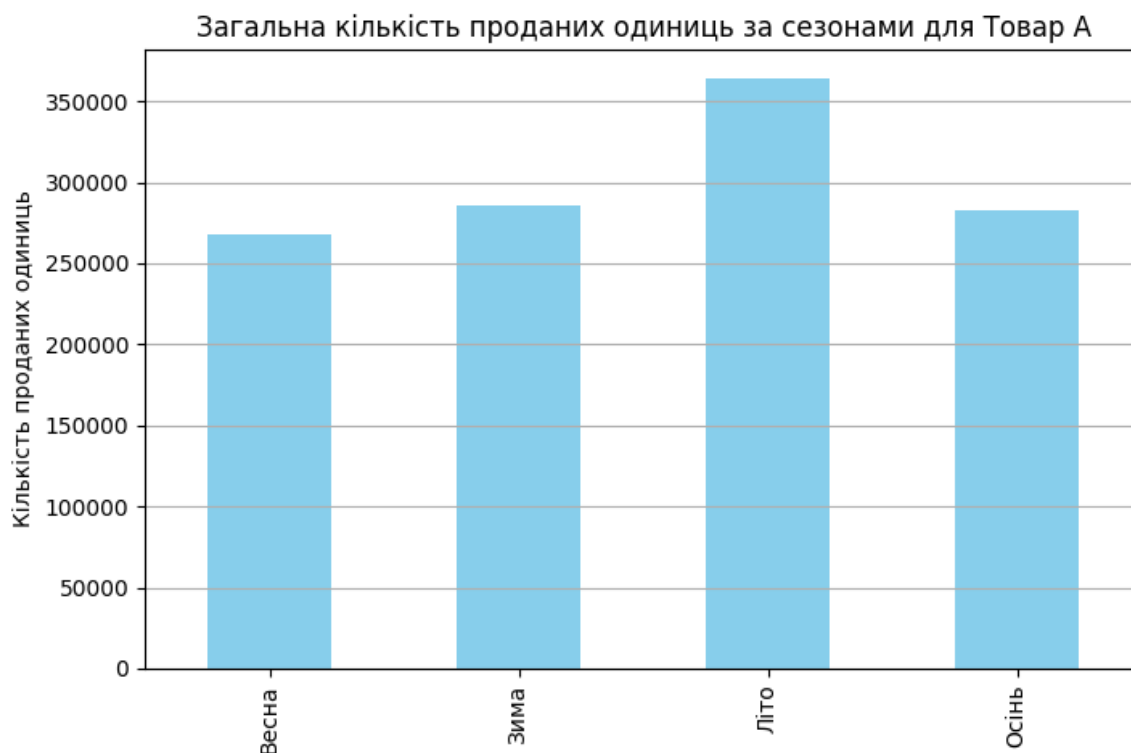


Рисунок 3.2. Кількість продаж за сезонами

Ці результати вказують на те, що товар має специфічну сезонність, яка, ймовірно, пов'язана з особливими потребами клієнтів у літній період.

Наприклад, у літній період може спостерігатися збільшення попиту на певні товари, що залежить від їх призначення, таких як пляжне приладдя чи літній одяг. Крім того, в теплі місяці споживачі часто мають більше вільного часу, що може сприяти зростанню обсягів покупок, оскільки люди частіше займаються шопінгом чи відвідують магазини.

Для оптимізації діяльності підприємства рекомендовано враховувати ці особливості сезонності. Зокрема, перед літнім сезоном необхідно збільшити обсяги виробництва відповідних товарів та планувати маркетингові кампанії, щоб стимулювати продажі. Восени, коли попит на деякі товари може бути нижчим, доцільно організовувати акції чи розпродажі для того, щоб зменшити складські залишки та підвищити обсяги продажу.

Аналіз сезонності є важливим інструментом, що дозволяє адаптувати бізнес-процеси до змін попиту та максимізувати дохід.

Окрім графіків, для кожного товару проводився ABC-XYZ аналіз на основі даних за два останні роки(див. рис. 3.3). Цей аналіз дозволяє оцінити фінансову значимість товарів і стабільність їх попиту.

Товар	Кількість_продано	Прибуток	Накопичувальний_прибуток	Категорія_ABC	Категорія_XYZ	Категорія_ABC_XYZ	Рік
Товар А	587366	18184703,4	1	C	Y	C-Y	2023
Товар Б	626267	33375846,2	0,905774935	B	Y	B-Y	2023
Товар В	592127	51976071,2	0,732836107	A	Y	A-Y	2023
Товар Г	598421	89455587,1	0,463519166	A	Y	A-Y	2023
Товар А	612188	17594419,9	1	C	Y	C-Y	2024
Товар Б	608423	32130043,9	0,912120459	B	Y	B-Y	2024
Товар В	585081	55140453,2	0,751639266	A	Y	A-Y	2024
Товар Г	609914	95345731,7	0,476227076	A	Y	A-Y	2024

Рисунок 3.3 Таблиця ABC-XYZ аналізу

Сукупність ABC і XYZ класифікацій дозволяє визначити найбільш пріоритетні товари, такі як ті, що належать до категорії AX (ключові товари зі стабільним попитом). Вони забезпечують основну частину доходів і мають низький рівень ризику. Наприклад, якщо товар залишається в категорії AX протягом кількох років, це свідчить про його стабільну популярність і важливість для бізнесу. Товари з категорії BX, хоч і менш значущі, але також є важливими через передбачуваність їх попиту.

З іншого боку, товари з категорії BZ чи CZ вимагають окремого підходу. Їх нестабільність може свідчити про зниження популярності, зміну сезонного попиту або залежність від зовнішніх факторів. Такі товари слід уважно аналізувати, щоб знайти можливі шляхи стабілізації їх продажів, наприклад, через рекламні кампанії або зміну позиціонування на ринку.

Загальний висновок із таблиці полягає в тому, що ABC-XYZ аналіз допомагає глибоко зрозуміти особливості кожної товарної позиції та адаптувати управлінські рішення. Товари з високою значимістю і стабільним попитом (категорії AX, BX) слід підтримувати за рахунок інвестицій у їхню популяризацію, оновлення асортименту та розширення ринку збуту. Другорядні товари (групи C) можуть бути використані для акцій, розпродаж або навіть виключені з асортименту, якщо вони не приносять суттєвої вигоди. Завдяки

такому підходу можна оптимізувати процеси виробництва, продажів і маркетингу, що сприятиме зростанню ефективності та прибутковості підприємства.

3.3. Аналіз результатів прогнозування продажів

Дата	Товар	Прогноз_Кількість_продано
2024-12-31 00:00:00	Товар А	1462
2025-01-01 00:00:00	Товар А	1172
2025-01-02 00:00:00	Товар А	964
2025-01-03 00:00:00	Товар А	1130
2025-01-04 00:00:00	Товар А	1135
2025-01-05 00:00:00	Товар А	1082
2025-01-06 00:00:00	Товар А	1183
2025-01-07 00:00:00	Товар А	1330
2025-01-08 00:00:00	Товар А	1156
2025-01-09 00:00:00	Товар А	1107
2025-01-10 00:00:00	Товар А	1116
2025-01-11 00:00:00	Товар А	1059
2025-01-12 00:00:00	Товар А	1172
2025-01-13 00:00:00	Товар А	1224
2025-01-14 00:00:00	Товар А	1162

Рисунок 3.4. Результати прогнозу продажів

Аналіз результатів прогнозування продажів «Товару А» за наступний місяць(див. рис. 3.4) демонструє очікувану динаміку попиту, що дозволяє краще планувати діяльність компанії. Загальний прогнозований обсяг продажів допоможе оцінити майбутній рівень доходу, а також підготувати достатній запас товару для задоволення потреб покупців. Середній рівень щоденних продажів показує на стабільність прогнозу, проте окремі дні можуть показати значні відхилення, які варто врахувати для поточного планування. Наприклад, прогнозовані піки можуть сигналізувати про можливість періоду підвищеного попиту, а спади — про періоди, коли активність покупців буде нижчою.

Важливим аспектом цього аналізу є точність прогнозування, яка становить 77,24%. Це означає, що модель здатна добре точно передбачити майбутній

попит, хоча залишаються ризики певних похибок. Такий рівень точності можна вважати достатнім для формування загальної стратегії продажів, проте уникнути варіацій в реальних умовах.

Додатковий аналіз сезонності може надати глибше розуміння змін у попиті, наприклад, підвищення активності в певні дні або зниження в інших. Це буде корисним для підготовки маркетингових дій та ефективного розподілу ресурсів. Загалом результат прогнозу є цінним інструментом, який дозволяє компанії бути більш гнучкою та готовою до можливих змін у поведінці покупців.

3.4. Огляд інтерфейсу

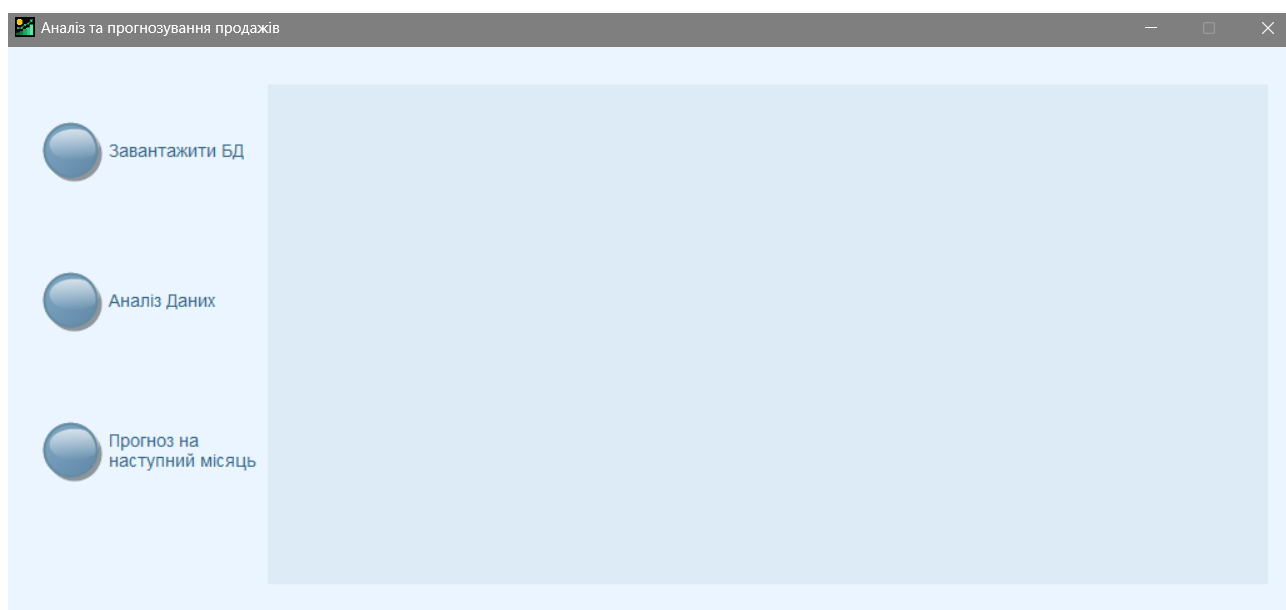


Рисунок 3.5. Головне меню

При запуску програми головне меню виглядає лаконічно та зрозуміло (див. рис. 3.5). Ліворуч розташовані три круглі кнопки: «Завантажити БД», «Аналіз даних» та «Прогноз на наступний місяць». Вони мають зручне розташування для швидкого доступу. Праворуч велика порожня область, яка призначена для відображення даних та результатів роботи.

Після натискання на кнопку «Завантажити БД» відкривається вікно з таблицею (див. рис. 3.6), що заповнюється даними з вибраного excel файлу. У верхній частині буде показано шлях до файлу бази даних. У таблиці є колонки:

дані, товар, кількість продажів, ціна закупівлі, ціна продажу, сезон та рекламна акція. Інтерфейс дозволяє зручно переглядати інформацію, що завантажена.

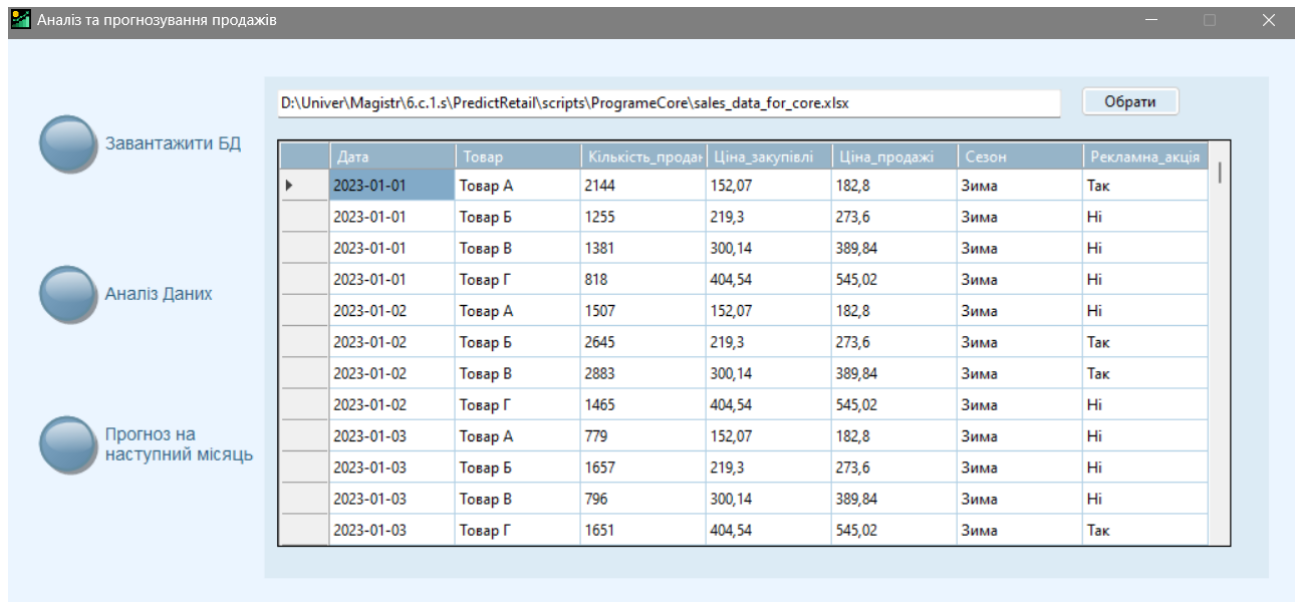


Рисунок 3.6. Панель завантаження БД

Після натискання на кнопку «Аналіз даних» підтягується раніше завантажена БД та запускається скрипт, що виконує аналіз та виводить у верхню частину таблицю з ABC-XYZ аналізом, а після графіки впливу рекламних акцій та сезонності на продажі(див. рис. 3.7).

Графіки виводяться для кожного товару окремо, щоб можна було переглянути інформацію про конкретний товар.

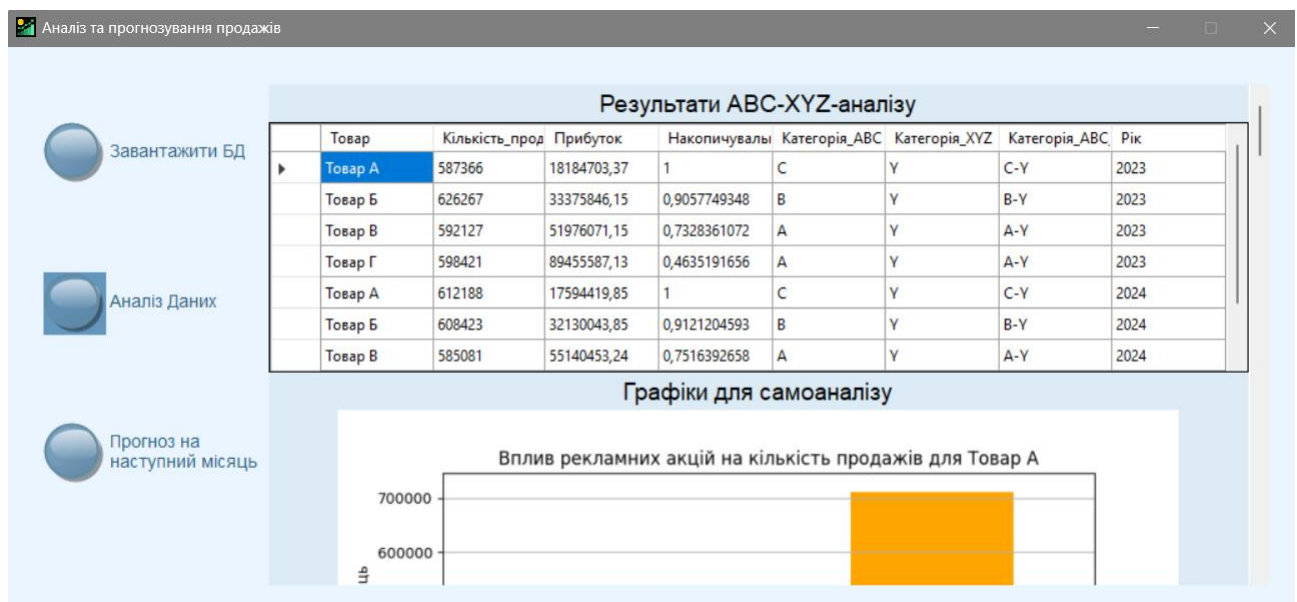
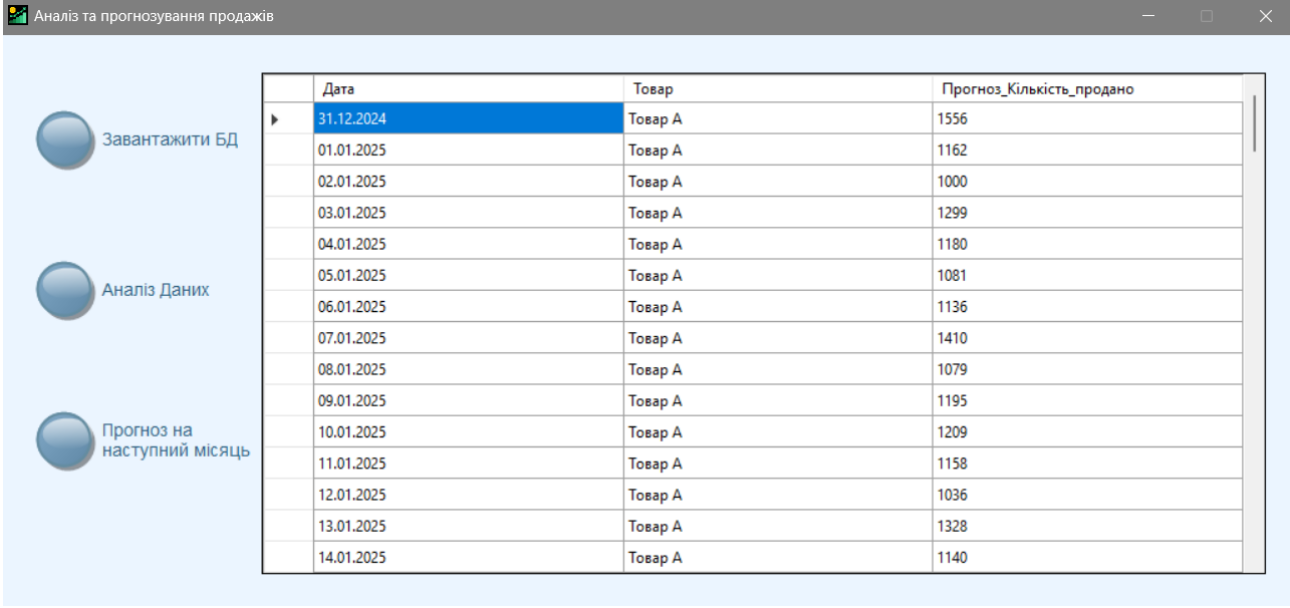


Рисунок 3.7. Панель аналізу даних

При натисканні на кнопку «Прогноз на наступний місяць» запускається скрипт, що навчає відповідні моделі та виконує прогноз для кожного товару, а результат виводить на панель(див. рис. 3.8).



Дата	Товар	Прогноз_Кількість_продано
31.12.2024	Товар А	1556
01.01.2025	Товар А	1162
02.01.2025	Товар А	1000
03.01.2025	Товар А	1299
04.01.2025	Товар А	1180
05.01.2025	Товар А	1081
06.01.2025	Товар А	1136
07.01.2025	Товар А	1410
08.01.2025	Товар А	1079
09.01.2025	Товар А	1195
10.01.2025	Товар А	1209
11.01.2025	Товар А	1158
12.01.2025	Товар А	1036
13.01.2025	Товар А	1328
14.01.2025	Товар А	1140

Рисунок 3.8. Панель прогнозу

Висновки до розділу 3

У третьому розділі роботи проаналізовано результати розробленої системи прогнозування продажів у роздрібній торгівлі. Основна увага приділяється оцінці ефективності використаних моделей, таких як ARIMA та LightGBM, а також функціональних модулів для аналізу сезонності, рекламних акцій і ABC-XYZ аналізу.

Система показала точність прогнозування на рівні 77%, що є достатнім для формування стратегій управління продажами. Алгоритм LightGBM виявився більш ефективним порівняно з ARIMA, особливо при роботі з великими обсягами даних і різноманітністю товарів. Оптимізація параметрів моделі за допомогою бібліотеки Optuna дозволила досягти високої точності без значного збільшення часу обчислень.

Розроблений інтерфейс користувача забезпечує зручну взаємодію із системою, дозволяючи виконувати аналіз даних, прогнозувати продажі та

отримувати результати у вигляді графіків і таблиць. Відображення результатів за допомогою графіків надає користувачам можливість легко аналізувати вплив різних факторів, таких як сезонність чи рекламні акції, на продажі окремих товарів.

Загалом, система є практичним інструментом для покращення управління продажами у роздрібній торгівлі, дозволяючи бізнесу ефективно планувати запаси, адаптувати маркетингові стратегії та швидко реагувати на зміни попиту.

ВИСНОВКИ

У кваліфікаційній роботі розроблено систему для аналізу та прогнозування продажів у роздрібній торгівлі, яка автоматизує ключові процеси і допомагає бізнесу адаптуватися до змін у попиті та ринкових умовах. В ході дослідження проведено детальний огляд теоретичних аспектів аналізу продажів, зокрема методів машинного навчання, таких як ARIMA, LightGBM, та підходів управління продуктовим портфелем (ABC-XYZ аналіз). Це дозволило обрати оптимальні моделі та інструменти для вирішення поставлених завдань.

У процесі роботи створено програмне забезпечення, яке поєднує аналіз історичних даних, прогнозування майбутніх продажів, оцінку впливу цінової політики, сезонності та рекламних акцій. Програма має зручний графічний інтерфейс, розроблений на мові C#, і використовує Python для проведення обчислень та аналізу даних. Точність системи більше 77%.

Результати роботи мають практичне значення для роздрібної торгівлі. Використання розробленої системи дозволяє оптимізувати управління запасами, зменшити ризики дефіциту або надлишку товарів, підвищити рентабельність і задовольнити потреби клієнтів. Впровадження сучасних методів аналізу та прогнозування забезпечує бізнесу конкурентні переваги і сприяє стійкому розвитку в умовах мінливого ринку.

Таким чином, мету роботи досягнуто, завдання виконано, а результати дослідження підтвердили ефективність розробленого підходу.

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Mukherjee M., Cuthbertson R., Howard E. Retailing in Emerging Markets: A policy and strategy perspective. Hardcover Book. 280 p. 2018.
2. Геммонд Р. Мистецтво роздрібної торгівлі. Книга у твердій палітурці. 416 с. 2019.
3. Thomas N. Ingram, Raymond W. LaForge, Ramon A. Avila, Charles H. Schwepker Jr, Michael R. Williams. Sales Management: Analysis and Decision Making. 10th Edition. Paperback Book. 358 p. October 31, 2019.
4. John C. Brocklebank, David A. Dickey. SAS for Forecasting Time Series (second edition). Paperback Book. 424 p. 2006.
5. IBM Corporation. IBM Planning Analytics: Overview [Electronic resource]. Доступно: <https://www.ibm.com/products/planning-analytics>.
6. Oracle Corporation. Oracle Retail Predictive Application Server: Overview [Electronic resource]. Доступно: https://docs.oracle.com/cd/E12478_01/rpas/pdf/1411/html/sg/overview.htm.
7. Dinesh C.S. Bisht, Mangey Ram. Recent Advances in Time Series Forecasting. Hardcover Book. 228 p. 2021.
8. Wes McKinney. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython 2nd Edition. Paperback Book. 524 p. 2017.
9. Gerardus Blokdyk. Bass Diffusion Model: A Complete Guide - 2020 Edition. Paperback Book. 305 p. March 28, 2021.
10. Christopher R. Sprague. Exponential Smoothing: An Extension (Classic Reprint). Hardcover Book. 37 p. August 24, 2018.
11. Rohit Lala. Learn ARIMA Model Using Eviews: Auto Regressive Integrated Moving Average Model. Paperback Book. 55 p. November 14, 2022.
12. Gerardus Blokdyk. ABC analysis Standard Requirements. Paperback Book. 305 p. January 4, 2022.
13. Erwin Opalla. How-To-Manual: Inventory Applications. 3rd Edition. Paperback Book. 493 p. May 1, 2019.

14. Joseph Albahari. C# 10 in a Nutshell: The Definitive Reference. Paperback Book. 1058 p. 2022.
15. Sunil Chopra, Peter Meindl. Supply Chain Management: Strategy, Planning, and Operation 7th Edition. Paperback Book. 540 p. April 11, 2019.
16. Hang Li, Lu Lin, Huanqiang Zeng. Machine Learning Methods 1st ed. Paperback Book. 548 p. 2023.
17. Hugh Bishop, Chris Bishop. Deep Learning: Foundations and Concepts. Paperback Book. 670 p. 2024.
18. Moritz Hardt, Benjamin Recht. Patterns, Predictions, and Actions: Foundations of Machine Learning. Hardcover Book. 320 p. 2022.
19. Joel Grus. Data Science from Scratch: First Principles with Python 2nd Edition. Paperback Book. 403 p. June 11, 2019.
20. Aston Zhang, Zachary C. Lipton, Mu Li, Alexander J. Smola. Dive into Deep Learning 1st Edition. Paperback Book. 574 p. December 7, 2023.

ДОДАТКИ

Код функції для оптимізації LightGBM з Optuna

```
def optimize_lightgbm(X_train, y_train):
    def objective(trial):
        params = {
            'num_leaves': trial.suggest_int('num_leaves', 20, 300),
            'max_depth': trial.suggest_int('max_depth', 4, 20),
            'learning_rate': trial.suggest_loguniform('learning_rate', 0.01, 0.3),
            'n_estimators': trial.suggest_int('n_estimators', 100, 1000),
            'min_child_samples': trial.suggest_int('min_child_samples', 5, 100),
            'subsample': trial.suggest_uniform('subsample', 0.5, 1.0),
            'colsample_bytree': trial.suggest_uniform('colsample_bytree', 0.5, 1.0),
            'reg_alpha': trial.suggest_loguniform('reg_alpha', 1e-5, 10),
            'reg_lambda': trial.suggest_loguniform('reg_lambda', 1e-5, 10),
            'min_split_gain': trial.suggest_loguniform('min_split_gain', 1e-5, 1),
            'min_child_weight': trial.suggest_loguniform('min_child_weight', 1e-5, 10),
            'feature_fraction': trial.suggest_uniform('feature_fraction', 0.5, 1.0)
        }

        model = lgb.LGBMRegressor(**params, random_state=42)
        scores = cross_val_score(model, X_train, y_train, cv=3,
            scoring='neg_mean_squared_error')
        return -scores.mean()

    study = optuna.create_study(direction='minimize')
    study.optimize(objective, n_trials=300)

    best_params = study.best_params
    best_model = lgb.LGBMRegressor(**best_params, random_state=42)
    best_model.fit(X_train, y_train)
    return best_model
```

Код функції для навчання і прогнозів

```
def train_and_predict(data):
    data = add_features(data)
    products = data['Товар'].unique()
    results = []
    accuracies = []

    for product in products:
        product_data = data[data['Товар'] == product]
        features = [col for col in product_data.columns if col not in ['Дата', 'Товар',
'Кількість_продано']]
        X = product_data[features]
        y = product_data['Кількість_продано']

        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
        best_model = optimize_lightgbm(X_train, y_train)

        y_pred = best_model.predict(X_test)
        y_pred = np.round(y_pred)
        accuracy = 100 - mean_absolute_error(y_test, y_pred) / y_test.mean() * 100
        accuracies.append(accuracy)

    results.append({
        'Product': product,
        'Test Score': best_model.score(X_test, y_test),
    })

    last_date = product_data['Дата'].max()
    future_dates = [last_date + timedelta(days=i) for i in range(1, 31)]
    future_data = pd.DataFrame({'Дата': future_dates})
    future_data['Товар'] = product
    future_data = add_features(pd.concat([product_data, future_data],
ignore_index=True)).iloc[-30:]
    future_X = future_data[features]

    next_month_pred = best_model.predict(future_X)
    next_month_pred = np.round(next_month_pred)

    pred_df = pd.DataFrame({
        'Дата': future_dates,
        'Товар': product,
        'Прогноз_Кількість_продано': next_month_pred
```

```

    })
    pred_df.to_excel(os.path.join(results_directory,
f'{product}_next_month_predictions.xlsx'), index=False)

    average_accuracy = sum(accuracies) / len(accuracies)
    with
open("D:/Univer/Magistr/6.c.1.s/PredictRetail/scripts/ProgrameCore/predictions_resu
lts/accuracy_result.txt", "w") as f:
        f.write(f"{average_accuracy:.2f}")
    return pd.DataFrame(results)

```

Код, що здійснює ABC-XYZ аналіз

```

annual_results = []

for year, year_data in df.groupby('Рік'):
    # Річний обсяг продажу та прибуток для кожного товару
    annual_data = year_data.groupby('Товар').agg(
        Кількість_продано=('Кількість_продано', 'sum'),
        Прибуток=('Прибуток', 'sum')
    ).reset_index()

    # ABC-аналіз за прибутком
    annual_data = annual_data.sort_values('Прибуток', ascending=False)
    annual_data['Накопичувальний_прибуток'] =
annual_data['Прибуток'].cumsum() / annual_data['Прибуток'].sum()

    # Категоризація за ABC (А - 80%, В - 15%, С - 5%)
    def abc_category(row):
        if row['Накопичувальний_прибуток'] <= 0.8:
            return 'A'
        elif row['Накопичувальний_прибуток'] <= 0.95:
            return 'B'
        else:
            return 'C'

    annual_data['Категорія_ABC'] = annual_data.apply(abc_category, axis=1)

    # XYZ-аналіз за коефіцієнтом варіації продажів
    monthly_sales = year_data.groupby(['Товар',
year_data['Дата'].dt.to_period('M')])['Кількість_продано'].sum().unstack().fillna(0)
    monthly_variability = monthly_sales.std(axis=1) / monthly_sales.mean(axis=1)
    monthly_variability = monthly_variability.fillna(0) # Замінюємо NaN на 0 для
уникнення проблем

    # Категоризація за XYZ (Х - низька варіабельність, Y - середня, Z - висока)
    def xyz_category(value):
        if value < 0.1:
            return 'X'
        elif value < 0.25:
            return 'Y'
        else:
            return 'Z'

    # Додавання категорії XYZ для кожного товару

```

```
    annual_data['Категорія_XYZ'] = annual_data['Товар'].map(lambda x:
xyz_category(monthly_variability.get(x, 0)))
    annual_data['Категорія_ABC_XYZ'] = annual_data['Категорія_ABC'] + '-' +
annual_data['Категорія_XYZ']
    annual_data['Рік'] = year

# Додавання результатів за рік до списку
annual_results.append(annual_data)
```