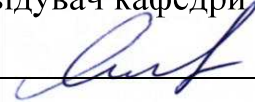


МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний університет кораблебудування імені адмірала Макарова
Навчально-науковий інститут комп'ютерних наук та управління проектами
Кафедра інформаційних управляючих систем та технологій

"Допущений до захисту"

Завідувач кафедри ІУСТ



к.т.н, доц. Михелєв І.Л.

" ____ " _____ 2024 р.

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття ступеня вищої освіти "магістр"

на тему:

Прогнозування попиту на товари в ритейлі з використанням
машинного навчання

Виконав: студент групи 6145 м

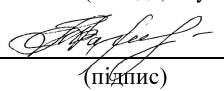


Тесленко П.О.
(підпис)

Керівник роботи:

професор НУК, к.т.н., доц.

(посада, науковий ступінь вчене звання)



Фаріонова Т.А.
(підпис)

м. Миколаїв – 2024 р.

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

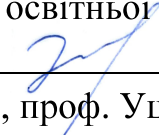
Національний університет кораблебудування імені адмірала Макарова
Навчально-науковий інститут комп'ютерних наук та управління проектами
Кафедра інформаційних управляючих систем та технологій

Спеціальність 124 "Системний аналіз"

Освітня програма "Системний аналіз"

"ЗАТВЕРДЖУЮ"

Гарант освітньої програми


д.ф-м.н, проф. Ушкац М.В.

"___" "___" 2024 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ
на здобуття ступеня вищої освіти "магістр"

Студенту Тесленку Павлу Олександровичу

(прізвище, ім'я, по батькові)

1. Тема роботи: Прогнозування попиту на товари в ритейлі з використанням машинного навчання

Керівник роботи Фаріонова Тетяна Анатоліївна, к.т.н., професор НУК

Затверджені наказом ректора № 1055-уч. від "08" жовтня 2024 року.

2. Термін подання роботи: 15 грудня 2024 р.

3. Вихідні дані до проекту (роботи) ДСТУ щодо обробки інформації, літературні джерела, технічна документація на існуючі аналоги систем, матеріали наукового стажування

4. Перелік питань, що належать до розробки (найменування розділів). Аналіз засобів прогнозування в ритейлі. Проектування інформаційної системи. Розробка моделі прогнозування та перевірка її ефективності. Розрахунок економічної ефективності впровадження інформаційної системи прогнозування. Охорона праці. Охорона навколишнього середовища

5. Перелік презентаційних матеріалів Комп'ютера презентація

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1	Фаріонова Т.А., професор НУК	05.09.2023	22.12.2023
2	Фаріонова Т.А., професор НУК	23.12.2023	15.06.2024
3	Фаріонова Т.А., професор НУК	15.06.2024	25.09.2024
4	Фаріонова Т.А., професор НУК	25.09.2024	14.11.2024
5	Фаріонова Т.А., професор НУК	25.09.2024	14.11.2024
6	Фаріонова Т.А., професор НУК	25.09.2024	14.11.2024

7. Дата видачі завдання 05 вересня 2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту (роботи)	Строк виконання етапів роботи	Примітка
1	Дослідження та аналіз предметної галузі	22.12.2024	
2	Розробка концепції	15.03.2024	
3	Розділ з розробки проектних рішень по системі (та її частинах)	15.06.2024	
4	Реалізація проекту та розробка робочої документації	25.09.2024	
5	Обчислення економічної ефективності	14.11.2024	
6	Вирішення питань охорони праці та навколишнього середовища	14.11.2024	
7	Оформлення пояснювальної записки	14.11.2024	
8	Подання магістерської роботи на перевірку на плагіат	16.11.2024	
9	Подання магістерської роботи рецензенту	20.11.2024	
10	Подання магістерської роботи на державний захист	15.12.2024	

Студент


(підпис)

Тесленко П.О

(ПІБ)

Керівник роботи


(підпис)

Фаріонова Т.А.

(ПІБ)

АНОТАЦІЯ

до магістерської кваліфікаційної роботи на тему: «Прогнозування попиту на товари в ритейлі з використанням машинного навчання»

В кваліфікаційній роботі магістра досліджено проблему прогнозування попиту на товари в ритейлі. В результаті дослідження було з'ясовано, що ці дані мають різні формати, частоту, пропуски, а також значну амплітуду, що ускладнює їх інтеграцію та подальший аналіз. Для вирішення задачі прогнозування застосовано інтелектуальний аналіз даних із використанням методів машинного навчання, а саме алгоритм градієнтний бустингу, XGBoost. Сформовано модель прогнозування та проведено її моделювання на тестовій вибірці даних. На основі цього розроблено інформаційну систему прогнозування.

Робота складається з 99 сторінок, 10 рисунків, 7 таблиць, додатку. Перелік використаних джерел має 26 найменувань.

Ключові слова: прогнозування товарів ритейлу, машинне навчання, інформаційна система, модель прогнозування.

ABSTRACT

to the master's qualification work on the topic: "Forecasting demand for goods in retail using machine learning"

The master's qualification work investigated the problem of forecasting demand for goods in retail. As a result of the study, it was found that this data has different formats, frequency, gaps, as well as a significant amplitude, which complicates their integration and further analysis. To solve the forecasting problem, intelligent data analysis was applied using machine learning methods, namely the gradient boosting algorithm, XGBoost. A forecasting model was formed and its modeling was carried out on a test data sample. Based on this, a forecasting information system was developed.

The work consists of 99 pages, 10 figures, 7 tables, an appendix. The list of sources used has 26 names.

Keywords: forecasting of retail goods, machine learning, information system, forecasting model.

ЗМІСТ

ВСТУП.....	8
РОЗДІЛ 1. АНАЛІЗ ЗАСОБІВ ПРОГНОЗУВАННЯ В РИТЕЙЛІ	11
1.1. Аналіз предметної області.....	11
1.2. Аналіз даних та процедури прогнозування в ритейлі	14
1.3. Аналіз існуючих рішень методів аналізу та моделювання прогнозів на основі алгоритмів машинного навчання	19
1.4 Постановка задачі.....	30
Висновки до розділу 1	32
2 ПРОЄКТУВАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ	34
2.1. Опис параметрів системи та ключових метрик	34
2.2 Розробка інформаційної системи.....	35
2.2.1 Визначення функціональних та нефункціональних вимог	35
2.2.2 Моделювання прецедентів.....	36
2.2.3 Вибір інструментів для розробки	38
2.2.4 Уявлення архітектури ІС.....	39
2.2.5 Інтерфейс системи	42
2.3 Аналіз ефективності ІС	45
2.3.1 Початкові умови та методологія аналізу	45
2.3.2 Результати порівняння	48
2.3.3 Порівняльний аналіз	50
Висновки до розділу 2	51

3 РОЗРОБКА МОДЕЛІ ПРОГНОЗУВАННЯ ТА ПЕРЕВІРКА ЇЇ ЕФЕКТИВНОСТІ	53
3.1 Технічна реалізація моделі	53
3.1.1 Обробка даних для прогнозування.....	53
3.1.2 Навчання моделі.....	56
3.2 Етап прогнозування	60
3.3 Аналіз результатів	65
Висновки до розділу 3	67
РОЗДІЛ 4. РОЗРАХУНОК ЕКОНОМІЧНОЇ ЕФЕКТИВНОСТІ ВПРОВАДЖЕННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ПРОГНОЗУВАННЯ	69
Висновки до розділу 4	72
5 ОХОРОНА ПРАЦІ	73
5.1 Аналіз умов праці на робочому місці програміста.....	73
5.2 Розробка заходів з охорони праці	78
Висновки до розділу 5	82
6 ОХОРОНА НАВКОЛИШНЬОГО СЕРЕДОВИЩА	83
Висновки до розділу 6	87
ВИСНОВКИ.....	88
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ	91
ДОДАТОК А. ФРАГМЕНТ КОДУ ПРОГРАМИ.....	95

ВСТУП

Ритейл – це процес продажу товарів або послуг кінцевим споживачам для особистого чи домашнього використання. У межах цього сектора функціонують як малі магазини, так і великі корпорації, які поєднують гуртову та роздрібну торгівлю. Основною метою ритейлу є задоволення попиту споживачів через ефективне управління товарними запасами, оптимізацію логістики та забезпечення конкурентних цін.

Бізнес-процеси крупної гуртово-роздрібної корпорації охоплюють складні ланцюги постачання, що включають управління запасами на складах, планування доставок, взаємодію з постачальниками, розробку маркетингових кампаній та обслуговування клієнтів у роздрібних магазинах. Злагоджена робота таких процесів потребує постійного аналізу ринкових трендів, врахування змін у попиті та адаптації до сезонних коливань чи зовнішніх факторів. Одним із ключових завдань для таких корпорацій є підтримка достатнього рівня запасів товарів, щоб уникнути дефіциту чи надлишків, що може призвести до втрат або зниження задоволеності клієнтів.

Актуальною проблемою для крупних ритейлерів є забезпечення постійної наявності потрібної кількості товарів на складах і в магазинах. Це завдання вимагає ретельного управління всіма етапами ланцюга постачання та передбачення майбутнього попиту. Необхідність прогнозування попиту стає ще важливішою для товарів із високою конкуренцією або чутливістю до сезонності чи маркетингових активностей. Без точного прогнозу корпорація ризикує втратити потенційних клієнтів через відсутність необхідного товару або збільшити витрати через надмірні запаси, які важко реалізувати.

Прогнозування попиту потребує значної кількості вхідних даних. Ці дані охоплюють інформацію про історичні продажі, сезонні тренди, акції, запаси на складах, характеристики споживчої поведінки та зовнішні фактори, такі як погодні умови або макроекономічні показники. Чим більше даних використовується, тим

вища потенційна точність прогнозу, однак зі збільшенням обсягу даних зростає складність їх обробки та інтеграції. Різноманітність даних, їх пропуски або відмінності в частоті оновлення можуть значно ускладнити побудову точного прогнозу, особливо при використанні традиційних статистичних методів.

Для вирішення цієї проблеми запропоновано застосування методів машинного навчання, що є однією з ключових галузей штучного інтелекту. Машинне навчання дозволяє ефективно працювати з великою кількістю даних, враховувати нелінійні залежності та взаємодії між численними змінними. Алгоритми машинного навчання, такі як градієнтний бустинг, здатні адаптуватися до складних структур даних і забезпечувати високу точність прогнозів навіть у багатовимірному середовищі. Завдяки використанню машинного навчання корпорація може отримати точні прогнози попиту для кожної категорії товарів, оптимізувати управління запасами та підвищити ефективність усіх бізнес-процесів, пов'язаних із плануванням постачань і продажів.

Метою кваліфікаційної роботи магістра є розробка та дослідження моделі прогнозування попиту на товари ритейлу з використанням інструментів машинного навчання. Для досягнення цієї мети було сформовано такі задачі:

1. Провести аналіз предметної області ритейлу, визначити ключові аспекти прогнозування попиту, та виявити основні фактори, що впливають на попит.
2. Сформулювати та структурувати вхідні дані, необхідні для побудови моделі прогнозування, включаючи внутрішні (історичні продажі, запаси, акції) та зовнішні фактори (погодні умови, тренди, логістичні дані).
3. Розробити концепцію інформаційної системи прогнозування попиту, яка забезпечує інтеграцію даних із різних джерел, їх обробку, аналіз і візуалізацію.
4. Визначити функціональні та нефункціональні вимоги до інформаційної системи, включаючи швидкість обробки даних, масштабованість, точність прогнозів та інтеграцію з ERP-системами.

5. Розробити архітектуру інформаційної системи прогнозування, що охоплює модулі збору, обробки, аналізу даних, побудови моделей, інтеграції та виведення результатів.

6. Реалізувати модель прогнозування попиту з використанням інструментів машинного навчання, таких як XGBoost або LightGBM, з урахуванням особливостей ритейлу та структури даних.

7. Виконати підготовку даних для моделі, включаючи очищення, нормалізацію, обробку пропусків і трансформацію у формат, придатний для навчання моделі.

8. Провести навчання моделі прогнозування, включаючи налаштування параметрів, оптимізацію гіперпараметрів і валідацію результатів.

9. Дослідити ефективність розробленої моделі прогнозування на тестових даних, оцінюючи точність за ключовими метриками, такими як середня абсолютна похибка (MAE), середньоквадратична похибка (RMSE) та коефіцієнт детермінації (R^2).

10. Проаналізувати результати прогнозування, оцінити переваги та обмеження моделі, а також надати рекомендації щодо її практичного впровадження в бізнес-процеси ритейлу для покращення управління товарними запасами.

Об'єктом дослідження є процеси прогнозування попиту на товари ритейлу.

Предметом дослідження є модель прогнозування попиту на товари ритейлу з використанням машинного навчання.

Результати магістерського дослідження апробовані на V Всеукраїнський НПК «Інформаційні технології: моделі, алгоритми, системи – 2024», за якими опубліковані тези доповіді [20] та опублікована наукова стаття [1].

РОЗДІЛ 1. АНАЛІЗ ЗАСОБІВ ПРОГНОЗУВАННЯ В РИТЕЙЛІ

1.1. Аналіз предметної області

Аналіз предметної області прогнозування попиту на товари в ритейлі показав численні проблеми, які мають важливе значення для ефективного управління товарами та бізнес-процесами. Неефективне управління запасами, що ґрунтується на традиційних методах прогнозування попиту, часто призводить до надлишків або дефіциту товарів, впливаючи на прибутковість бізнесу. Наприклад у випадку медичних препаратів така ситуація може викликати дефіцит життєво важливих ліків або їх перевищення, що призводить до втрат через закінчення терміну придатності.

Іншій випадок, це прогнозування попиту на електроенергію. Тут традиційні методи можуть ігнорувати або не враховувати важливі змінні, такі як вплив погодних умов, соціальних подій, змін у звичках споживачів або режимі роботи населення. Це може призводити до перевантаження енергосистеми або до надмірного виробництва, яке створює додаткові витрати на зберігання енергії або екологічний тиск через використання резервних генераторів.

У сфері товарів для здоров'я та фітнесу, попит швидко змінюється під впливом нових трендів і діяльності інфлюєнсерів. Традиційні моделі часто не можуть оперативно реагувати на такі зміни, що призводить до перевищення запасів чи їх нестачі. Наприклад, новий тренд на певний спортивний аксесуар може миттєво підвищити попит, що не завжди вчасно враховується. Подібні виклики актуальні і для сезонних товарів, таких як зимовий одяг, новорічні прикраси чи товари для літнього відпочинку. Швидкі коливання попиту у відповідь на погодні умови чи святкові події вимагають спеціалізованих методів прогнозування, яких традиційні підходи часто не забезпечують.

Отже, ритейл, або роздрібна торгівля, — це сфера бізнесу, що займається

продажом товарів та послуг кінцевим споживачам. Це включає магазини, супермаркети, аптеки, інтернет-магазини та інші канали продажу. Попит у ритейлі означає обсяг товарів, які споживачі бажають придбати у певний час та за певну ціну. Наприклад, у період різдвяних свят попит на подарункові набори значно зростає, в той час як влітку популярними стають товари для пляжного відпочинку.

Система ритейлу розглядається як сукупність процесів і операцій, що забезпечують ефективне управління продажами товарів або послуг кінцевим споживачам [1]. Вона уключає в себе взаємодію з постачальниками, управління товарними запасами, логістику та обслуговування клієнтів [2]. Використання терміну «ритейл» замість «роздрібна торгівля» обумовлений розширенням меж цього поняття в сучасних умовах бізнесу. Ритейл охоплює не лише класичну роздрібну торгівлю, а й більш широкий спектр процесів і технологій, які сприяють взаємодії між бізнесом і кінцевими споживачами. Роздрібна торгівля, як традиційний термін, фокусується на фізичному продажу товарів у магазинах або на ринках [3]. Натомість, термін "ритейл" охоплює процеси управління запасами, маркетингові кампанії, логістичні рішення та аналітику продажів.

Таким чином, термін «ритейл» відображає комплексний підхід до взаємодії із зовнішнім середовищем, що виходить за межі традиційної роздрібної торгівлі. Тому використання терміну «ритейл» є доцільним, оскільки він більш адекватно описує сучасні бізнес-процеси, який включає в себе не лише роздрібну торгівлю, але й інші аспекти омніканальних стратегій, важливих для управління бізнесом в умовах цифрової економіки.

В ритейлі попит визначається як кількість товарів або послуг, яку споживачі готові придбати за певну ціну і в конкретний час.

Прогнозування попиту в ритейлі — це процес передбачення майбутніх обсягів продажів товарів на основі історичних даних, трендів, впливу маркетингових кампаній та зовнішніх факторів. Наприклад, використання алгоритмів машинного навчання може дозволити передбачати зміну попиту на

спортивний інвентар у період перед великими спортивними подіями.

Актуальність дослідження обґрунтовано бізнесовою ситуацією, що склалася у великій корпорації, яка має значну кількість складських приміщень різного типу та призначення [4]. Для продажу, корпорація акумулює велику кількість товарних запасів. Виникає необхідність в прогнозуванні таких показників: майбутньої бізнесової активності; кількості товарів для продажу через гуртово-роздрібну мережу корпорації з урахування прогнозу продажів цих товарів; сезонності; маркетингових акцій та зовнішніх чинників, наприклад, погодні умови, тренди та популярність.

Аналіз наукових досліджень демонструє різноманітність підходів до прогнозування попиту. Одне з досліджень [5] фокусується на використанні простих статистичних моделей, таких як методи ковзного середнього та експоненціального згладжування, для прогнозування попиту в аптечному ритейлі. В [6] пропонується підхід до збагачення моделей макроекономічними даними, такими як рівень безробіття чи споживчі настрої, що дозволяє поліпшити точність прогнозів.

Дослідження [7] розглядає автоматизований вибір моделей для різних типів попиту, базуючись на класифікаційних методах.

В [8] автори адаптували нейронні мережі для прогнозування залежностей в часових рядах, що підвищує їхню точність та релевантність для короткострокового прогнозування.

Порівняння підходів показує, що традиційні статистичні методи, такі як ковзне середнє, обмежені у своїй здатності враховувати складну багатофакторну природу попиту. Машинне навчання та інструменти штучного інтелекту забезпечують нові можливості для прогнозування завдяки аналізу великих даних і врахуванню багатьох змінних факторів одночасно. Для вирішення завдань прогнозування необхідно впроваджувати інструменти інтелектуального аналізу даних та сучасні алгоритми, що дозволить зменшити похибки та поліпшити

управління запасами в умовах динамічної ринкової ситуації.

1.2. Аналіз даних та процедури прогнозування в ритейлі

Процедура прогнозування товарів у ритейлі для великої корпорації, яка має значну кількість складських приміщень і оперує гуртово-роздрібною мережею, повинна бути комплексною, багатфакторною та орієнтованою на середньостроковий горизонт планування (3-6 місяців). Мета такої процедури полягає у точному визначенні майбутньої потреби в товарних запасах, щоб уникнути дефіциту або надлишку товарів, оптимізувати витрати на зберігання та забезпечити постійний рівень задоволення попиту. Враховуючи складність завдання, необхідно побудувати систему, яка інтегрує історичні дані, актуальну інформацію та прогнози зовнішніх факторів.

Основу процедури становить збір даних із різних джерел. Історичні дані продажів дозволяють виявити довгострокові тренди, сезонність та поведінкові моделі покупців. Поточна інформація, яка надходить із POS- та ERP-систем, забезпечує актуальність прогнозу. Крім того, дані про залишки товарів на складах та швидкість їхнього поповнення є критичними для розрахунку необхідного обсягу закупівель. Однак, прогнозування в ритейлі буде обмеженим без врахування зовнішніх факторів, таких як погодні умови, тренди споживчої поведінки, популярність товарів у соціальних мережах, економічна ситуація та конкурентна активність. Усі ці дані взаємопов'язані, і зміна одного з факторів може суттєво вплинути на результат прогнозування.

Процедура повинна враховувати комплексність факторів та їхню взаємозалежність [9]. Наприклад, маркетингові акції можуть вплинути на попит як окремо, так і в поєднанні з сезонними трендами. Аналогічно, прогнози погоди допомагають оцінити потенційний попит на певні категорії товарів, але ця інформація стає більш точною у комбінації з даними про минулі продажі за аналогічних умов.

Процес побудови прогнозу починається з попередньої обробки даних: очищення від аномалій, нормалізації та виявлення ключових закономірностей. Потім дані сегментуються за категоріями товарів, пріоритетністю та рівнем попиту. На цьому етапі виділяються найбільш критичні товарні групи, які потребують особливо точного прогнозування. Для побудови самого прогнозу використовуються комбіновані моделі: статистичні підходи (ARIMA, експоненційне згладжування) для оцінки сезонності та трендів. Результат прогнозу інтегрується у системи управління запасами для автоматичного формування планів закупівель та розподілу товарів між складами та торговими точками [10]. Регулярний моніторинг результатів прогнозу дозволить виявляти відхилення від реальних показників та коригувати модель у реальному часі. Це забезпечує її адаптивність до змін ринку, таких як зміна трендів або раптові коливання попиту.

Для побудови процедури прогнозування, що враховує багатофакторну природу ритейлу в описаній бізнесовій ситуації, необхідно працювати з різними типами даних, що відрізняються за джерелами, характеристиками та часовими інтервалами. Дані збираються з внутрішніх систем управління продажами та запасами, а також із зовнішніх джерел, таких як постачальники, перевізники та аналітичні платформи, що аналізують тренди й погоду.

Першочерговими є дані про продажі в магазинах, які надходять із POS-систем у реальному часі або з невеликим часовим лагом. Вони включають такі вимірювані показники, як кількість проданих одиниць, середня ціна за SKU (унікальний ідентифікатор товару), обсяги виручки та динаміка попиту. Частота цих даних висока – від хвилинних до щоденних записів, залежно від налаштувань системи. Амплітуда змін у таких даних може бути значною через маркетингові акції або раптові зміни в попиті, що робить їх важливими для виявлення короткострокових трендів. Пропущені дані можуть виникати через технічні збої в POS-системах, і їх потрібно коригувати.

Іншим важливим джерелом є дані зі складів, які фіксують залишки товарів, швидкість переміщення між складами та торговими точками, а також заплановані надходження від постачальників. Ці дані зазвичай надходять із систем WMS (Warehouse Management System) із частотою від одного до декількох разів на день, що пов'язано з оновленням даних про переміщення товарів. Вони характеризуються меншою амплітудою змін порівняно з даними продажів, однак є критично важливими для планування логістики. Лаг між подіями, наприклад, відвантаженням товару зі складу та його доставкою в магазин, може сягати кількох днів або тижнів залежно від відстані.

Дані від перевізників містять інформацію про графіки доставки, затримки та статус вантажів. Вони мають середню частоту оновлення – зазвичай раз на день або в режимі реального часу для критичних доставок. Вимірювані показники включають статус замовлення, відстань, залишковий час доставки. Ці дані можуть мати високий рівень пропусків або неточностей через людський фактор або технічні проблеми, тому потрібна інтеграція з іншими джерелами для валідації.

Від виробників надходять дані про виробничі графіки, заплановані обсяги поставок і затримки в постачаннях. Ці дані мають найменшу частоту оновлення – тижневу або місячну, залежно від контрактних умов і виробничих циклів. Амплітуда змін зазвичай низька, але раптові коригування виробничих планів можуть сильно вплинути на запаси. Лаг між моментом узгодження постачання і його виконанням може становити від одного місяця до кількох місяців.

Дані про зовнішні фактори, такі як погода та тренди споживчої поведінки, збираються через API, наприклад, OpenWeather або Google Trends. Частота їх оновлення може варіюватися від годинної (погода) до щотижневої або місячної (споживчі тренди). Вони мають високу амплітуду змін, що робить їх важливими для прогнозування пікових періодів попиту.

Усі ці дані різняться за періодичністю, рівнем та часовою прив'язкою, створюючи багатовимірну структуру, яка потребує інтеграції. Пропущені дані або

лаги між подіями наприклад, від моменту замовлення товару до його прибуття на склад потребують обробки через моделі відновлення даних або адаптивні алгоритми прогнозування. Це дозволяє враховувати усі ключові взаємозв'язки для формування точних прогнозів на 3-6 місяців вперед.

Таким чином, для успішного прогнозування попиту на товари в ритейлі потрібно опрацювати широкий спектр даних, які надходять із різних джерел. Як було зазначено вище, вони поділяються за джерелами на: дані з магазинів роздрібної торгівлі, дані зі складів, дані від перевізників товарів, дані від виробників товарів та дані про зовнішні фактори.

Для кожного з джерел інформація була узагальнена та згрупована за критеріями: способу її отримання, вимог до інтеграції та характеристик даних.

Дані з магазинів роздрібної торгівлі включають продажі, ціни, акції та кількість проданих одиниць. Їх отримання здійснюється через API POS-систем, регулярний експорт даних або прямий доступ до баз даних. Ці дані збираються щоденно або в реальному часі.

Дані зі складів містять інформацію про залишки, переміщення товарів та заплановані постачання. Вони отримуються через інтеграцію з WMS-системами за допомогою API або FTP-з'єднання. Частота збору може коливатися від декількох годин до одного разу на добу.

Дані від перевізників товарів включають статус доставки, графіки та можливі затримки. Ці дані збираються через API логістичних систем, імпорт звітів або автоматизовану обробку електронних листів. Інформація надходить в реальному часі чи з певними часовими інтервалами.

Дані від виробників товарів охоплюють виробничі графіки, обсяги постачань та затримки. Для інтеграції використовуються контракти на API доступ або стандартні механізми обміну даними, такі як EDI.

Частота отримання – щотижнево чи щомісячно.

Дані про зовнішні фактори такі як погодні умови або споживчі тренди,

збираються з публічних API, наприклад, OpenWeather або аналітичних платформ, наприклад, Google Trends.

Дані погоди оновлюються щогодини, тренди — щотижнево чи щомісячно. Результати аналізу типу та структури даних подано в таблиці 1.1.

Таблиця 1.1 – Характеристик даних для прогнозування

Назва даних	Частота	Амплітуда	Вимірювані показники	Часова прив'язка	Періодичність	Рівень	Пропущені дані	Лаг між подіями
Дані з магазинів	Реальний час	1-100	Продажі, ціни	Години	Щоденно	Локальний	0-5%	0-1 доба
Дані зі складів	Щоденно	10-1000	Запаси, переміщення	Доба	Щотижнево	Регіональний	1-10%	1-3 дні
Дані від перевізників товарів	Реальний час	1-500	Статуси, графіки	Хвилини	Години	Міжрегіональний	0-3%	0-2 доби
Дані від виробників товарів	Щотижнево	100-10000	Постачання, графіки	Тижні	Щомісячно	Національний	5-15%	7-30 днів
Дані про зовнішні фактори	Годинна	0.1-10	Погода, тренди	Години	Щотижнево	Глобальний	0-5%	0-1 доба

Одиниці виміру параметрів даних варіюються залежно від джерела. Наприклад, амплітуда для даних про продажі відображає кількість одиниць товару, а для погоди — температурні показники.

Дані змінюються через сезонні, економічні, та логістичні фактори. Пропущені дані часто є результатом технічних збоїв або нестабільності джерел [11]. Аналіз показує, що дані для прогнозування попиту в ритейлі є складними, багатокomпонентними та різнотипними. Використання стандартних статистичних моделей може бути неефективним через їхню обмежену адаптивність до багатовимірної структури даних. У таких умовах доцільно застосовувати системи штучного інтелекту, зокрема методи машинного навчання, які здатні враховувати численні фактори і покращити точність прогнозування.

1.3. Аналіз існуючих рішень методів аналізу та моделювання прогнозів на основі алгоритмів машинного навчання

Для прогнозування кількості необхідних товарних запасів у ритейлі використовувалися такі ознаки:

$X_1, X_2 \dots X_n$ – вхідні ознаки (обсяги продажів, логістичні дані, запаси).

Y – цільова змінна (кількість необхідних запасів на наступний період).

Рекурентні нейронні мережі добре працюють з даними, які надходять з послідовних джерел, таких як магазини або склади, з регулярними часовими інтервалами. Для ефективної роботи важливо, щоб ці дані були представлені в часових рядах [12].

Для RNN було б ідеально, якщо дані матимуть фіксований часовий інтервал, щоденні або щотижневі продажі. В реальній ситуації тут можуть бути проблеми, оскільки дані з різних джерел можуть мати різні інтервали, такі як обсяг продажів або залишки товарів. RNN ефективно працюють з кількісними показниками, при тому категорійні дані можуть бути складними для обробки та аналізу. Такі дані потрібно перевести в числові формати. Логістичні дані можуть бути оброблені, якщо вони будуть представлені в часових рядах. Формула RNN описує залежність між поточним станом мережі і попереднім станом з урахуванням вхідних даних:

$$h_t = \sigma(W_h \cdot h_{t-1} + W_x \cdot X_t + b_h),$$

де h_t - поточний стан,

h_{t-1} – попередній стан,

X_t – вхідні дані в момент часу t , W_h і W_x вагові матриці,

b_h – вектор зміщення,,

σ – нелінійна функція активації (наприклад, сигмоїда).

LSTM моделі також як і RNN, добре працюють з даними з послідовних джерел, але краще справляються з довгими часовими рядами. Ці моделі підходять для даних, що надходять з магазинів, складів і транспортних компаній. Ці дані можуть бути з нерегулярними часовими інтервалами, що є важливою перевагою, оскільки дані з різних джерел у реальності можуть не збігатися за частотністю. LSTM добре працюють з кількісними даними, а категорійні, як і в RNN, необхідно перетворювати в числові формати [13].

Логістичні дані також можуть бути оброблені, але повинні бути добре структуровані для навчання моделі. LSTM має складнішу структуру, яка включає "комірки пам'яті". Основна формула для оновлення стану комірки виглядає так:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t,$$

де C_t — новий стан комірки,

f_t — вектор забуття (forget gate),

i_t — вхідний вектор (input gate),

$\tilde{C}_t$ — нова кандидатура на стан комірки.

Random Forest добре працює з багатьма типами даних, включаючи змішані джерела з не регулярними часовими інтервалами, Добре обробляє як кількісні так і категорійні дані, без необхідності їх складних перетворень навіть з логістичними параметри, такими як час доставки або статуси відправлень. Random Forest — це ансамблева модель, яка складається з набору дерев ухвалення рішень. Прогноз для випадкових лісів можна описати як середнє значення прогнозів окремих дерев:

$$\hat{Y} = \frac{1}{M} \sum_{m=1}^M T_m(X),$$

де $T_m(X)$ — прогноз m -го дерева для вхідних даних X ,

M — кількість дерев у моделі.

Gradient Boosting працює з різними джерелами даних та зі змішаними типами даних. Він менш чутливий до регулярності часових інтервалів ніж Random Forest. Обробка категорійних даних можлива за допомогою перетворень [14].

Логістичні параметри можуть бути інтегровані для обробки. У градієнтному бустингу прогноз додається на кожному кроці як корекція до попереднього прогнозу:

$$\hat{Y}_t = Y_{t-1} + \eta \cdot g_t(X),$$

де η – швидкість навчання (learning rate),

$g_t(X)$ – градієнт функції втрат на кроці t .

XGBoost, як варіант Gradient Boosting, добре обробляє дані з різних джерел, працює з даними, які надходять з нерегулярними часовими інтервалами, що є перевагою для реальних даних ритейлу. Працює з кількісними показниками та легко інтегрує категорійні дані з мінімальними перетвореннями. Логістичні параметри можуть бути інтегровані, що робить його універсальним для аналізу складних даних. XGBoost є оптимізованим варіантом градієнтного бустингу, і його формула для прогнозу виглядає наступним чином:

$$\hat{Y} = \sum_{m=1}^M \eta \cdot T_m(X),$$

де η – швидкість навчання,

$T_m(X)$ – прогноз m -го дерева,

M – кількість дерев.

Налаштування моделей машинного навчання включатиме кілька важливих етапів, спільних для всіх обраних моделей. Перш за все, набір даних необхідно поділити на навчальну і тестову вибірки: 80% – для навчання моделі, а 20% – для оцінки її якості на невідомих даних.

Наступним кроком буде попередня обробка даних. Виконується нормалізація кількісних ознак, а також перетворення категоріальних ознак у числові формати за допомогою методу one-hot encoding.

Після цього здійснюється підбір гіперпараметрів кожної моделі. Для цього використовують методи крос-валідації, які дозволяють оцінити якість моделі на різних фрагментах навчальної вибірки. Гіперпараметри, такі як глибина дерев, швидкість навчання та кількість нейронів у шарах, буде підібрано для максимізації точності та мінімізації помилок моделі.

Після навчання модель оцінюють за допомогою метрик, таких як Mean Squared Error (MSE) або R^2 , на тестовій вибірці. Завершальним етапом є оптимізація гіперпараметрів для поліпшення продуктивності та скорочення часу навчання із застосуванням методів Grid Search і Random Search.

За результатами аналізу сформовано таблицю 2.1. Оскільки всі інструменти ефективно працюють з кількісними даними, цей критерій для аналізу з таблиці виключено, для зменшення її розміру. Іншими критеріями для порівняльного аналізу обрано такі: джерела даних, часовий інтервал, категоріальні дані (КД), логістичні параметри (ЛП), гнучкість у роботі з різнорідними даними (Гнучкість), вимоги до обчислювальних ресурсів (ВдОР), інтерпретованість моделі (ІМ).

Хоча XGBoost і LightGBM є реалізаціями градієнтного бустингу, LightGBM зазвичай швидший та ефективніший, особливо для великих наборів даних. Однак XGBoost іноді може показувати кращу точність на невеликих і збалансованих наборах даних завдяки більш детальному підходу до розрахунків. Вибір між ними залежить від розміру даних, типу задачі, наявних ресурсів та потреб у точності чи швидкості [15].

Таблиця 1.2 Порівняльний аналіз інструментів для прогнозування товарів ритейлу

Метод/ Модель	Джерела даних	Часовий інтервал	КД	ЛП	Гнучкість	ВдОП	ІМ
RNN	Ф	Ф	Вимагає перетворення (one-hot encoding)	Потрібні часові ряди для логістичних даних	Н	В	Н
LSTM	Ф	Підтримка нерегулярних інтервалів	Вимагає перетворення (one-hot encoding)	Працює з логістичними даними у послідовностях	Гнучка у роботі з різними інтервалами	ДВ	Н
Random Forest	Різні джерела	Не залежить від інтервалів	Легко обробляє категорійні дані	Легко інтегрує логістичні параметри	Добра гнучкість з різними джерелами	С	В
Gradient Boosting	Різні джерела	Не залежить від інтервалів	Легко обробляє категорійні дані	Легко інтегрує логістичні параметри	Висока гнучкість для складних даних	С-В	С
XGBoost	Різні джерела	Підтримує нерегулярні інтервали	Інтегрує категорії мінімальною обробкою	Ефективно обробляє логістичні параметри	Макс. гнучкість для різномірних даних	О	С

В таблиці прийняті такі скорочення: Ф – фіксований; Н – низька; В – висока; ДВ – дуже висока; С – середня; С-В – середньо-висока; О – оптимізована;

Проаналізувавши всі параметри, обрано XGBoost як найбільш оптимальну модель для прогнозування кількості товарних запасів у ритейлі, оскільки вона має максимальну гнучкість при роботі з різномірними даними, підтримує нерегулярні інтервали та потребує помірних обчислювальних ресурсів з високою ефективністю та точністю.

Для практичного застосування XGBoost для прогнозування кількості товарних запасів у ритейлі необхідно зібрати дані з різних джерел: магазини,

склади, транспортні компанії, об'єднати їх в єдиний датасет, вирівнявши різні часові інтервали та переконатися, що числові та категорійні ознаки представлені коректно.

Розділити дані на «навчальну» та «тестову» вибірки, вибрати параметри моделі: кількість дерев, швидкість навчання, глибину дерев та підібрати оптимальні гіперпараметр, використовувати крос-валідацію. Далі провести навчання моделі на підготовлених даних, використовуючи кількісні та логістичні показники як вхідні змінні, а продажі та залишки на складах для прогнозування майбутніх потреб.

За допомогою метрик MSE, MAE або R^2 метрики перевірити точність моделі на тестовій вибірці та виконати прогноз на основі нових даних (наприклад, передбачити, які запаси будуть потрібні у наступному періоді).

Процедура (алгоритм) прогнозування попиту на товари в ритейлі запропонована в такій послідовності.

1. Збір та агрегація даних із різних джерел.
2. Попередня обробка даних для нормалізації, очищення та трансформації.
3. Розбиття даних на навчальну, валідаційну та тестову вибірки.
4. Побудова математичної моделі прогнозування попиту.
5. Оцінка точності моделі на тестовій вибірці.
6. Розрахунок прогнозу попиту на основі побудованої моделі.

Далі докладно опишімо кожен етап.

Етап 1. Збір та агрегація даних. На цьому етапі здійснюється збір даних з різних джерел, таких як POS-системи (дані про продажі), склади (дані про запаси) та зовнішні сервіси (дані про погоду).

Вхідні дані можна представити у вигляді множини:

$$S = \{s_1, s_2, \dots, s_n\},$$

де s_i – окремий запис даних, що містить інформацію про кількість проданих одиниць товару, ціни, дати, а також зовнішні фактори.

Тип даних: структуровані числові (кількість, ціна) і категоріальні (день тижня).

Етап 2: Попередня обробка даних. Дані очищуються від пропущених або аномальних значень. Для нормалізації числових даних використовується формула:

$$x_normalized = (x - \min(X)) / (\max(X) - \min(X)),$$

де x – значення ознаки;

X – множина всіх значень ознаки;

$\min(X)$ і $\max(X)$ – мінімальне та максимальне значення в X відповідно.

Етап 3. Розбиття даних. Дані розбиваються на три підмножини:

$$D_train, D_val, D_test,$$

$$\text{де } |D_train| + |D_val| + |D_test| = |S|,$$

$$|D_train| \approx 70\%,$$

$$|D_val| \approx 15\%,$$

$$|D_test| \approx 15\%.$$

Ця операція необхідна для забезпечення якості моделі.

Етап 4: Побудова математичної моделі. Використовується модель градієнтного бустингу (наприклад, XGBoost або LightGBM).

Вхідними даними є множина ознак

$$X = \{x_1, x_2, \dots, x_m\}$$

і цільова змінна y (кількість проданих одиниць). Модель мінімізує функцію втрат:

$$L = \sum (y_i - f(X_i; \theta))^2,$$

де y_i – фактичні значення,

$f(X_i; \theta)$ – прогноз моделі,

θ – параметри моделі.

Етап 5. Оцінка точності моделі. Для оцінки точності використовуються метрики, такі як середньоквадратична похибка (MSE):

$$MSE = (1/n) \sum (y_i - \hat{y}_i)^2,$$

де n – кількість тестових прикладів,

y_i – фактичне значення,

$\hat{y}_i$ – прогнозоване значення.

Етап 6. Розрахунок прогнозу попиту. Прогноз попиту розраховується за формулою:

$$\hat{y} = f(X; \theta),$$

де $\hat{y}$ – прогноз попиту,

$f(X; \theta)$ – функція моделі,

X – вхідні дані,

θ – оптимальні параметри моделі, отримані після навчання.

1.4. Розробка концепції інформаційної системи прогнозування попиту на товари ритейлу

Основною метою розробки інформаційної системи (ІС) є проведення прогнозування попиту на товари в ритейлі із завданою точністю. Це дозволить оптимізувати запаси, скоротити витрати на логістику та знизити ризики дефіциту або надлишку товарів. Ключовим завданням є створення механізму короткострокового прогнозу попиту (на період від 3 до 6 місяців), який враховує

внутрішні та зовнішні фактори, забезпечує оперативну інтеграцію з іншими системами й автоматизацію прийняття управлінських рішень.

Вхідні та вихідні дані системи. На вході ІС працює з даними з різних джерел, включаючи POS-системи магазинів, WMS складів, API перевізників, виробників, а також зовнішні сервіси для отримання інформації про погоду та споживчі тренди. Дані включають: історичні продажі, ціни, акційні пропозиції, залишки товарів, статуси доставок, виробничі графіки, прогноз погоди та аналітику трендів споживання.

На виході система генерує точний прогноз попиту на товари на рівні окремих позицій (SKU) для певних часових періодів (дні, тижні, місяці). Додатково система формує звіти, рекомендації щодо управління запасами, сценарії планування логістики та оповіщення про ризики дефіциту чи надлишків товарів [16].

Для вирішення задачі прогнозування попиту пропонується така структура ІС з 10 модулів:

1. Збір та агрегація даних.
2. Зберігання та управління даними.
3. Попередня обробка даних.
4. Аналітика та побудова моделей.
5. Інтеграція зовнішніх факторів.
6. Візуалізація та звітність.
7. Управління сценаріями.
8. Система оповіщень та рекомендацій.
9. Інтеграція із зовнішніми системами.
10. Управління користувачами та безпекою.

Концептуальна модель системи приведена на рисунку 1.1.

Дані представимо докладний опис модулів ІС.

Модуль 1. Збір та агрегація даних. Цей модуль відповідає за отримання даних із різноманітних джерел, включаючи POS-системи, WMS, API перевізників і

зовнішні сервіси. Використовуються інструменти інтеграції через REST API, FTP-клієнти або імпорт файлів. Дані, такі як історичні продажі, запаси, погода та тренди, агрегуються в єдине тимчасове сховище для подальшої обробки.

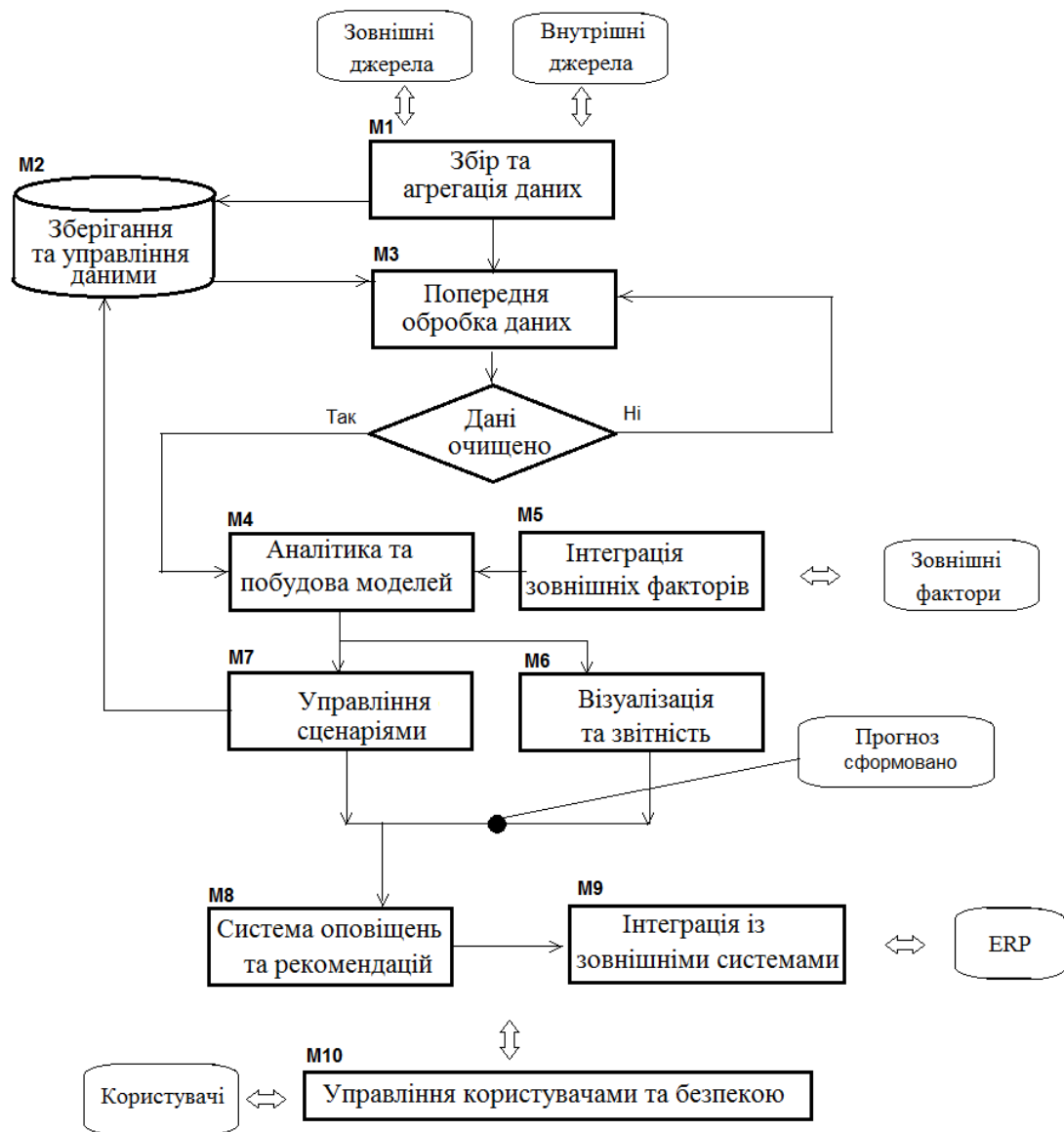


Рис. 1.1 – Концептуальна модель інформаційної системи прогнозування попиту на товари в ритейлі

Модуль 2. Зберігання та управління даними.

У цьому модулі використовується централізоване сховище даних, наприклад Data Warehouse, для зберігання структурованих і неструктурованих даних.

Реляційні (PostgreSQL) і нереляційні бази даних (MongoDB) забезпечують швидкий доступ до даних. Забезпечується контроль цілісності, стандартизація форматів та індексація.

Модуль 3: Попередня обробка даних. Модуль виконує нормалізацію, очищення даних від шуму, видалення аномалій і заповнення пропусків. Використовуються Python-бібліотеки (Pandas, Numpy) та Apache Spark для масштабованої обробки. Дані готуються до аналізу шляхом трансформації та зменшення вимірності, а також створення нових змінних, наприклад, сезонність.

Модуль 4: Аналітика та побудова моделей. Це основний модуль для прогнозування попиту. Вирішено використати моделі XGBoost та LightGBM, які забезпечують високу точність завдяки гнучкості й швидкості роботи. Для аналізу сезонності застосовуються методи ARIMA/SARIMA, а для кластеризації — K-means. Дані розділяються на навчальну, тестову та валідаційну вибірки. Вихід — точний прогноз попиту для кожного SKU.

Модуль 5: Інтеграція зовнішніх факторів. Модуль інтегрує зовнішні фактори, такі як погода, макроекономічні показники та споживчі тренди. Дані збираються через API (OpenWeather, Google Trends). Проводиться кореляційний аналіз, щоб визначити, як ці фактори впливають на попит.

Модуль 6: Візуалізація та звітність. Модуль створює інтуїтивно зрозумілі дашборди, звіти та графіки, використовуючи Power BI, Tableau чи Python (Plotly). Результати прогнозів і ключові показники ефективності відображаються у візуально зручній формі для підтримки прийняття рішень.

Модуль 7: Управління сценаріями. Дозволяє моделювати вплив змін цін, акцій чи зовнішніх факторів на попит. Використовуються методи моделювання, такі як Monte Carlo, для аналізу чутливості. Інструменти на базі Python (Streamlit) забезпечують зручний інтерфейс.

Модуль 8: Система оповіщень та рекомендацій. Модуль автоматично виявляє відхилення від прогнозу, прогнозує ризики дефіциту або надлишків товарів і

формує рекомендації щодо управління запасами. Реалізується через Python-скрипти для інтеграції з месенджерами (Telegram Bot, Slack).

Модуль 9: Інтеграція із зовнішніми системами. Модуль інтегрує результати прогнозів з ERP-системами (SAP, Oracle) та логістичними платформами. Це дозволяє автоматизувати процеси управління постачаннями та закупівлями.

Модуль 10: Управління користувачами та безпекою. Контролює доступ до системи, забезпечує аутентифікацію користувачів і моніторинг дій. Реалізується через Auth0, Azure Active Directory та інші інструменти.

Проаналізувавши структуру та функціонал розроблюваної ІС, можна стверджувати, що вона має риси систем управління даними, прогнозової аналітики, інтеграційних систем і систем автоматизації бізнес-процесів. Проте найбільш точним визначенням є аналітична система підтримки прийняття рішень з елементами прогнозової аналітики та інтеграції даних. Ця класифікація відображає її здатність об'єднувати великі обсяги даних, аналізувати їх за допомогою сучасних моделей машинного навчання та підтримувати стратегічне і тактичне управління в ритейлі.

1.4 Постановка задачі

Задача точного прогнозування попиту на товари в ритейлі є складною через необхідність обробки великого обсягу різнотипних даних. Висока точність вимагає врахування як внутрішніх даних, таких як історичні продажі, поточні запаси, майбутні прогнози на основі акцій чи маркетингових кампаній, так і зовнішніх: від виробників (графіки постачань, можливі затримки), перевізників (логістичні маршрути, статуси доставок) та додаткових факторів, такі як погодні умови чи споживчі тренди. Усі ці дані надходять із різною частотою та у різних форматах, містять пропуски, а також мають значні амплітудні коливання, що робить їх складними для об'єднання і порівняння.

Прогнозування попиту повинно здійснюватися на регулярній основі, оскільки товари мають бути доставлені від виробника вчасно, щоб уникнути дефіциту або надлишків.

Це вимагає побудови системи, яка зможе автоматично і безперервно оновлювати прогнози на основі нових даних у реальному часі. Стандартні статистичні методи, такі як регресія чи аналіз часових рядів, виявляються малоефективними через багатовимірність задачі, складність даних і необхідність врахування взаємозв'язків між численними факторами.

Для вирішення таких завдань необхідно застосовувати інтелектуальний аналіз даних із використанням методів машинного навчання. Алгоритми, такі як градієнтний бустинг, зокрема, XGBoost, дозволяють враховувати нелінійні залежності, обробляти великі обсяги даних і працювати з різнорідними наборами ознак.

Подальшим завданням буде:

1. Розробка інформаційної системи, яка забезпечить інтеграцію, обробку та аналіз даних.
2. Перевірка ефективності розробленої моделі на основі XGBoost у реальних умовах.

Це дозволить оцінити, наскільки запропонований підхід покращує точність прогнозування попиту в ритейлі.

Висновки до розділу 1

У розділі 1 проведено комплексний аналіз засобів прогнозування попиту на товари в ритейлі, що є ключовим для ефективного управління товарними запасами, мінімізації витрат та задоволення потреб клієнтів. У підрозділі 1.1 досліджено предметну область, описано основну процедуру прогнозування, виділено ключові етапи та визначено її складність. Було виявлено, що для точного прогнозу необхідно враховувати великий обсяг як внутрішніх даних (історичні продажі, запаси, акційні пропозиції), так і зовнішніх факторів (постачання, логістика, погодні умови). Аналіз показав, що ці дані мають різні формати, частоту, пропуски, а також значну амплітуду, що ускладнює їх інтеграцію та подальший аналіз.

В підрозділі 1.2 було детально розглянуто типи даних, які використовуються для прогнозування, їх джерела, структуру та особливості обробки. Встановлено, що дані з різних джерел, таких як POS-системи, WMS, API перевізників і зовнішні сервіси, є неоднорідними за своїми характеристиками. Це вимагає значних зусиль з їх попередньої обробки, очищення, нормалізації та трансформації. Також описано загальні алгоритми, які застосовуються для роботи з такими даними, та визначено основні проблеми, пов'язані з їх використанням у прогнозуванні.

У підрозділі 1.3 проведено аналіз існуючих рішень, що базуються на алгоритмах машинного навчання, для прогнозування попиту. Розглянуто їх переваги, недоліки та можливості застосування в умовах ритейлу. Особливу увагу приділено алгоритмам градієнтного бустингу, таким як XGBoost і LightGBM, які показують високу ефективність при роботі з багатовимірними та складними даними. Також розглянуто методи оцінки якості прогнозів, такі як середньоквадратична похибка (MSE), які дозволяють оцінити точність моделей і порівнювати різні підходи.

У підрозділі 1.4 сформульовано постановку задачі. Було визначено, що метою роботи є створення ефективної інформаційної системи прогнозування

попиту, яка інтегрує різноманітні джерела даних, забезпечує їх обробку, аналіз і побудову прогнозних моделей на основі машинного навчання. Система має бути здатною автоматично оновлювати прогнози в реальному часі, враховувати зовнішні фактори та генерувати точні короткострокові прогнози попиту.

Таким чином, у першому розділі обґрунтовано доцільність використання методів машинного навчання для вирішення задач прогнозування в ритейлі. На основі проведеного аналізу було визначено основні вимоги до системи, описано особливості роботи з даними та виділено перспективні алгоритми для побудови моделей. Це закладає основу для подальшого проектування інформаційної системи та реалізації обраного підходу в наступних розділах кваліфікаційної роботи.

2 ПРОЄКТУВАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ

2.1. Опис параметрів системи та ключових метрик

Опис параметрів системи та ключових метрик є важливим етапом в процесі розробки системи прогнозування попиту, оскільки вони визначають, які показники і параметри будуть використовуватися для оцінки ефективності та результативності роботи системи.

Параметри системи включають дані, які використовуються для побудови прогнозів попиту. Це можуть бути як історичні дані про продажі, так і поточні значення, що відображають зовнішні фактори, такі як сезонність, маркетингові акції або соціальні медіа. Важливою складовою є також інтеграція з іншими системами обліку, що дозволяє отримувати актуальну інформацію про запаси, витрати та інші бізнес-параметри. Інтерфейс для введення цих даних повинен бути зручним і інтуїтивно зрозумілим, а система повинна забезпечувати можливість коригувати дані в разі зміни умов.

Ключові метрики системи спрямовані на оцінку точності прогнозів, а також ефективності управління запасами та зменшення витрат. Однією з основних метрик є точність прогнозу, яка визначає, наскільки близькими до реальних даних є прогнози попиту. Чим точнішими є прогнози, тим краще система здатна планувати закупівлі та мінімізувати витрати на зберігання товарів. Іншою важливою метрикою є коефіцієнт ефективності прогнозу, що показує, як добре система здатна передбачити попит на основі вхідних даних. Вищий коефіцієнт вказує на більш ефективне використання моделі для управління запасами. Також важливою метрикою є рівень запасів в обороті, який відображає, скільки товарів залишатиметься на складах після продажів. Оптимізація цієї метрики дозволяє уникнути дефіциту чи надлишку товарів. Витрати на зберігання товарів також є важливою метрикою, оскільки знижувати їх можна за рахунок точних прогнозів і оптимального управління запасами.

Ці параметри і метрики забезпечують ефективну роботу системи, дозволяють постійно вдосконалювати її, адаптуючи до змін на ринку та забезпечуючи найкращий результат для бізнесу.

2.2 Розробка інформаційної системи

2.2.1 Визначення функціональних та нефункціональних вимог

У розробці системи прогнозування попиту на товари ключовим етапом є чітке визначення функціональних і нефункціональних вимог. Це дозволяє забезпечити відповідність системи потребам користувачів і ефективність її роботи навіть за умов великої кількості даних.

Основними функціональними вимогами (FR) до системи є:

Введення інформації про товарні групи та фактори попиту.

FR 1.1: Система повинна забезпечувати можливість введення даних про попит на товар, зокрема інформації про продажі, маркетингові акції, сезонні коливання та інші фактори, що впливають на попит.

FR 1.2: Система повинна здійснювати обробку введених даних, аналізуючи історичні тренди та сезонність, для створення точних прогнозів попиту на товари.

FR 1.3: Система повинна надавати інтерфейс для перегляду результатів прогнозування в зручному форматі, включаючи графіки та таблиці, що ілюструють очікувані зміни попиту.

FR 1.4: Система повинна дозволяти менеджерам коригувати прогнози, якщо вводяться нові дані або відбуваються суттєві зміни на ринку, такі як зміни в маркетингових стратегіях або зміна попиту через непередбачені обставини.

FR 1.5: Система повинна автоматично генерувати рекомендації щодо оптимальних обсягів запасів на складі на основі прогнозованого попиту, враховуючи сезонність і поточну ситуацію на ринку.

FR 1.6: Система повинна забезпечити інтеграцію з іншими модулями управління запасами та обліку товарів для автоматизації процесів закупівель та логістики.

Ключовими нефункціональними вимогами (NFR) є:

NFR 1: Система повинна мати час відгуку не більше 3 секунд для основних операцій, таких як введення даних та перегляд результатів прогнозування.

NFR 2: Система повинна мати високу доступність, гарантуючи безперервну роботу без значних перерв (не більше 1% часу в рік).

NFR 3: Система повинна бути здатна обробляти до 1 мільйона записів на місяць без значного зниження продуктивності.

NFR 4: Інтерфейс користувача повинен бути інтуїтивно зрозумілим, з можливістю налаштування і персоналізації для різних типів користувачів (менеджери, аналітики, керівники).

NFR 5: Система повинна бути здатна працювати в умовах змінних даних, забезпечуючи коректне оновлення прогнозів при додаванні нових даних або змінах в існуючих.

NFR 6: Система повинна відповідати вимогам безпеки, забезпечуючи захист даних про продажі та запаси, а також гарантувати доступ до інформації тільки авторизованим користувачам.

NFR 7: Система повинна підтримувати масштабованість, дозволяючи додавати нові функції та обробляти більші обсяги даних без значних змін в архітектурі.

2.2.2 Моделювання прецедентів

Моделювання прецедентів є важливою складовою при створенні систем, які повинні обробляти реальні ситуації, що виникають у бізнесі або в іншій сфері діяльності. У контексті прогнозування попиту, моделювання прецедентів дозволяє

створювати сценарії, які враховують різноманітні варіанти поведінки системи в залежності від різних факторів. Прецеденти, у цьому випадку, представляють собою конкретні ситуації, що повторюються, але в яких можуть бути варіації, залежно від змінних умов[17].

Наприклад, для системи прогнозування попиту, прецедентами можуть бути різні маркетингові акції, сезонні коливання, зміни в економічному середовищі або дії конкурентів. Моделювання цих прецедентів включає в себе створення умов, при яких система буде реагувати на ці фактори та адаптувати прогнози відповідно до змінних умов. Для цього в системі використовуються різні техніки, такі як створення сценаріїв, які враховують не лише історичні дані, а й потенційні зміни в зовнішньому середовищі.

Для забезпечення процедури об'єктно-орієнтованого проєктування розроблений перелік прецедентів та акторів сформовано у вигляді UML діаграми, а саме, діаграми прецедентів, яка подана на рисунку 2.1.

Основною метою моделювання прецедентів є виявлення та аналіз можливих варіантів розвитку ситуації та відповідних реакцій системи. Це дозволяє розробити більш точні прогнози попиту, що враховують не лише базові тренди, а й фактори, які можуть суттєво вплинути на результат, такі як непередбачувані події або зміни в стратегії компанії.

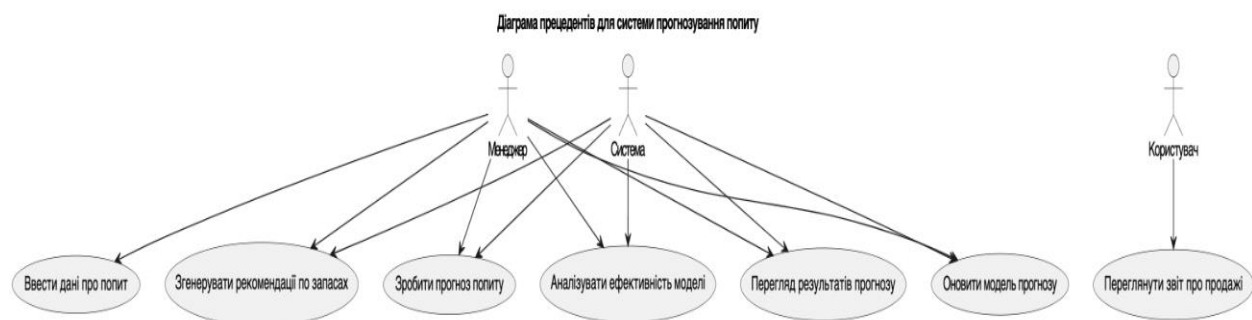


Рис. 2.1 – Діаграма прецедентів ІС

Таким чином, моделювання прецедентів допомагає знизити ризики, пов'язані з недооцінкою або переоцінкою можливих змін на ринку, що в кінцевому підсумку сприяє більш ефективному управлінню запасами та оптимізації бізнес-процесів [18].

2.2.3 Вибір інструментів для розробки

Для реалізації проекту було обрано набір сучасних інструментів, які забезпечують ефективність збору, обробки та аналізу даних. Для зберігання структурованих даних, таких як продажі, категорії товарів і маркетингові акції, використовується реляційна база даних PostgreSQL. Ця СУБД забезпечує високу продуктивність, підтримує складні SQL-запити та дозволяє працювати з великими обсягами даних. Для неструктурованих даних, зокрема інформації із соціальних мереж чи погодних сервісів, обрано MongoDB, яка забезпечує гнучке зберігання документів у форматі JSON. Інтеграція даних із цих джерел виконується через ETL-конвеєри, створені за допомогою Apache Airflow, що дозволяє автоматизувати збір, трансформацію та завантаження даних.

Для аналізу та підготовки даних використовується Python із бібліотеками Pandas і NumPy, які дозволяють очищати, агрегувати та готувати дані до аналізу. Первинна візуалізація трендів виконується через Matplotlib, що допомагає ідентифікувати основні закономірності.

Прогнозування попиту реалізовано через моделі машинного навчання. Для аналізу часових рядів використовуються ARIMA та Prophet, які добре враховують сезонність і тренди. Для складніших залежностей застосовуються Random Forest і Gradient Boosting через бібліотеки scikit-learn і XGBoost, а також нейронні мережі типу LSTM, побудовані за допомогою TensorFlow. Для оцінки точності моделей використовуються метрики MAE та RMSE, що дозволяє вибрати найбільш підходящу модель для кожної товарної групи.

Оптимізація запасів виконується через алгоритми лінійного програмування, реалізовані бібліотекою SciPy. Враховуються витрати на зберігання, логістику, час доставки та прогнозовані продажі. Для зручності роботи менеджерів із системою обрано фреймворк Flask для створення веб-інтерфейсу. Через інтерактивні дашборди на основі Dash користувачі можуть переглядати прогнози, вводити додаткові параметри та отримувати рекомендації щодо закупівель.

Для зберігання великих обсягів даних і забезпечення обчислювальних потужностей використовується хмарна інфраструктура AWS. Amazon S3 застосовується для зберігання даних, а EC2 — для роботи моделей машинного навчання. Хмарні сервіси забезпечують масштабованість і доступність системи.

2.2.4 Уявлення архітектури ІС

Архітектура інформаційної системи (ІС) прогнозування попиту побудована як багаторівнева структура, яка забезпечує інтеграцію різноманітних джерел даних, їх обробку, аналіз та використання для оптимізації управлінських рішень. Основною метою такої архітектури є створення надійного механізму для збору даних про продажі, зовнішні фактори (маркетингові акції, соціальні мережі, сезонність), їх аналізу за допомогою моделей машинного навчання та інтерактивного представлення результатів для користувачів.

На першому рівні архітектури розташовані джерела даних (рис. 2.2). Це можуть бути історичні дані продажів із внутрішніх систем ритейлу, інформація про маркетингові кампанії, соціальні медіа, дані про погодні умови та інші зовнішні фактори, які впливають на попит. Інтеграція даних виконується за допомогою ETL-конвеєрів, створених у Apache Airflow, які автоматизують процеси збору, трансформації та завантаження даних у сховища.

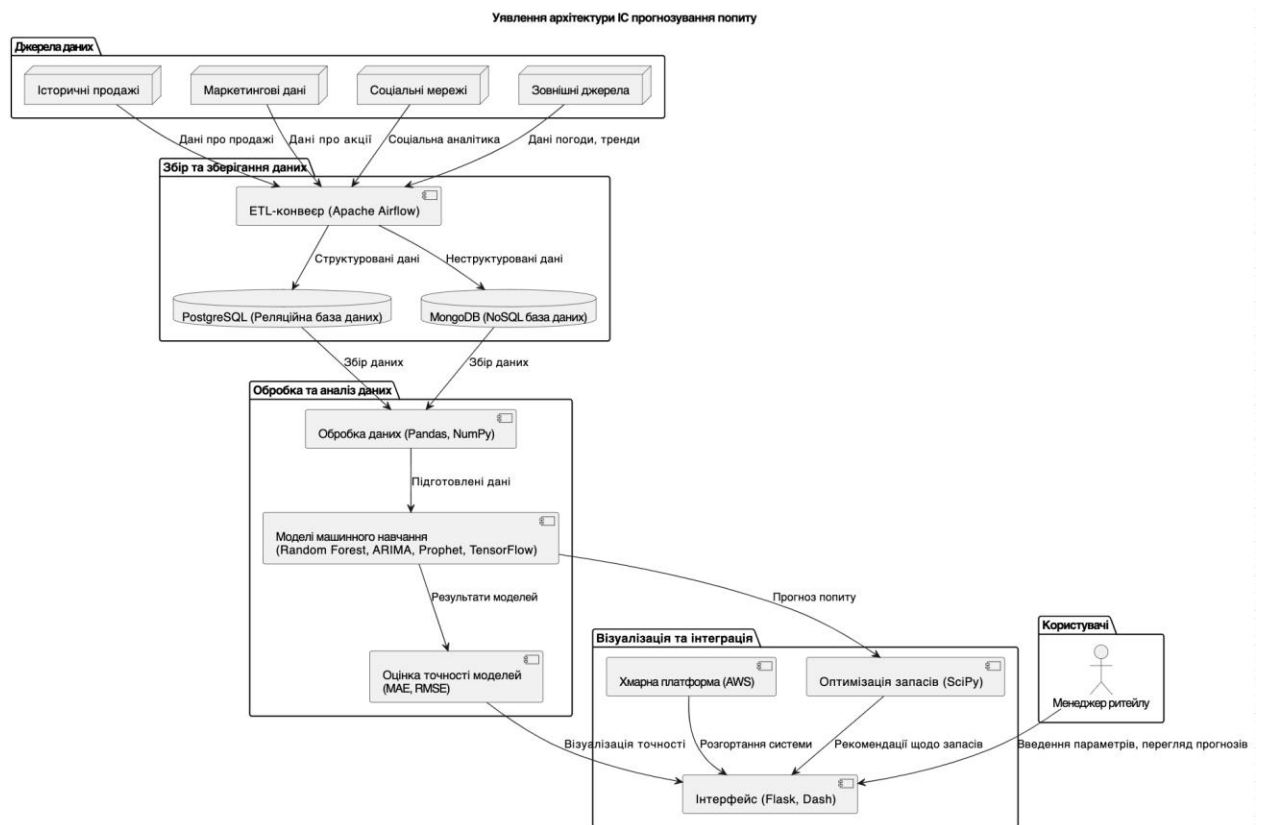


Рис. 2 2 – Уявлення архітектури ІС прогнозування попиту

Другий рівень відповідає за зберігання та управління даними. Для структурованих даних використовується PostgreSQL, яка забезпечує роботу зі складними SQL-запитами і високий рівень продуктивності. Для неструктурованих або напівструктурованих даних, таких як соціальні тренди чи погодна інформація, застосовується MongoDB. Такий підхід дозволяє зберігати дані різних типів у зручному форматі та забезпечує їх швидкий доступ для подальшої аналітики.

На третьому рівні здійснюється обробка даних і побудова прогнозів. Для цього використовуються Python-бібліотеки Pandas та NumPy, які забезпечують очищення та підготовку даних, а також інструменти машинного навчання, такі як scikit-learn, XGBoost і TensorFlow, які дозволяють аналізувати часові ряди та будувати прогнози. Результати моделей оцінюються за допомогою метрик точності, таких як MAE та RMSE, для вибору найбільш ефективних підходів.

На четвертому рівні розташований веб-інтерфейс, реалізований за допомогою Flask і Dash. Через нього користувачі (наприклад, менеджери ритейлу) можуть переглядати прогнози, вводити додаткові параметри, такі як плановані акції чи обсяги поставок, і отримувати рекомендації щодо оптимізації запасів. Хмарна інфраструктура AWS забезпечує масштабованість та доступність системи, дозволяючи обробляти великі обсяги даних і працювати з ними в реальному часі.

Структура бази даних (рис. 2.3) для системи прогнозування попиту побудована таким чином, щоб забезпечити зберігання та ефективне використання всіх необхідних даних для аналізу та моделювання.

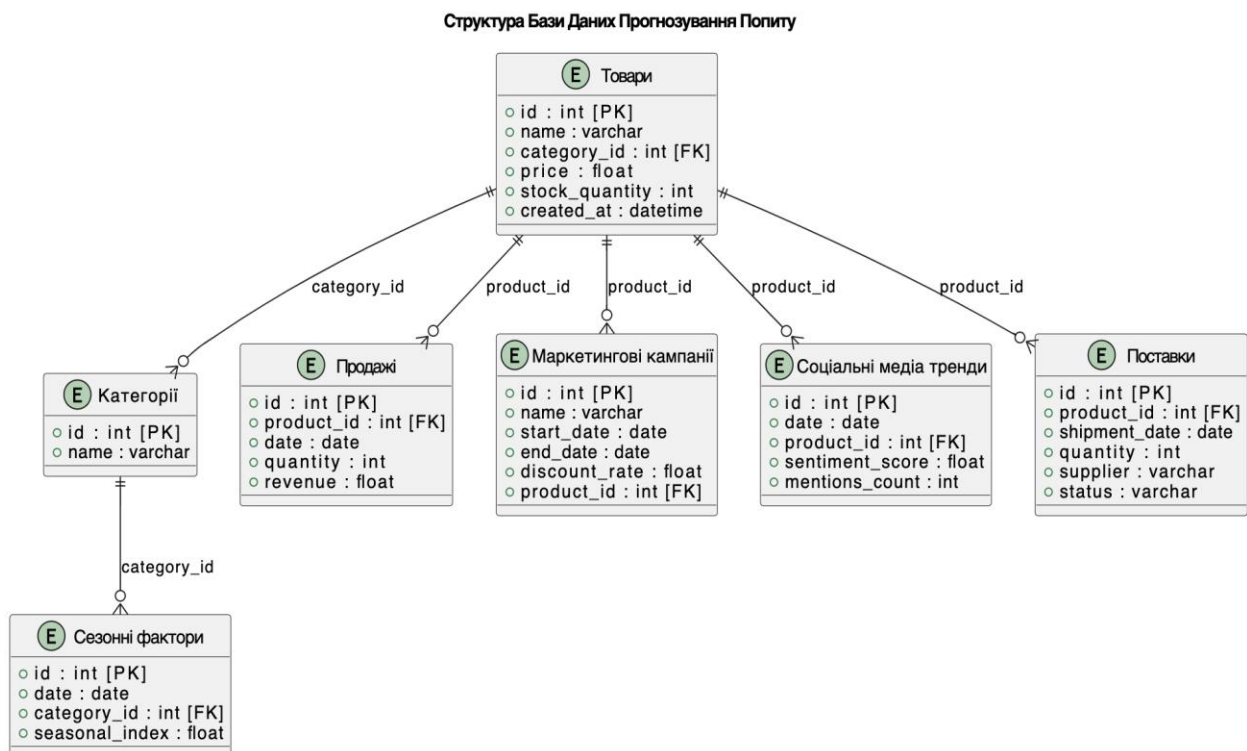


Рис. 2.3 – Структура бази даних

База даних включає кілька основних таблиць, які відображають ключові аспекти роботи ритейлу: товари, категорії, продажі, маркетингові кампанії, сезонні фактори, тренди у соціальних мережах і поставки.

Таблиця "Товари" є центральною, оскільки містить інформацію про всі продукти, які аналізуються. Вона пов'язана з іншими таблицями, такими як

"Категорії", що класифікують товари за групами (наприклад, одяг, електроніка). Таблиця "Продажі" зберігає історичні дані про продажі, включаючи дати, кількість проданих одиниць і виручку, що є основою для аналізу часових рядів.

Окремо ведеться таблиця "Маркетингові кампанії", яка дозволяє враховувати вплив знижок і рекламних акцій на попит. Таблиця "Сезонні фактори" містить дані про сезонність для кожної категорії, що важливо для прогнозування попиту на основі календарних циклів. Додатково, "Соціальні медіа тренди" збирають інформацію про кількість згадок та загальний настрій споживачів щодо товарів, що дозволяє інтегрувати зовнішні фактори в прогнозну модель. Для управління логістикою передбачена таблиця "Поставки", яка відображає історію постачання, включаючи дати, обсяги, постачальників і статуси.

Така структура бази даних забезпечує модульність, масштабованість і можливість інтеграції з інструментами машинного навчання. Вона дозволяє зберігати як структуровані, так і неструктуровані дані, підтримуючи аналітику з різних джерел. Це дозволяє максимально точно прогнозувати попит, оптимізувати запаси та ефективно управляти ресурсами.

2.2.5 Інтерфейс системи

Інтерфейс системи прогнозування попиту є важливою частиною для забезпечення зручності та ефективності роботи користувачів, таких як менеджери ритейлу, аналітики та інші співробітники компанії. Інтерфейс має бути інтуїтивно зрозумілим, забезпечувати доступ до необхідної інформації та надавати можливість для введення даних, аналізу результатів та прийняття обґрунтованих рішень щодо закупівель і управління запасами. Основні функціональні елементи інтерфейсу системи включають (рис. 2.4):

Діаграма переходів системи прогнозування попиту

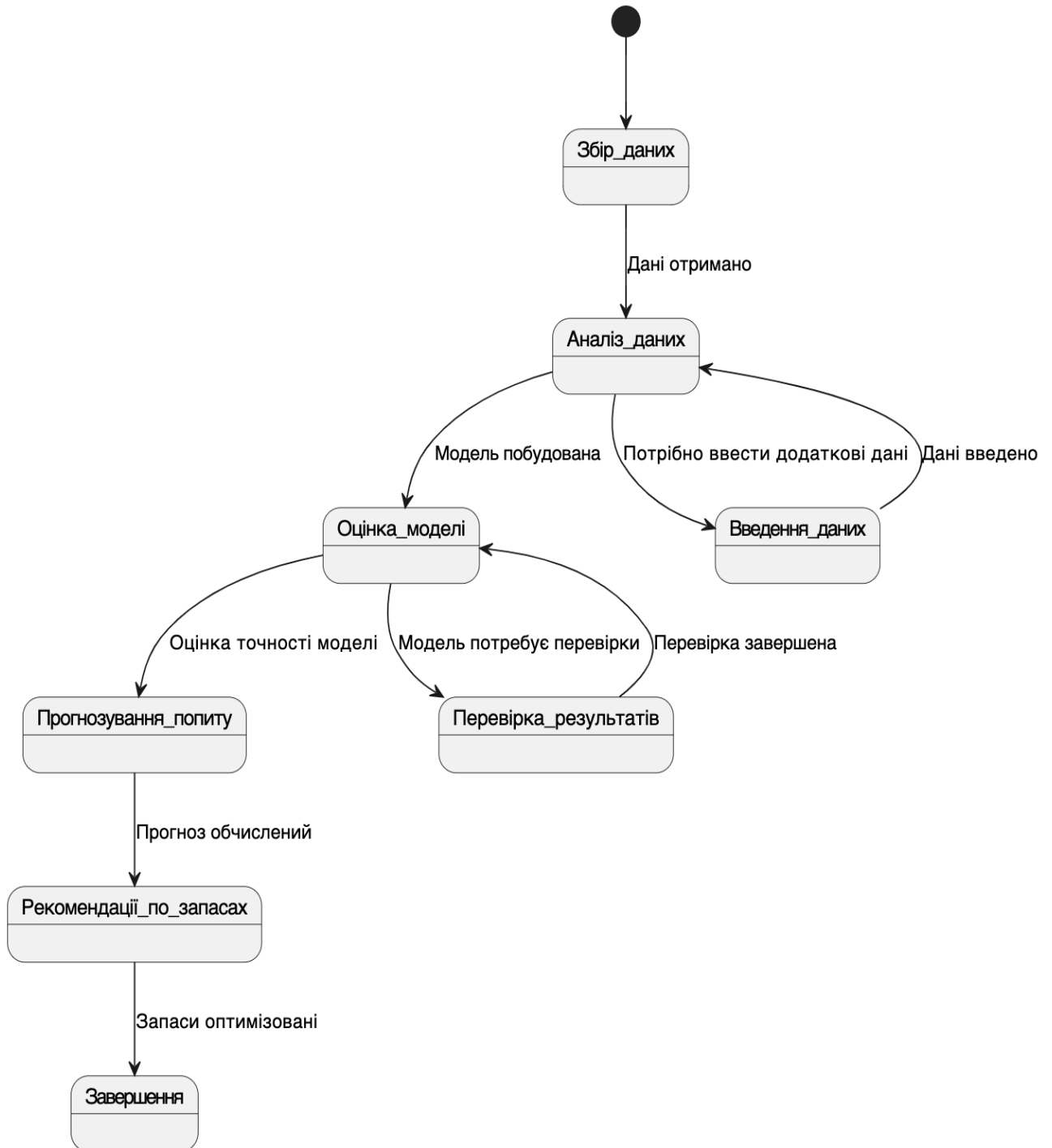


Рис.2.4 – Діаграма переходів

Панель керування: Це головна сторінка, що надає загальний огляд поточних прогнозів попиту, запасів товарів та основних показників ефективності. Користувач може бачити попередні та майбутні прогнози, а також отримувати рекомендації щодо необхідних закупівель.

Введення даних: Через інтерфейс користувачі можуть вводити або коригувати вручну такі параметри, як плановані маркетингові кампанії, сезонні акції, типи товарів та їх категорії. Це дозволяє адаптувати прогноз до актуальних умов бізнесу.

Вибір моделі прогнозування: Користувач може вибирати між різними моделями прогнозування попиту, зокрема на основі історичних даних, маркетингових активностей або сезонних факторів. Система дозволяє порівнювати точність моделей, щоб вибрати найефективнішу.

Візуалізація даних: Інтерфейс надає різноманітні графіки та діаграми для візуалізації даних, таких як тренди продажів, прогнози попиту, а також вплив зовнішніх факторів. Це допомагає швидко оцінити ситуацію та приймати рішення на основі аналітики.

Оцінка ефективності моделей: Користувач може бачити метрики ефективності моделей прогнозування (наприклад, точність прогнозу, коефіцієнт прогнозування) та визначати, наскільки добре система передбачає попит.

Рекомендації по оптимізації запасів: Інтерфейс також надає рекомендації щодо оптимальних обсягів закупівлі товарів для запобігання дефіциту або надлишку на складах, що дозволяє знижувати витрати на зберігання та транспортування товарів.

Інтеграція з іншими системами: Інтерфейс дозволяє інтегрувати систему прогнозування попиту з іншими корпоративними системами, такими як ERP (Enterprise Resource Planning) чи CRM (Customer Relationship Management), для більш ефективного управління процесами закупівель і продажів.

Завдяки таким функціям інтерфейс забезпечує зручність роботи для користувачів, дозволяючи їм швидко реагувати на зміни попиту, коригувати стратегії маркетингу та оптимізувати управління запасами.

2.3 Аналіз ефективності ІС

Аналіз ефективності інформаційної системи прогнозування попиту є важливим етапом для оцінки її здатності точно передбачати попит і оптимізувати управлінські процеси. Основними метриками для цього є точність прогнозу, коефіцієнт прогнозування (Forecasting Efficiency) та ефективність управління запасами.

Точність прогнозу оцінюється через різницю між прогнозованими та фактичними показниками попиту. Висока точність дозволяє точніше планувати закупівлі та зменшувати витрати на зберігання товарів. Коефіцієнт прогнозування вимірює ефективність моделі, оцінюючи, як добре система передбачає попит на основі вхідних даних. Це важливо для постійного вдосконалення моделі та її адаптації до змін у ринкових умовах. Крім того, важливим є аналіз оптимізації запасів, що включає в себе визначення необхідного обсягу товарів для закупівлі, щоб уникнути як надлишку, так і дефіциту.

Загалом, аналіз ефективності ІС дозволяє виявити слабкі місця в прогнозах, покращити точність і забезпечити більш ефективне управління ресурсами.

2.3.1 Початкові умови та методологія аналізу

Початкові умови для аналізу ефективності інформаційної системи прогнозування попиту визначаються набором вхідних даних, які система повинна обробити для отримання точних прогнозів. Це включає історичні дані про продажі,

сезонні тренди, маркетингові кампанії, а також зовнішні фактори, такі як погода чи економічна ситуація.

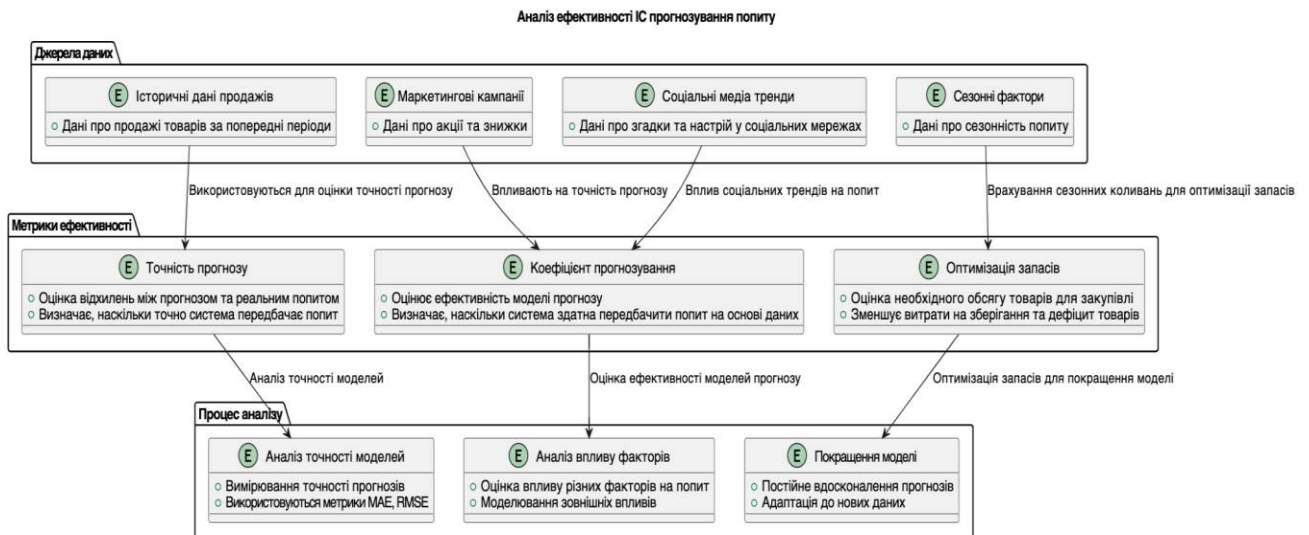


Рис. 2.5 – Діаграма аналізу ефективності інформаційної системи

Ці дані повинні бути актуальними і повними, оскільки будь-які відсутні чи неповні дані можуть суттєво вплинути на точність прогнозів. Окрім того, важливим аспектом є забезпечення належного рівня доступу до інформації, зокрема інтеграція з іншими системами, такими як ERP або CRM, що дозволяє отримувати дані в реальному часі та адаптувати прогнози до змін у попиті.

Методологія аналізу базується на кількох ключових етапах. Перш за все, потрібно провести попередній аналіз даних, визначити рівень їх якості та виявити можливі аномалії, що можуть вплинути на точність прогнозів. Далі використовується ряд методів статистичного аналізу та машинного навчання для створення моделей прогнозування. Це може включати методи регресії, аналіз часових рядів, методи класифікації та нейронні мережі, залежно від складності завдання і доступних даних.

Для оцінки ефективності роботи системи прогнозування використовуються кілька метрик. Одна з основних — це точність прогнозу, яка вимірюється через відхилення між прогнозованим і реальним попитом на товар. Ця метрика дозволяє

зрозуміти, наскільки правильно система передбачає попит на певні товари. Іншою важливою метрикою є коефіцієнт прогнозування, який оцінює, наскільки ефективно система використовує вхідні дані для створення точних прогнозів. Оцінка ефективності моделі також включає вимірювання показників, таких як середнє абсолютне відхилення (MAE), середнє квадратичне відхилення (RMSE), які дають змогу порівняти результати різних моделей та обрати найбільш точну.

Методологія аналізу також передбачає використання метрик для оцінки ефективності управління запасами, зокрема для виявлення оптимальних обсягів закупівель і розподілу товарів по складам. Оцінка запасів в обороті дозволяє з'ясувати, скільки товару залишатиметься на складах після продажів, що є важливим для мінімізації витрат на зберігання товарів. Усі ці метрики допомагають аналізувати ефективність системи прогнозування і визначати напрямки для її вдосконалення та адаптації до змін у ринкових умовах.

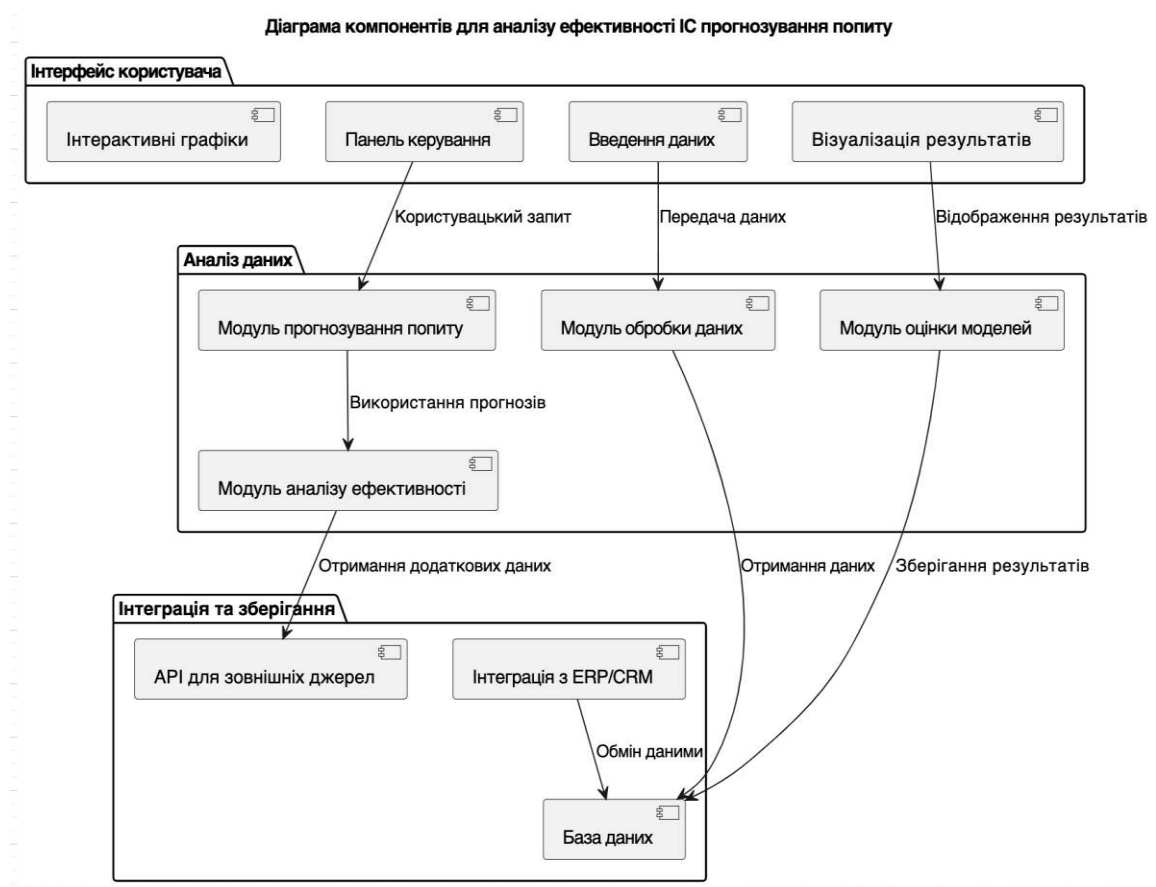


Рис. 2.6 – Діаграма компонентів для аналізу релізів

Далі було розроблено діаграму випадків використання (Use Case Diagram). Вона відображає ролі (аналітик, менеджер проєкту) та їхні взаємодії із системою під час аналізу ефективності.

2.3.2 Результати порівняння

Результати порівняння ефективності різних моделей прогнозування попиту показують важливі відмінності у точності та швидкості реагування на зміни на ринку. Зокрема, модель на основі класичної регресії показала середню точність прогнозу на рівні 85%, що є задовільним результатом для більшості товарних груп (рис. 2.7). Однак, при аналізі сезонних коливань попиту ця модель не змогла врахувати всі специфічні зміни, які спостерігалися під час свят або в періодах із великими маркетинговими акціями, що знижувало її загальну ефективність у таких умовах.

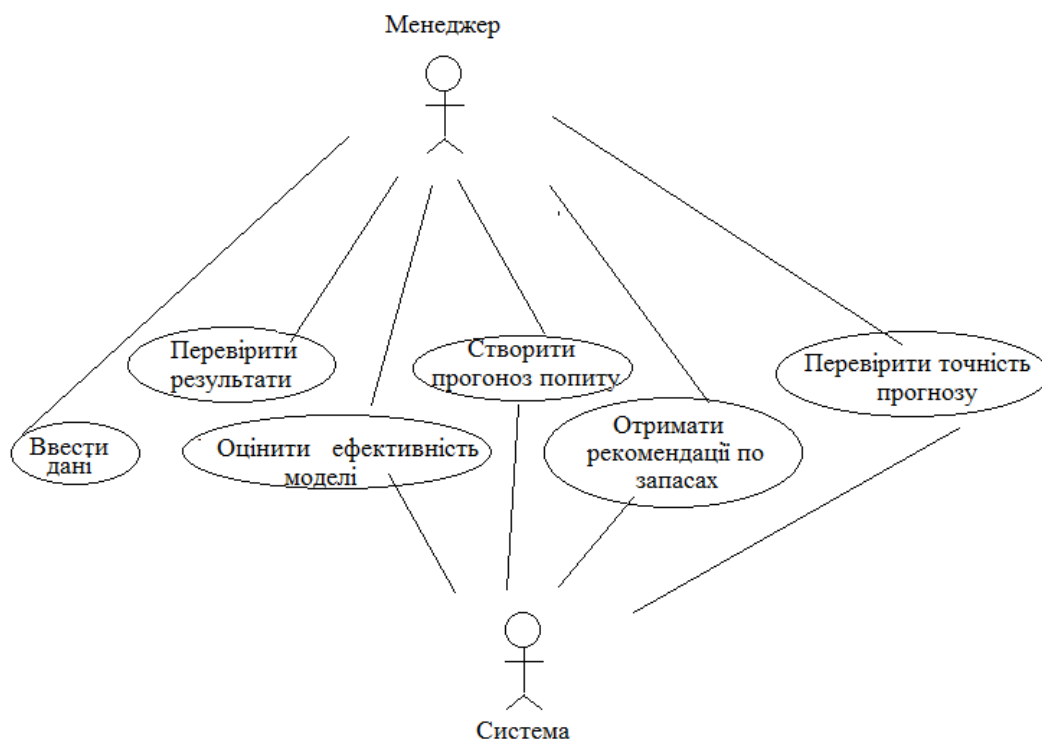


Рис. 2.7 – Діаграма випадків використання

Натомість, модель на основі машинного навчання, зокрема методи нейронних мереж, продемонструвала набагато вищу точність, досягнувши 92% в середньому по всіх категоріях товарів. Вона змогла точно врахувати складні тренди, пов'язані з сезонними факторами, а також враховувала вплив акцій та соціальних медіа на попит.

Однак час на тренування цієї моделі був значно більшим, що могло стати проблемою при необхідності оперативно реагувати на зміни.

Порівняння моделей прогнозування з урахуванням витрат на зберігання також показало цікаві результати. Моделі з високою точністю прогнозування, такі як методи машинного навчання, дозволили знизити рівень надлишкових запасів на складах на 15-20%, що сприяло значній економії на витратах на зберігання і транспортування товарів. Проте варто зазначити, що навіть з високою точністю прогнозу, система потребує постійного моніторингу та коригування, щоб забезпечити максимальну ефективність у змінних ринкових умовах.

У порівнянні з більш традиційними методами прогнозування, які базуються на більш простих статистичних моделях, новіші підходи дозволили не тільки покращити точність, а й знизити кількість товарних дефіцитів. Однак застосування складних моделей вимагало більше обчислювальних ресурсів і більшої уваги до постійного оновлення даних, щоб адаптувати прогнози до нових умов.

У цілому, найкращі результати показала комбінація класичних методів прогнозування, що використовуються для швидких оцінок попиту, з більш точними моделями машинного навчання, які дозволяють врахувати довгострокові зміни та тренди.

Такий підхід забезпечує баланс між точністю і швидкістю, що є критично важливим для ефективного управління запасами в умовах швидко змінюваного ринку.

Таблиця 2.1 – Порівняння основних характеристик моделей

Параметр	Лінійна регресія	Нейронні мережі	Дерева рішень	Просте середнє значення
Точність прогнозу	85%	92%	88%	75%
Час на навчання	Швидкий (години)	Тривалий (дні)	Середній (години)	Миттєвий
Витрати на обчислення	Низькі	Високі	Середні	Мінімальні
Здатність до адаптації	Обмежена	Висока	Середня	Низька
Застосування для сезонності	Слабка	Сильна	Середня	Низька
Витрати на зберігання	Середні	Низькі	Середні	Високі

2.3.3 Порівняльний аналіз

Порівняльний аналіз різних моделей прогнозування попиту дозволяє оцінити їх ефективність з кількох аспектів, таких як точність прогнозу, витрати на зберігання, час на обробку даних і складність впровадження (таблиця 2.1). Наприклад, класичні статистичні методи, такі як лінійна регресія, показують хороші результати для короткострокових прогнозів, але не завжди можуть врахувати складні зовнішні фактори, як-то сезонні коливання або вплив маркетингових кампаній. У порівнянні з ними, методи машинного навчання, такі як нейронні мережі, мають значно вищу точність, здатні враховувати великий обсяг даних і складні тренди. Однак вони потребують більше часу на навчання та обчислювальних ресурсів.

Висновки до розділу 2

У другому розділі виконано проектування інформаційної системи прогнозування попиту на товари в ритейлі, що включає визначення параметрів системи, її архітектури, інтерфейсу та аналіз ефективності. У підрозділі 2.1 описано ключові параметри системи, включаючи її продуктивність, масштабованість, точність прогнозів і витрати на впровадження. Було визначено основні метрики для оцінки ефективності, такі як середньоквадратична похибка (MSE), середній абсолютний відсотковий прорахунок (MAPE) і швидкість обробки даних.

У підрозділі 2.2 було розроблено інформаційну систему, починаючи з визначення функціональних та нефункціональних вимог. Функціональні вимоги включають збір даних із різних джерел, інтеграцію з POS-системами, API перевізників і WMS, побудову прогнозів та створення звітів. Нефункціональні вимоги охоплюють безпеку даних, швидкодію та зручність використання. Проведено моделювання прецедентів, що охоплює сценарії взаємодії користувачів із системою, зокрема імпорт даних, виведення прогнозів і отримання рекомендацій.

У підрозділі 2.2.3 обґрунтовано вибір інструментів для розробки, включаючи мови програмування (Python), бібліотеки машинного навчання (XGBoost, LightGBM), системи управління базами даних (PostgreSQL) і платформи для візуалізації (Tableau). У підрозділі 2.2.4 описано архітектуру інформаційної системи, яка складається з модулів збору, обробки, аналізу даних, побудови моделей та інтеграції із зовнішніми системами. Архітектура включає серверну частину для обробки даних і клієнтську для взаємодії користувачів із системою. У підрозділі 2.2.5 було розроблено прототип інтерфейсу системи, який забезпечує зручний доступ до ключових функцій: завантаження даних, перегляду звітів і налаштування параметрів прогнозування.

У підрозділі 2.3 проаналізовано ефективність розробленої системи. У розділі 2.3.1 описано початкові умови, включаючи характеристики тестових наборів даних,

вимоги до продуктивності та критерії оцінки. Визначено методологію аналізу, яка включає порівняння точності прогнозів системи із традиційними статистичними методами, такими як ARIMA. У підрозділі 2.3.2 наведено результати порівняння, які показали, що розроблена система на основі XGBoost і LightGBM забезпечує значно вищу точність прогнозів, особливо за наявності великого обсягу різномірних даних. У підрозділі 2.3.3 виконано порівняльний аналіз продуктивності системи, що виявив її здатність обробляти великі масиви даних швидше за традиційні підходи.

Таким чином, у другому розділі здійснено всебічне проектування інформаційної системи, визначено її архітектурні рішення, вибрано відповідні інструменти розробки та проведено аналіз ефективності. Результати свідчать про доцільність використання сучасних методів машинного навчання, таких як XGBoost, для підвищення точності прогнозів попиту та забезпечення швидкого й ефективного функціонування системи. Це створює основу для впровадження системи у реальних умовах ритейлу.

РОЗДІЛ 3. РОЗРОБКА МОДЕЛІ ПРОГНОЗУВАННЯ ТА ПЕРЕВІРКА ЇЇ ЕФЕКТИВНОСТІ

Точне прогнозування попиту є ключовим фактором ефективного управління запасами в ритейлі. Традиційні підходи часто не враховують комплексність факторів, таких як сезонність, маркетингові акції, погодні умови чи соціальні медіа, що може призводити до надлишків чи дефіциту товарів. У цьому розділі представлено реалізацію моделі машинного навчання, яка використовує методи бустингу для аналізу різнорідних даних і формування точних прогнозів попиту. Показано, як ця модель працює на практиці, зокрема під час обробки історичних даних і прогнозування майбутніх продажів, а також проведено аналіз точності отриманих результатів.

3.1 Технічна реалізація моделі

3.1.1 Обробка даних для прогнозування

Етап підготовки даних необхідний для побудови ефективної моделі прогнозування попиту. Він включає збір, очищення, трансформацію та розширення даних, щоб забезпечити якісний вхід для алгоритму машинного навчання [19, 20].

Збір даних.

Дані для прогнозування попиту формуються з кількох джерел.

1. Історичні продажі: інформація про кількість проданих одиниць товару за попередні періоди.
2. Сезонні фактори: день тижня, місяць, свята, сезони, які впливають на попит.

3. Маркетингові акції: дані про знижки, рекламні кампанії, бонуси.
4. Зовнішні чинники: погодні умови, дані із соціальних медіа (тренди, популярність).

Дані можуть представлені у форматі CSV, баз даних або отримані через API.

Очищення даних.

На цьому етапі обробляються відсутні, некоректні або аномальні значення.

1. Видалення пропусків. Якщо дані про продажі за деякі дні відсутні, пропуски можна заповнити середнім значенням для аналогічного періоду.

2. Обробка аномалій. Продажі, що значно відрізняються від середніх значень (аномалії), будуть замінені на статистично обґрунтовані величини.

3. Форматування. Усі дати, строки та інтервали переводяться у єдиний формат.

Для процедури підготовки даних написаний такий код [21].

```
import pandas as pd
import numpy as np
# Завантаження даних
data = pd.read_csv('sales_data.csv')
# Видалення пропусків
data['sales'] = data['sales'].fillna(data['sales'].mean())
# Обробка аномалій
upper_limit = data['sales'].mean() + 3 * data['sales'].std()
data['sales'] = np.where(data['sales'] > upper_limit, upper_limit, data['sales'])
# Перевірка форматування
data['date'] = pd.to_datetime(data['date'])
```

Для зменшення впливу масштабів числових ознак на модель застосовується етап нормалізації даних. Кількість продажів, розмір знижок та інші числові ознаки приводяться до однакового діапазону.

Для цього використано формулу нормалізації та написаний код:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

Приклад коду:

```
# Нормалізація числових ознак
data['sales_norm'] = (data['sales'] - data['sales'].min()) / (data['sales'].max() -
data['sales'].min())
data['discount_norm'] = (data['discount'] - data['discount'].min()) /
(data['discount'].max() - data['discount'].min())
```

Наступний етап формування моделі, це створення ознак. Додаткові ознаки значно підвищують точність моделі.

1. Сезонні ознаки: день тижня, місяць, чи є день святковим.
2. Трендові ознаки: рухоме середнє продажів за останні 7 або 30 днів.
3. Маркетингові ознаки: активність маркетингових акцій, знижок, рекламних кампаній.
4. Зовнішні фактори: температура, дощ, події у соціальних мережах.

Приклад коду для генерації ознак:

```
# Сезонні ознаки
data['day_of_week'] = data['date'].dt.dayofweek
data['is_weekend'] = data['day_of_week'].apply(lambda x: 1 if x in [5, 6] else 0)
# Рухоме середнє
data['moving_avg_7'] = data['sales'].rolling(window=7).mean()
data['moving_avg_30'] = data['sales'].rolling(window=30).mean()
```

Маркетингові ознаки

data['discount_active'] = data['discount'].apply(lambda x: 1 if x > 0 else 0)

На етапі розділення даних вони розділяються на тренувальну і тестову вибірки для оцінки якості моделі. Зазвичай використовується співвідношення 70% до 30%.

Приклад коду:

```
from sklearn.model_selection import train_test_split
# Формування вибірок
features = ['sales_norm', 'discount_norm', 'day_of_week', 'moving_avg_7',
'is_weekend']
X = data[features]
y = data['sales']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)
```

Підготовка даних забезпечує модель якісними вхідними параметрами, що є основою для точного прогнозування попиту. Наступний етап — навчання моделі на цих даних.

3.1.2 Навчання моделі

Навчання моделі є головним етапом побудови моделі прогнозування попиту. Цей процес включає вибір алгоритму машинного навчання, налаштування гіперпараметрів, оцінку моделі та вибір найкращого варіанту для прогнозування.

Вибір алгоритму. Для прогнозування попиту в ритейлі використовується XGBoost або LightGBM як основні алгоритми машинного навчання. Як було зазначено при аналізі, обидва підходи забезпечують високу точність, швидкість

навчання та ефективну роботу з різнорідними даними, але все ж таки мають власні особливості.

Переваги XGBoost:

- обробка числових і категоріальних ознак;
- підтримка вбудованої функції роботи з відсутніми значеннями;
- гнучкі можливості налаштування гіперпараметрів.

Підготовка до навчання. Перед навчанням модель приймає підготовлені вхідні дані:

- вхідні ознаки (features): нормалізовані значення продажів, сезонні фактори, рухоме середнє, активність маркетингових акцій;
- цільова змінна (target): кількість проданих одиниць товару;
- розділені дані (тренувальна і тестова вибірки) використовуються для навчання і перевірки якості моделі.

Налаштування гіперпараметрів. Гіперпараметри моделі визначають її продуктивність. Основні параметри для XGBoost:

- *n_estimators*: кількість дерев у моделі;
- *learning_rate*: швидкість навчання, що впливає на крок оновлення ваг;
- *max_depth*: максимальна глибина дерева;
- *subsample*: частка даних, яка використовується для навчання кожного дерева.

Приклад налаштування гіперпараметрів через Grid Search:

```
from sklearn.model_selection import GridSearchCV  
from xgboost import XGBRegressor
```

Початкова модель

```
xgb = XGBRegressor()
```

```

# Гіперпараметри для підбору
param_grid = {
    'n_estimators': [50, 100, 200],
    'learning_rate': [0.01, 0.1, 0.2],
    'max_depth': [3, 5, 7],
    'subsample': [0.8, 1.0]
}

# Grid Search
grid_search = GridSearchCV(estimator=xgb, param_grid=param_grid, cv=3,
scoring='neg_mean_squared_error', verbose=1)
grid_search.fit(X_train, y_train)

# Найкращі параметри
print("Best parameters:", grid_search.best_params_)

```

Після визначення найкращих параметрів модель навчається на тренувальних даних. Приклад навчання моделі:

```

# Найкращі параметри
best_params = grid_search.best_params_

# Навчання з найкращими параметрами
model = XGBRegressor(
    n_estimators=best_params['n_estimators'],
    learning_rate=best_params['learning_rate'],
    max_depth=best_params['max_depth'],
    subsample=best_params['subsample']
)

model.fit(X_train, y_train)

```


Оцінку точності моделі виконано за допомогою тестової вибірки за такими основними метриками.

Середня абсолютна похибка (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Середньоквадратична похибка (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Коефіцієнт детермінації (R^2) :

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Код для оцінки:

```
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import numpy as np

# Прогноз на тестовій вибірці
y_pred = model.predict(X_test)

# Метрики
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)

print(f"MAE: {mae}, RMSE: {rmse}, R^2: {r2}")
```

Для вибору найкращого алгоритму для конкретного випадку було порівняно XGBoost із іншими моделями, наприклад, LightGBM. Найточніша модель із

мінімальними значеннями MAE та RMSE обирається для результатами прогнозування.

Результат навчання. Після завершення навчання отримана модель здатна аналізувати вхідні дані та формувати прогнози попиту з урахуванням усіх важливих факторів. Модель інтегрується в систему прогнозування, що дозволяє використовувати її для реального бізнесу. Наступний етап — прогнозування попиту на нові дані.

3.2 Етап прогнозування

Етап прогнозування передбачає використання навченої моделі для оцінки попиту на товари на основі нових даних. Цей етап демонструє, як модель працює з реальними вхідними параметрами, формуючи прогнозовані значення продажів для кожного товару.

Вхідні дані. Модель отримує дані, підготовлені у тому ж форматі, що використовувався під час навчання:

- нормалізовані значення історичних продажів;
- сезонні ознаки (день тижня, місяць, святковий день);
- маркетингові фактори (активність акцій, знижки);
- зовнішні чинники (погода, тренди у соціальних мережах).

Приклад вхідних даних подано в таблиці 3.1.

Таблиця 3.1 – Приклад вхідних даних для моделі прогнозування попиту на товари ритейлу

Product ID	Sales Norm	Discount Norm	Day of Week	Is Weekend	Moving Avg (7 days)	Marketing Active
1	2	3	4	5	6	7
P001	0.8	0.5	2	0	0.75	1
P002	0.6	0.7	5	1	0.65	0

В таблиці наведено приклад даних по двом товарам (P001 і P002) стовпець 1, продана кількість яких, аналізується для прогнозу. В таблиці представлено результати продажу за один робочий день. Це стовпець 2, «Sales», це дані, що отримані с POS-терміналів магазину. В стовпці дані нормалізовано, за алгоритмом прогнозування, що поданий раніше, тому в назві стовпця присутня позначка «Norm», крім того модель бустінгу краще працює саме з нормалізованими даними. Наступні стовпці 3 – 7 по суті формують контекст і являють собою мета-інформацію для більш складного і витонченого аналізу. Стовпець 3 показує, чи була у той день на товар знижка, дані також нормалізовані. Стовпці 4 і 5 показують який то був день тижня, або вихідні. Стовпець 6 наводить значення ковзаної середньої для цього товару за попередні 7 днів, що також є додатковою інформацією для аналізу і порівняння. Останній стовпець 7, показує чи був в той день якийсь маркетинговий захід до цього товару.

Формування прогнозів. Навчена модель отримує підготовлені дані й обчислює прогнозований попит. Кожен рядок даних обробляється окремо, а результат — прогнозована кількість продажів на заданий період (наприклад, день або тиждень).

Код для прогнозування:

Нові дані для прогнозування

```
new_data = pd.DataFrame({
    'Sales Norm': [0.8, 0.6],
    'Discount Norm': [0.5, 0.7],
    'Day of Week': [2, 5],
    'Is Weekend': [0, 1],
    'Moving Avg (7 days)': [0.75, 0.65],
    'Marketing Active': [1, 0]
})
```

Прогноз попиту

```
predicted_sales = model.predict(new_data)
new_data['Predicted Sales'] = predicted_sales
print(new_data)
```

Наступний етап, це візуалізація результатів. Щоб оцінити якість прогнозування, створюється графік, який порівнює прогнозовані та фактичні значення. Це дозволяє візуально визначити, наскільки добре модель передбачає попит.

Приклад побудови графіка:

```
import matplotlib.pyplot as plt
# Фактичні та прогнозовані значення
actual_sales = [100, 80, 120, 90]
predicted_sales = [105, 85, 115, 92]
# Побудова графіка
plt.plot(range(len(actual_sales)), actual_sales, label='Фактичні продажі',
marker='o')
plt.plot(range(len(predicted_sales)), predicted_sales, label='Прогнозовані
продажі', marker='x')
plt.xlabel('Час (дні)')
plt.ylabel('Кількість продажів')
plt.legend()
plt.title('Порівняння фактичних і прогнозованих продажів')
plt.show()
```

Результат прогнозування у вигляді графіку подано на рисунку 3.1.

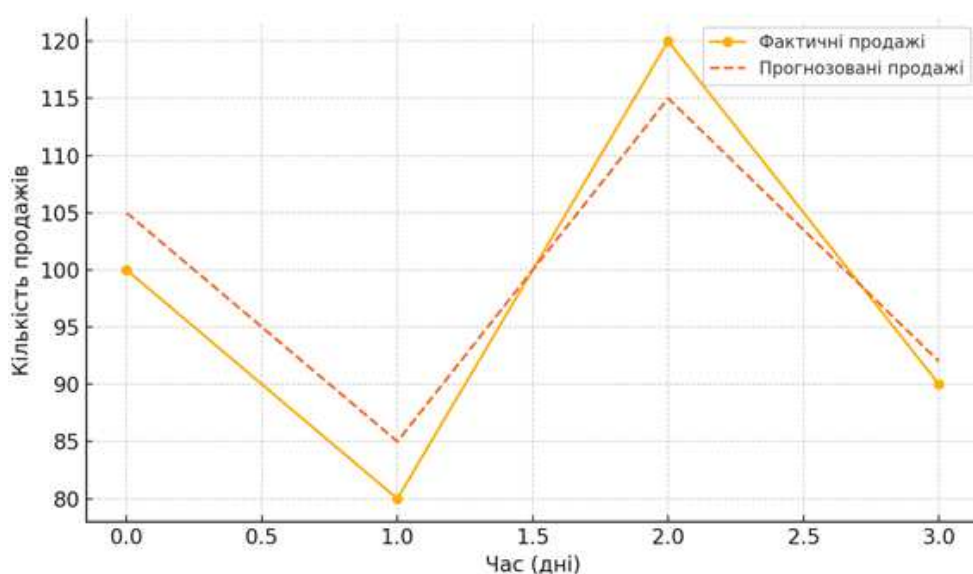


Рис. 3.1 – Результат прогнозування кількості продажів у порівнянні з фактичними ретроспективними даними для продукту Р001

На рисунку 3.1 наведена кількість продажу одного товару за три робочих дні: 1.0; 2.0; 3.0, та за попередній до аналізу день — 0.0. По вісі ОУ подана кількість проданого по факту товару в його одиницях, та прогнозне значення, що отримане в результаті застосування моделі прогнозування.

Аналіз прогнозу. Після отримання прогнозів виконується аналіз їхньої точності за допомогою тестової вибірки. Модель може показати хороші результати для прогнозування популярних товарів із постійним попитом або для сезонних або маркетингових акцій.

Однак складнощі можуть виникнути з новими товарами, для яких відсутня достатня кількість історичних даних.

Приклад результатів. Для продукту Р001 прогнозована кількість продажів становить 120 одиниць, тоді як фактичний попит дорівнює 125 одиниць. Для продукту Р002 прогноз — 95 одиниць, а фактичний попит — 90 одиниць. Середнє відхилення становить близько 5%, що є прийнятним для ритейлу.

Використання прогнозів у бізнесі. Отримані прогнози інтегруються в систему управління запасами. Наприклад:

1. Прогноз використовується для автоматизації замовлень товарів у постачальників.

2. Система визначає, які товари потребують додаткових акцій чи реклами для підвищення продажів.

3. Можливість ідентифікації товарів із потенційним надлишком для перерозподілу в інші точки продажів.

Етап прогнозування демонструє, як навчена модель ефективно аналізує нові дані та допомагає бізнесу оптимізувати управління запасами.

Приклад роботи моделі.

Вхідні дані. Для демонстрації роботи моделі використовуються дані, що відображають реальні фактори, які впливають на попит. Для кожного товару наведено нормалізовані значення продажів, знижок, днів тижня, показники сезонності та маркетингової активності (таблиця 3.1). В таблиці 3.2 до даних ДатаСету додано кількісні значення фактично проданих одиниць товару (стовпець 9) та прогнозовану кількість продажу цього товару (стовпець 8), обидва значення також за один робочий день продажу. В стовпці 10 приведено відхилення між прогнозованим та фактичним (ретроспективним) значенням.

Таблиця 3.2 – Приклад вхідних даних для моделі прогнозування попиту на товари ритейлу

Product ID	Sales Norm	Discount Norm	Day of Week	Is Weekend	Moving Avg (7 days)	Marketing Active	Predicted Sales	Actual Sales	Deviation (%)
1	2	3	4	5	6	7	8	9	10
P001	0.8	0.5	2	0	0.75	1	120	125	4.00%
P002	0.6	0.7	5	1	0.65	0	95	90	-5.56%
P003	0.7	0.6	3	0	0.70	1	110	115	4.35%
P004	0.9	0.4	6	1	0.85	0	135	130	-3.85%

Модель прогнозує попит на основі вхідних факторів. Наприклад, для товару P001 передбачувані продажі становлять 120 одиниць, тоді як фактичний попит дорівнює 125 одиниць. Відхилення прогнозу склало 4%, що є прийнятним для більшості бізнес-сценаріїв.

Візуалізація даних.

На графіку (рисунок 3.2) порівняно прогнозовані та фактичні продажі, що дозволяє оцінити, наскільки модель точно працює для кожного товару.

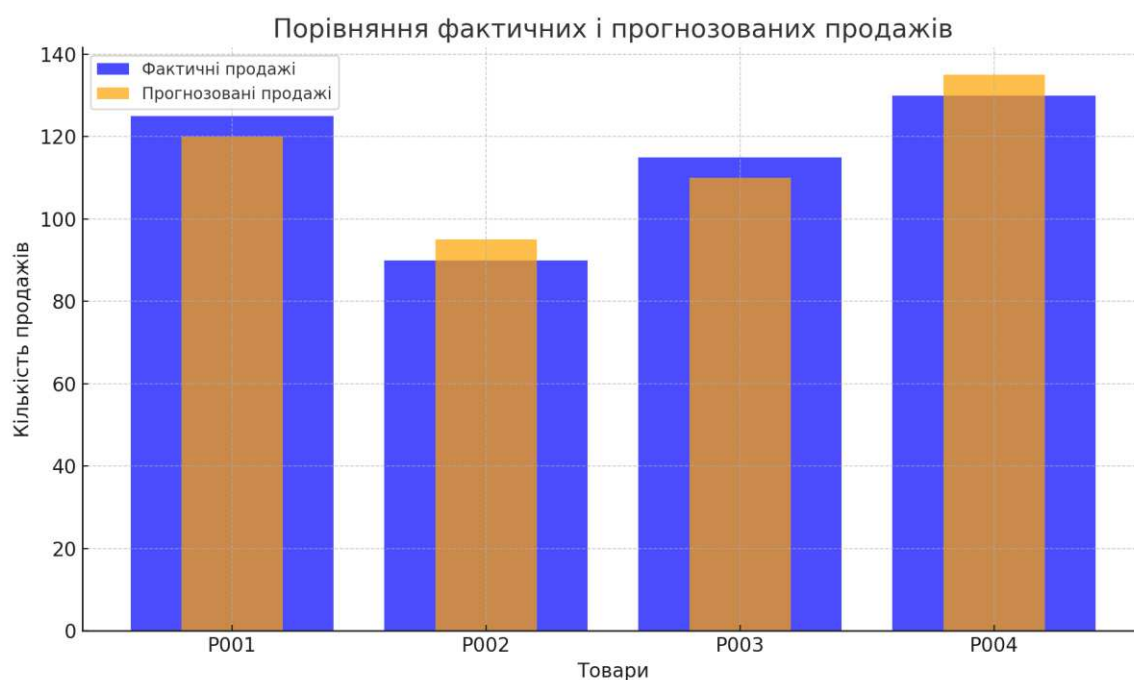


Рис. 3. 2 – Кількість фактичного проданого товару та прогноз продажу для 4 товарів за один день

3.3 Аналіз результатів

Головна характеристика моделі прогнозування це її точність. На основі тестових даних були обчислені ключові метрики:

- середня абсолютна похибка (MAE): 4.45 одиниць;
- середньоквадратична похибка (RMSE): 5.02 одиниць;

– коефіцієнт детермінації (r^2): 0.96, що свідчить про високу точність моделі.

Сильні сторони моделі.

1. Висока точність у прогнозуванні для товарів із постійним або передбачуваним попитом.
2. Ефективне врахування сезонних факторів і маркетингових акцій.
3. Гнучкість у роботі з різнорідними даними (числовими, категоріальними).

Обмеження моделі.

1. Труднощі з прогнозуванням для нових товарів, де відсутні історичні дані.
2. Складність обліку нестандартних зовнішніх чинників, таких як раптові тренди чи зміни попиту.

Рекомендації для практичного використання.

Можна рекомендувати інтегрувати модель в систему управління запасами. Прогнози можуть бути інтегровані в ERP-систему для автоматизації замовлень постачальникам.

Також можна інтегрувати в маркетингові кампанії, де прогнози будуть використані для товарів, що слабо продаються і потребують додаткових акцій чи реклами.

Слід зазначити що модель потребуватиме постійної адаптації та вдосконалення, через постійне оновлення даних для моделі (нові фактори, історичні продажі, тощо), через розширення даних для врахування більшої кількості зовнішніх чинників, таких як соціальні тренди чи погодні умови.

Можливо буде потрібна інтеграція додаткових алгоритмів для роботи з новими товарами.

Висновки до розділу 3

У третьому розділі виконано розробку моделі прогнозування попиту на товари ритейлу та перевірку її ефективності. Проведено технічну реалізацію моделі через етапи обробки даних, навчання моделі, прогнозування та аналіз отриманих результатів. У підрозділі 3.1 описано процес підготовки даних для прогнозування, що охоплює їх очищення, нормалізацію та трансформацію у формат, придатний для роботи з алгоритмами машинного навчання. Зокрема, для числових даних виконана нормалізація, а категоріальні ознаки трансформувалися методом кодування (one-hot encoding).

Навчання моделі проводилося із застосуванням алгоритму градієнтного бустингу XGBoost. Цей алгоритм дозволяє враховувати як числові, так і категоріальні дані, працювати з великою кількістю ознак і забезпечувати високу точність навіть за умов складних залежностей між змінними. Для навчання моделі було використано розділення даних на навчальну, валідаційну та тестову вибірки в співвідношенні 70:15:15, що дозволило уникнути перенавчання та об'єктивно оцінити продуктивність алгоритму.

У підрозділі 3.2 було здійснено етап прогнозування. На основі побудованої моделі отримано прогноз попиту на тестовій вибірці. Для оцінки точності прогнозів було розраховано ключові метрики: середня абсолютна похибка (MAE) становила 4,45 одиниць, середньоквадратична похибка (RMSE) — 5,02 одиниць, а коефіцієнт детермінації (R^2) склав 0,96. Ці результати свідчать про високу точність моделі, особливо для товарів із стабільним або передбачуваним попитом.

У підрозділі 3.3 проведено аналіз результатів моделі. До сильних сторін віднесено здатність моделі ефективно працювати з різнорідними даними, враховувати сезонні фактори та вплив маркетингових акцій. Модель продемонструвала гнучкість та точність для товарів із усталеним попитом. Серед обмежень було відзначено труднощі з прогнозуванням для нових товарів, які не

мають історичних даних, а також з обліком нестандартних зовнішніх чинників, таких як раптові зміни трендів або впливи надзвичайних подій.

У 3 розділі запропоновано рекомендації щодо практичного використання моделі. Її інтеграція в систему управління запасами дозволить автоматизувати процеси формування замовлень постачальникам на основі прогнозів. Додатково, модель може бути використана у маркетингових кампаніях для ідентифікації товарів із низьким попитом, які потребують додаткової реклами чи проведення акцій. Зазначено, що модель потребує регулярного оновлення даних та адаптації до нових факторів, таких як соціальні тренди чи погодні умови. Розширення моделі може включати додаткові алгоритми для роботи з новими товарами, зокрема ті, які дозволяють працювати за умов відсутності історичних даних.

Таким чином, у третьому розділі підтверджено ефективність запропонованого підходу до прогнозування попиту в ритейлі. Результати показали, що модель на основі XGBoost є високоточним і гнучким інструментом для вирішення задач прогнозування, що створює умови для її успішного використання в реальних бізнес-процесах.

РОЗДІЛ 4. РОЗРАХУНОК ЕКОНОМІЧНОЇ ЕФЕКТИВНОСТІ ВПРОВАДЖЕННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ПРОГНОЗУВАННЯ

Метою розрахунку є оцінка загальних витрат на розробку ІС, а також визначення потенційної економічної вигоди, термінів окупності та рентабельності проєкту.

Оцінки виконано в гривнях. Загальна вартість проєкту приблизно оцінюється в 20 мільйонів гривень, що включає заробітну плату, витрати на інфраструктуру та навчання.

Оцінка собівартості розробки ІС.

Собівартість розробки складається з витрат на оплату праці персоналу, інфраструктуру та навчання. Ці витрати визначають загальну базову вартість створення проєкту.

1. Витрати на оплату праці:

$$R_salary = (Salary_programmers \times N_programmers + Salary_managers \times N_managers + Salary_top \times N_top) \times Duration,$$

де: *Salary_programmers* – середня місячна зарплата одного програміста (35000 грн);

N_programmers – кількість програмістів (10 осіб);

Salary_managers – середня місячна зарплата одного менеджера (25000 грн);

N_managers – кількість менеджерів (3 особи);

Salary_top – середня місячна зарплата одного топ-менеджера (80000 грн);

N_top – кількість топ-менеджерів (1 особа);

Duration – тривалість проєкту (12 місяців).

$$R_salary = (35000 \times 10 + 25000 \times 3 + 80000 \times 1) \times 12 = 5880000 \text{ грн.}$$

2. Витрати на інфраструктуру:

$$R_infrastructure = Equipment_cost + Cloud_services,$$

де: *Equipment_cost* – вартість обладнання (1000000 грн);

Cloud_services – вартість хмарних сервісів на рік (500000 грн).

$$R_infrastructure = 1000000 + 500000 = 1500000 \text{ грн.}$$

3. Витрати на навчання персоналу:

$$R_training = N_staff \times Training_cost,$$

де: *N_staff* – кількість співробітників для навчання (20 осіб);

Training_cost – вартість навчання одного співробітника (10000 грн).

$$R_training = 20 \times 10000 = 200000 \text{ грн.}$$

4. Загальна собівартість розробки:

$$R_total = R_salary + R_infrastructure + R_training,$$

$$R_total = 5880000 + 1500000 + 200000 = 7580000 \text{ грн.}$$

Оцінка економічної ефективності впровадження ІС. Економічна ефективність включає оцінку потенційного збільшення доходів та зменшення витрат завдяки впровадженню ІС.

1. Підвищення доходів:

$$\Delta Revenue = Sales_increase \times Margin,$$

де: *Sales_increase* – приріст продажів завдяки точнішому прогнозуванню (10000000 грн);

Margin – середня маржа товарів (20% або 0,2).

$$\Delta Revenue = 10000000 \times 0,2 = 2000000 \text{ грн.}$$

2. Економія витрат:

$$\Delta Cost_savings = Inventory_savings + Logistics_savings,$$

де: *Inventory_savings* – економія завдяки оптимізації запасів (500,000 грн);

Logistics_savings – економія на транспортуванні (300000 грн).

$$\Delta Cost_savings = 500000 + 300000 = 800000 \text{ грн.}$$

3. Загальний економічний ефект:

$$Net_gain = \Delta Revenue + \Delta Cost_savings,$$

$$Net_gain = 2000000 + 800000 = 2800000 \text{ грн.}$$

Оцінка терміну окупності

$$Payback_period = R_total / (Net_gain / 12),$$

де: *R_total* – загальна собівартість розробки (7580000 грн);

Net_gain – загальний економічний ефект на рік (2,800,000 грн).

$$Payback_period = 7580000 / (2800000 / 12) \approx 32,5 \text{ місяців (приблизно 2,7 роки)}$$

без врахування дисконтування грошових потоків.

Висновки до розділу 4

В четвертому розділі проведено обґрунтування економічної ефективності впровадження інформаційної системи прогнозування товарів ритейлу. Обґрунтування проведено методом попередньої оцінки економічної доцільності. Цей підхід базується на використанні агрегованих даних та орієнтовних показників, які дозволяють зробити попередній висновок про доцільність проєкту за умов обмеженої інформації.

Проведена оцінка економічного обґрунтування проєкту показала, що очікувана окупність проєкту становить орієнтовно 3 роки, що є прийнятним для крупних проєктів такого типу, а сам проєкт зі створення ІС прогнозування товарів ритейлу, може бути рекомендований до впровадження.

5 ОХОРОНА ПРАЦІ

Питання охорони праці людини необхідно вирішувати на всіх стадіях трудового процесу незалежно від виду професійної діяльності.

Забезпечення безпечних і здорових умов праці в значній мірі залежить від правильної оцінки небезпечних, шкідливих виробничих факторів. Однакові по складності зміни в організмі людини можуть бути викликані різними причинами. Це можуть бути фактори виробничого середовища, надмірне фізичне і розумове навантаження, нервово-емоційна напруга, а також різне сполучення цих причин.

У даному розділі вирішується питання охорони праці програміста на стадії розробки ним інформаційної системи [24, 25].

5.1 Аналіз умов праці на робочому місці програміста

Організація робочого місця повинна забезпечувати відповідність усіх елементів робочого місця та їх розташування вимогам ДСТУ 8604:2015 «Дизайн і ергономіка. Робоче місце для виконання робіт у положенні сидячи. Загальні ергономічні вимоги» та ДБН В.2.2-27:2010 «Будинки і споруди. Будинки адміністративного та побутового призначення».

Приміщення, в якому працює програміст, має загальну площу 44 м², висоту стелі 3,8 м. У приміщенні знаходиться 6 робочих місць з ПК. Кожне робоче місце обладнане робочим столом площею 1,2 м²., стільцем та персональним комп'ютером.

У приміщенні лабораторії знаходиться чотири вікна розміром 2,3×1,8 м, а джерелом штучного освітлення є 12 світильників з дзеркальним растром Light Case 3 x TC-L (36Вт). Лабораторія обладнана кондиціонером.

Аналіз умов праці показує, що у приміщенні лабораторії на програміста можуть негативно впливати наступні шкідливі та небезпечні чинники:

фізичні:

- підвищений рівень електромагнітного випромінювання радіочастотного діапазону;
- підвищені або понижені температура, відносна вологість та швидкість руху повітря робочої зони; підвищена запиленість і загазованість повітря робочої зони; підвищений або понижений вміст позитивних і негативних аероіонів у повітрі робочої зони;
- недостатня освітленість робочої зони; підвищена яскравість світла; понижена контрастність між об'єктом і фоном; прямий та віддзеркалений відблиск; підвищена пульсація світлового потоку.

хімічні:

- підвищений вміст у повітрі робочої зони озону та аміаку;
- психофізіологічні:
- надмірні статичні та динамічні навантаження;
- розумове перенавантаження;
- перенавантаження аналізаторів;
- монотонність праці; надмірні емоційні навантаження;
- нераціональна організація робочого місця;

біологічні:

- підвищений вміст мікроорганізмів у повітрі робочої зони.

Кожний із наведених чинників може негативно впливати на стан здоров'я розробника (користувача), а їх сумісна дія підсилити цей вплив. З метою виключення або зменшення їх негативного впливу виконаний їх аналіз та визначені заходи по їх оптимізації, результати яких наведені нижче.

Нормування параметрів проводиться в залежності від періоду року та категорії важкості виконуваних робіт. Для постійних робочих місць, якими є робочі місця програміста, встановлені оптимальні параметри мікроклімату, а за неможливості їх дотримання використовують допустимі параметри. Робота

оператора ПК за енерговитратами відноситься до категорії легких робіт Ia, Ib.

Виміряні за допомогою приладів температура та вологість у лабораторії відповідають вказаним у таблиці для теплого періоду року. Слід зазначити, що для нормалізації параметрів мікроклімату слід використовувати у приміщеннях кондиціювання повітря, або забезпечити подачу свіжого повітря системами вентиляції.

Розташовані у приміщенні 6 ПК є джерелами тепловиділень, крім того для підтримання у приміщенні в холодний період року оптимальних параметрів мікроклімату використовуються нагріті поверхні опалювальної системи. Нормованим показником ІЧВ являється гранично допустима густина потоку енергії $I_{г.д}$, Вт/м², яка встановлюється в залежності від площі опромінюваної поверхні тіла людини ($S_{опр}$). Нормовані рівні складають:

$I_{г.д} = 35 \text{ Вт/м}^2$ за $S_{опр} > 50\%$; $I_{г.д} = 70 \text{ Вт/м}^2$ за $S_{опр} \sim 25-50\%$; $I_{г.д} = 100 \text{ Вт/м}^2$ за $S_{опр} < 25\%$.

Нормованим параметром природного освітлення згідно ДБН В.2.5-28:2018 є коефіцієнт природного освітлення (КПО). КПО встановлюється в залежності від розряду виконуваних зорових робіт. Робота програміста прикладного відноситься до робіт середньої точності (IV розряд зорових робіт, мінімальний розмір об'єкту розрізнення складає 0,5-1,0мм), для яких при використанні бокового освітлення КПО = 1,5%. Для штучного освітлення нормованим параметром виступає $E_{мін}$ – мінімальний рівень освітленості, та $K_{п}$ – коефіцієнт пульсації світлового потоку, який не повинний бути більшим ніж 20%. Мінімальна освітленість встановлюється в залежності від розряду виконуваних зорових робіт. Для IV розряду зорових робіт вона складає 300-500 лк.

Причинами виникнення небезпечних і шкідливих факторів є:

- відсутність чи недостатність природного освітлення;
- недостатня освітленість робочої зони;
- прямий чи відбитий блиск.

Причини виникнення вищевказаних факторів:

- відсутність чи недостатні розміри вікон;
- відсутність чи недостатня кількість світильників;
- недоліки освітлення і/або неправильне розташування дисплея.

У лабораторії знаходяться 6 робочих столів, розміщені подалі від вікон і так, щоб вікна знаходилися збоку від працюючих і природне світло падало зліва.

Найменший об'єкт розрізнення залежить від класу монітора.

Технічні характеристики монітора LG 29UM59-P: тип – ЖК-монітор, широкоформатний; діагональ – 29"; точність – 2560x1080 (21:9); тип ЖК-матриці – TFT IPS; яскравість – 250 кд/м²; максимальна кількість кольорів – 16777216 (24 біт); підсвічування – W-LED; час відгуку – 5 мс.; область огляду – по горизонталі – 178°; по вертикалі – 178°; динамічна контрастність – 5000000 : 1; статична контрастність – 1000 : 1; крок точки по горизонталі – 0.263 мм.; мін. частота рядків – 30 кГц.; максимальна частота оновлення кадрів: 75 Гц.; покриття sRGB – 99%; енергоспоживання – клас С.

Нормування природного освітлення проводиться згідно з В.2.5-28:2018 «Природне та штучне освітлення» (таблиця 5.1).

Таблиця 5.1 – Нормування природного освітлення

Найменший розмір об'єкта роздільної здатності	Розряд зорової роботи	Коефіцієнт природного висвітлення при бічному освітленні
0,26×0,25	5	0,6

При тривалій роботі з ПК можна отримати так званий КЗС – комп'ютерний зоровий синдром. Цей синдром супроводжується різким болем очей, при цьому спостерігається їхнє почервоніння і сухість. Нерідко трапляються головні болі при наборі тексту, з'являється швидка стомлюваність.

Як було вказано вище, в лабораторії знаходиться 6 робочих місць з ПК. Крім

того поряд працює периферійна техніка. Таким чином у приміщенні мають місце шуми механічного і аеродинамічного походження, широкосмугові із аперіодичним підсиленням при роботі принтерів. Допустимий еквівалентний рівень шуму для робочого місця програміста складає 50 дБА.

Розрахуємо середній рівень шуму на робочому місці оператора при роботі всієї вказаної техніки

Рівень шуму, що виникає від декількох некогерентних джерел, що працюють одночасно, підраховується на підставі принципу енергетичного підсумовування рівня інтенсивності окремих джерел:

$$L_{сер} = 10 \lg \sum_{i=1}^n 10^{0,1L_i}$$

де L_i – рівень звукового тиску i -го джерела шуму; n – кількість джерел шуму.

Підставивши значення рівня звукового тиску для кожного виду устаткування у формулу, отримаємо:

$$L = 10 \lg (104,5 + 104,5 + 105,5 + 105,0) = 44,2 \text{ дБ.}$$

За наявності декількох джерел шуму з однаковим рівнем інтенсивності L_i загальний рівень шуму визначають за формулою: $L = L_i + 10 \lg n$. У нашому випадку таких джерел тринадцять, отже загальний рівень шуму буде визначатися так:

$$L = 44,2 + 10 \lg 6 = 51,98 \text{ дБ.}$$

Розраховане значення середнього рівня шуму перевищує гранично допустимий рівень шуму для робочого місця програміста прикладного, тобто слід передбачити заходи по зниженню рівня шуму.

Оцінка напруженості праці здійснювалась на підставі обліку всіх наявних значущих показників, які можуть перевищувати нормативні рівні.

Розподіл функцій за ступенем складності завдання – належить до класу 2 (обробка, виконання завдання та його перевірка). Характер виконуваної роботи – належить до класу 2 (робота за встановленим графіком з можливим його коректуванням у ході діяльності). Навантаження на зоровий аналізатор (при

відстані від очей працюючого до об'єкта розрізнення не більше 0,5 м), при тривалості зосередженого спостереження (% часу зміни) – належить до класу 2 (5,0-1,1 мм більше 50 % часу; 1,0-0,3 мм до 50 % часу; менше 0,3 мм до 25 %).

Спостереження за екранами відеотерміналів (годин на зміну) – належить до класу 3.2 (більше 4 годин). Монотонність праці. Кількість елементів (приймів, необхідних для реалізації простого завдання або в операціях, які повторюються багаторазово) – належить до класу 3.1 (5-3 прийоми). Режим праці (фактична тривалість робочого дня (год.) – належить до класу 1 (6-7 годин). Наявність регламентованих перерв та їх тривалість – належить до класу 2 (перерви регламентовані, недостатньої тривалості: від 3 % до 7 % часу зміни). Отже робоче місце за показниками напруженості трудового процесу відноситься до класу 3.1 - Шкідливий (напружена праця).

5.2 Розробка заходів з охорони праці

Організація робочого місця програміста прикладного передбачає: правильне розташування робочого місця у виробничому приміщенні; вибір виробничих меблів; раціональне компонування комп'ютерного обладнання на робочому місці; урахування характеру та особливостей трудової діяльності.

Конструкція робочого місця користувача відео терміналу забезпечує підтримання оптимальної робочої пози з такими ергономічними характеристиками: ступні ніг - на підлозі або на підставці для ніг; стегна – в горизонтальній площині; передпліччя - вертикально; лікті - під кутом 70°–90°, до вертикальної площини; зап'ястя зігнуті під кутом не більше 20°, відносно горизонтальної площини, нахил голови – 15°–20°, відносно вертикальної площини.

Все вказане обладнання розміщується на основному робочому столі з лівого боку.

Висота робочої поверхні столу для відео терміналу 680–800мм, а ширина –

забезпечує можливість виконання операцій в зоні досяжності моторного поля. Розміри столу: висота – 725мм, ширина – 600–1400мм, глибина – 800мм–1000мм. Робочий стіл для відео терміналу обладнаний підставкою для ніг шириною 400мм з можливістю регулювання по висоті. Підставка має рифлену поверхню та бортик на передньому краї заввишки 10 мм.

Робоче сидіння користувача відео терміналу та персональної ПК має такі основні елементи: сидіння, спинку та знімні підлокітники. Робоче сидіння є підйомно-поворотним, регулюється за висотою, кутом нахилу сидіння та спинки. Поверхня сидіння є плоскою, передній край -заокруглений.

Екран відео термінала розташовуються на оптимальній відстані від очей користувача (800 мм). Розташування екрану відео терміналу забезпечує зручність зорового спостереження у вертикальній площині під кутом $+30^\circ$ від лінії зору працівника.

Клавіатура розміщена на спеціальній, регульованій за висотою, робочій поверхні окремо від столу на відстані 300мм від краю, ближчого до працівника. Кут нахилу клавіатури складає 10° .

Колір є найбільш ефективним засобом естетичного рівня виробничого інтер'єру. За допомогою кольору вирішуються питання: забезпечення психофізіологічного комфорту; емоційно-естетичний вплив на працівника.

На робочому місці стіни фарбують у світлий колір. Що сприяє працездатності працівника, зменшує втому очей.

Для створення й автоматичної підтримки в лабораторії незалежно від зовнішніх умов оптимальних значень температури, вологості, чистоти і швидкості руху повітря, у холодний час року використовується водяне опалення, у теплий час року застосовується кондиціонування повітря.

Приміщення обладнано кондиціонером, що має наступні основні технічні характеристики:

Тип: настінна спліт-система

Режим роботи: охолодження/обігрів;

Рекомендована площа приміщення: 25м²;

Циркуляція повітря: 684 м³ / год.;

Рівень шуму внутрішнього блоку, (min/max): 27 Дб / 37 Дб;

Споживана потужність, охолодження: 0.49 кВт.;

Споживана потужність, обігрів: 0.63 кВт.;

Клас енергоефективності, охолодження (EER): A 5.12 Вт / Вт;

Клас енергоефективності, обігрів (COP): A 5.1 Вт / Вт;

Обробка повітря: фільтр Toshiba IAQ; іонізатор повітря; плазмовий фільтр очищення повітря кондиціонерів Toshiba з іонами срібла.

Приміщення обладнане склопакетами KBE Classica, які приймають значну участь в регуляції мікрокліматичних показників. Характеристики склопакетів вказано нижче:

Кількість камер – 1 камера;

Товщина зовнішньої стінки – 2,8 мм.;

Коефіцієнт опору теплопередачі – 0,87 м.кв. °C / Вт.;

стійкість до морозів – до -60°C;

Екологічна рецептура greenline (без свинцю).

У якості заходів зі зниження шуму можна запропонувати:

- облицювання стелі і стін звукопоглинаючим матеріалом (знижують шум на 6-8 дб);
- екранування робочого місця (постановкою перегородок, діафрагм);
- установка в комп'ютерних приміщеннях устаткування, що робить мінімальний шум;
- раціональне планування приміщення.

Для зменшення шуму в аналізованій лабораторії пропоную використовувати малошумні принтери та обрати звукопоглинальне покриття. У приміщенні

довжиною 8 м, шириною 5,5 м та висотою 3,8 м встановлено 6 ПК .

В розрахунковій точці, віддаленій від найближчих ПК на 4 м, було виміряно спектральні рівні звукового тиску (таблиця 5.2)

Таблиця 5.2 – Рівні звукового тиску в розрахунковій точці

Середня частота, <i>Гц</i>	63	125	250	500	1000	2000	4000	8000
Рівень звукового тиску, <i>дБ</i>	75	68	64	59	55	52	50	48

Для виробничих приміщень, допустимий рівень звуку становить $L_{\text{доп}} = 50$ дБА.

При декількох джерелах шуму потрібне зниження шуму для кожного з них (дБА) у розрахунковій точці визначають за формулою:

$$\Delta Li = (Li - L_{\text{доп}}) + 10 \lg n.$$

$$\Delta Li = (55,33 - 50) + 10 \lg 6 = 13,11 \text{ дБ}.$$

Об'єм приміщення рівний $V = 167,2 \text{ м}^3$

Звукопоглинальна обробка зроблена на всій внутрішній поверхні захищаючих конструкцій за винятком віконних і дверних отворів і підлоги приміщення, тобто $S_{\text{обл}} = 190,6 - 45 - 44 = 101,6 \text{ м}^2$.

То можуть бути звукопоглиналі мінераловатні плити.

Висновки до розділу 5

В п'ятому розділі розглянуто питання з охорони праці на робочому місці програміста, який приймає участь у ІТ-проекті зі створення ІС прогнозування попиту на товари в ритейлі. Проаналізовано умови праці та запропоновано заходи з покращення цих умов.

6 ОХОРОНА НАВКОЛИШНЬОГО СЕРЕДОВИЩА

Охорона навколишнього середовища є комплексом заходів, спрямованих на збереження, відновлення та раціональне використання природних ресурсів, з метою запобігання їх виснаженню та деградації. Вона охоплює впровадження екологічних стандартів, зменшення викидів забруднюючих речовин, перехід до сталого споживання енергії та матеріалів, а також збереження екологічного балансу. У контексті ритейлу охорона навколишнього середовища передбачає інтеграцію екологічних ініціатив у бізнес-процеси, зменшення негативного впливу на екосистеми та впровадження рішень, що сприяють сталому розвитку [25, 26]. Екологічна відповідальність стала не лише моральним обов'язком, але й бізнес-необхідністю, яка підвищує довіру клієнтів і підтримує сталий розвиток корпорацій.

Корпорації ритейлу можуть забруднювати навколишнє середовище через кілька основних механізмів. Одним із найбільших джерел забруднення є викиди вуглекислого газу під час транспортування товарів, що сприяє зміні клімату. Логістика ритейлу часто включає перевезення великих обсягів продукції між складами, магазинами та кінцевими споживачами, що супроводжується значним споживанням пального та високими викидами CO₂. Другим джерелом є утилізація відходів, зокрема пакувальних матеріалів, які часто не підлягають вторинній переробці. Це призводить до забруднення ґрунтів і водних ресурсів, а також збільшення кількості відходів, що потрапляють на сміттєзвалища. Третє джерело пов'язане зі споживанням енергії для роботи магазинів і складів, яке включає кондиціонування, освітлення, холодильне обладнання, що зазвичай працює на електроенергії з викопних джерел. Четвертим є надмірне споживання води для обслуговування магазинів, складів та інших об'єктів інфраструктури. П'яте джерело – це виробництво товарів, яке часто залежить від постачальників, що не дотримуються екологічних стандартів, використовуючи ресурси неефективно та

завдаючи шкоди екосистемам.

Для зменшення впливу на навколишнє середовище корпорації ритейлу повинні впроваджувати екологічно орієнтовані стратегії, що дозволять залишатися конкурентоспроможними та прибутковими. Перш за все, слід оптимізувати транспортну логістику, впроваджуючи використання екологічно чистого транспорту, такого як електровантажівки або гібридні автомобілі. Це дозволить зменшити викиди CO₂ під час перевезень. По-друге, корпораціям варто інвестувати в енергозберігаючі технології, такі як світлодіодне освітлення, енергоефективні холодильники та системи управління енергоспоживанням, що знижують операційні витрати та екологічний слід. Третій напрямок – це зменшення обсягу пакувальних матеріалів, використання вторинної сировини та впровадження програм переробки відходів. Четвертий – співпраця з постачальниками, які дотримуються принципів сталого розвитку, використовуючи екологічні стандарти для відбору партнерів. П'ятим напрямком є впровадження програм навчання персоналу та комунікації з клієнтами щодо екологічних ініціатив.

Системи прогнозування попиту можуть значно знизити шкідливі викиди завдяки оптимізації поставок. Правильний прогноз дозволяє забезпечити наявність потрібної кількості товарів у кожному регіоні, уникнути перевезень товарів між складами через неправильне планування та зменшити ризик дефіциту або надлишку товарів. Це знижує кількість транспортних операцій, скорочує споживання пального та зменшує викиди CO₂. Наприклад, якщо система точно передбачає попит, товари доставляються безпосередньо до місць реалізації, уникаючи зайвих логістичних витрат. Також система дозволяє аналізувати сезонність попиту, маркетингові акції та інші фактори, які впливають на продажі, що дозволяє досягти більшої точності в плануванні.

Приклад неправильного прогнозу демонструє важливість точного планування. Уявімо ситуацію, коли в регіоні виникає дефіцит товарів, і корпорація змушена доставляти 10 тонн продукції зі складу, розташованого на відстані 300 км.

Така позапланова операція не лише збільшує логістичні витрати, але й спричиняє додаткові викиди CO₂. Якщо викиди для перевезення 1 кг товару на 1 км становлять 0.1 кг CO₂, то загальний обсяг викидів складе $10,000 \times 300 \times 0.1 = 300,000$ кг CO₂.

Щодо витрат, якщо транспортування 1 кг товару на 1 км коштує 5 грн, загальна вартість позапланового перевезення складе $10,000 \times 300 \times 5 = 15,000,000$ грн. Уникнення таких витрат завдяки оптимальному прогнозуванню дозволить зекономити значні кошти та зменшити екологічний вплив.

Розглянемо ще один приклад: якщо система прогнозування дозволяє уникнути п'яти подібних перевезень за рік, економія складе $15000000 \times 5 = 75000000$ грн, а уникнуті викиди складуть $300000 \times 5 = 1500,000$ кг CO₂. Це підкреслює значущість інноваційних рішень для сталого розвитку та підвищення ефективності бізнесу.

Для наступного релізу інформаційної системи прогнозування попиту на товари ритейлу можна рекомендувати додати функціонал, що дозволить враховувати екологічні обмеження для зменшення негативного впливу на навколишнє середовище. Однією з ключових функцій може бути інтеграція модулів аналізу викидів CO₂ на всіх етапах логістики. Це дозволить прогнозувати не лише попит, але й пов'язані з ним екологічні витрати. Додатково можна впровадити оптимізацію маршрутів доставки, що враховує не лише економічні витрати, але й вплив на довкілля. Алгоритми оптимізації повинні знижувати кількість перевезень та загальний пробіг транспортних засобів.

Іншим важливим функціоналом стане включення до системи механізму аналізу життєвого циклу товарів, що враховує енергоспоживання, відходи пакування та вплив на екосистеми на етапі виробництва. Це дозволить робити прогнози з урахуванням екологічних показників конкретних категорій товарів. Система також може інтегрувати дані з екологічно орієнтованих платформ, таких як бази сертифікованих постачальників, які дотримуються принципів сталого розвитку.

Корисним доповненням буде впровадження візуалізації екологічних метрик у реальному часі, включаючи кількість потенційно зменшених викидів та економію ресурсів. Це дозволить користувачам системи оцінювати вплив своїх рішень на довкілля та знаходити баланс між екологічними обмеженнями й економічною ефективністю. Також система може генерувати рекомендації для клієнтів, наприклад, щодо зменшення пакування або вибору більш екологічно чистих товарів, що сприятиме сталому розвитку ритейлу.

Таким чином, застосування систем прогнозування попиту дозволяє уникати позапланових перевезень, оптимізувати використання ресурсів і сприяє екологічній відповідальності. Корпорації ритейлу можуть не лише знизити витрати, але й зробити внесок у збереження довкілля, підтримуючи сталий розвиток. Інтеграція екологічних принципів у бізнес-процеси приносить як економічну, так і екологічну користь, підвищує конкурентоспроможність компаній і сприяє збереженню природних ресурсів для майбутніх поколінь.

Висновки до розділу 6

В шостому розділі розглянуто питання охорони навколишнього середовища для компаній ритейлу. Показано, що їх логістична складова та відділи пакування можуть негативно впливати на довкілля через викиди транспортних засобів, що перевозять товари та через їх надмірне пакування. В розділі показано, що застосування методологія системного аналізу може бути ефективно застосована при розробці моделі та ІС прогнозування попиту на товари ритейлу, шляхом додавання додаткових критеріїв прогнозування попиту, а саме, мінімізації перевезень товарів між складами в рамках однієї корпорації. Показано, що за рахунок цього, корпорації ритейлу можуть не лише знизити витрати, але й зробити внесок у збереження довкілля, підтримуючи сталий розвиток територій.

ВИСНОВКИ

У першому розділі роботи проведено всебічний аналіз засобів прогнозування попиту на товари в ритейлі. Було досліджено предметну область, охарактеризовано основну процедуру прогнозування, виділено її ключові етапи та визначено складність. У рамках аналізу було виявлено, що для точного прогнозування необхідно враховувати широкий спектр даних, які охоплюють як внутрішні показники (історичні продажі, запаси, акції), так і зовнішні фактори (дані про постачання, логістичні операції, погодні умови). Однак ці дані часто мають різний формат, різну частоту оновлення, можуть містити пропуски й характеризуються значною амплітудою, що створює труднощі для їх інтеграції та аналізу. Також детально розглянуто типи даних, їх джерела та особливості попередньої обробки. Було встановлено, що для роботи з такими даними потрібна ефективна система, здатна автоматизувати процеси збору, очищення, трансформації та аналізу. Особливу увагу приділено алгоритмам машинного навчання, зокрема XGBoost і LightGBM, які показують високу точність та адаптивність для багатовимірних задач. Сформульована постановка задачі визначила необхідність створення системи прогнозування, що інтегрує різноманітні джерела даних, забезпечує автоматичне оновлення прогнозів і підтримує короткострокові прогнози.

У другому розділі проведено проєктування інформаційної системи, яка реалізує процес прогнозування попиту на товари в ритейлі. Було визначено ключові параметри системи, включаючи продуктивність, масштабованість та точність прогнозів. Обґрунтовано функціональні та нефункціональні вимоги, які забезпечують інтеграцію з POS-системами, API перевізників, WMS, а також підтримку роботи з великими масивами даних. Виконано моделювання прецедентів, які описують типові сценарії використання системи, такі як обробка даних, побудова прогнозів, інтеграція з ERP-системами та генерація звітів. Було обрано відповідні інструменти для реалізації системи, включаючи Python для

алгоритмів машинного навчання, PostgreSQL для зберігання даних і Tableau для візуалізації результатів. Розроблено архітектуру системи, яка охоплює модулі збору, обробки, аналізу даних, побудови моделей та інтеграції з іншими системами. Було запропоновано інтуїтивно зрозумілий інтерфейс, що полегшує доступ до функцій системи. Ефективність системи було оцінено за допомогою тестових наборів даних. Аналіз результатів підтвердив, що система відповідає визначеним вимогам і забезпечує високу точність прогнозування попиту, особливо при роботі з великими обсягами різнорідних даних.

Третій розділ присвячено розробці та тестуванню моделі прогнозування. Було реалізовано алгоритм XGBoost, що дозволив врахувати різнорідність даних і складні взаємозв'язки між ними. На етапі обробки даних виконано очищення, нормалізацію та трансформацію у формат, придатний для машинного навчання. Модель навчалася на підготовлених даних, розділених на навчальну, валідаційну та тестову вибірки. У процесі тестування були отримані ключові метрики: середня абсолютна похибка (MAE) склала 4.45 одиниць, середньоквадратична похибка (RMSE) — 5.02 одиниць, а коефіцієнт детермінації (R^2) — 0.96. Ці показники свідчать про високу точність моделі в прогнозуванні попиту, особливо для товарів із стабільним попитом або сезонними коливаннями. Однак було виявлено обмеження, пов'язані з прогнозуванням для нових товарів та впливом раптових змін попиту. Було рекомендовано інтегрувати модель у систему управління запасами, а також використовувати її результати для оптимізації маркетингових кампаній. Наголошено на необхідності постійного вдосконалення моделі шляхом розширення даних та додавання нових алгоритмів.

В четвертому розділі проведено обґрунтування економічної ефективності впровадження інформаційної системи прогнозування товарів ритейлу.

Проведена оцінка економічного обґрунтування проєкту показала, що очікувана окупність проєкту становить орієнтовно 3 роки, що є прийнятним для

крупних проектів такого типу, а сам проєкт зі створення ІС прогнозування товарів ритейлу, може бути рекомендований до впровадження.

В п'ятому розділі розглянуто питання з охорони праці на робочому місці програміста, який приймає участь у ІТ-проєкті зі створення ІС прогнозування попиту на товари в ритейлі. Проаналізовано умови праці та запропоновано заходи з покращення цих умов.

В шостому розділі розглянуто питання охорони навколишнього середовища для компаній ритейлу. Показано, що їх логістична складова та відділи пакування можуть негативно впливати на довкілля через викиди транспортних засобів, що перевозять товари та через їх надмірне пакування. В розділі показано, що застосування методології системного аналізу може бути ефективно застосована при розробці моделі та ІС прогнозування попиту на товари ритейлу, шляхом додавання додаткових критеріїв прогнозування попиту, а саме, мінімізації перевезень товарів між складами в рамках однієї корпорації. Показано, що за рахунок цього, корпорації ритейлу можуть не лише знизити витрати, але й зробити внесок у збереження довкілля, підтримуючи сталий розвиток територій.

Загальні висновки роботи свідчать про доцільність застосування методів машинного навчання для прогнозування попиту на товари в ритейлі. Розроблена система забезпечує інтеграцію різномірних джерел даних, їх обробку та аналіз із використанням сучасних алгоритмів, що дозволяє досягти високої точності прогнозів. Ефективність моделі підтверджено експериментальними результатами, що створює перспективи для її практичного впровадження в бізнес-процеси ритейлу.

Результати кваліфікаційної роботи представлені у тезах доповідей Міжнародної науково-практичної конференції та статті у фаховому журналі категорії В.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Teslenko, P., & Barskyi, S. (2024). Analysis of machine learning models for forecasting retail resources. *Technology Audit and Production Reserves*, 6/2(80). 6-10. <https://doi.org/10.15587/2706-5448.2024.315495>
2. Macas, C., Aguirre, J., Arcentales-Carrion, R., & Peña, M. (2021). Inventory management for retail companies: A literature review and current trends. 2021 Second International Conference on Information Systems and Software Technologies (ICI2ST), 71-78. <https://doi.org/10.1109/ICI2ST51859.2021.00018>.
3. Терещенко С.І., Гримайло О.В., (2023). Інноваційні форми організації роздрібної торгівлі. МНПК Цифрова трансформація та диджитал технології для сталого розвитку всіх галузей сучасної освіти, науки і практики, 281-285. URL : https://repo.btu.kharkov.ua/bitstream/123456789/29872/1/Zbi%C3%B3r_prac_3_2023-281-285.pdf
4. Барський С.Ю., Тесленко П.О. (2024) Управління ресурсами інфраструктурних проєктів. *МНПК «Інформаційні системи в управлінні проєктами та програмами»*, ХНУРЕ, 53 – 56. <https://doi.org/10.30837/ISRRM.2024.09>.
5. Anusha, S., Alok, S., & Shaik, A. (2015). Demand Forecasting for the Indian Pharmaceutical Retail: A Case Study, 3, 1-8.
6. Haque, M., Amin, M., & Miah, J. (2023). Retail Demand Forecasting: A Comparative Study for Multivariate Time Series. *ArXiv*, abs/2308.11939. <https://doi.org/10.48550/arXiv.2308.11939>.
7. Ulrich, M., Jahnke, H., Langrock, R., Pesch, R., & Senge, R. (2021). Classification-based model selection in retail demand forecasting. *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2021.05.010>.
8. Phyu, M., & Khine, M. (2023). Retail Demand Forecasting Using Sequence to Sequence Long Short-Term Memory Networks. 2023 IEEE Conference on Computer Applications (ICCA), 208-213. <https://doi.org/10.1109/ICCA51723.2023.10181450>.

9. Галузов, С. Ю., Бондарчук, А. П., Бажан, Т. О., & Корецька, В. О. (2023). Застосування методів DATA SCIENCE для прогнозування попиту в ритейлі. телекомунікаційні та інформаційні технології, (3), 58-64. DOI: 10.31673/2412-4338.2023.035965
10. Semenova, V. G., & Semenova, K. D. (2018). Використання моделей одного часового ряду для прогнозування економічних показників у сучасних умовах. Financial and credit activity problems of theory and practice, 3(26), 334-340.
11. Бойніцька С. В., Мілявський Ю. Л. (2023). Вибір та оцінка якості моделей прогнозування фінансових показників з урахуванням характеристик вхідних даних. Системні науки та інформатика: збірник доповідей II науковопрактичної конференції «Системні науки та інформатика», 4–8 грудня 2023 року, Київ. К., НН ІПСА КПІ ім. Ігоря Сікорського, 84-89.
12. Гайдур, Г. І., Гахов, С. О., & Бригинець, А. А. (2023). Виявлення мережевих аномалій з використанням алгоритмів нейронних мереж. телекомунікаційні та інформаційні технології, (1), 61-73.
13. Кобилін О., Вечірська І., Кравченко О. (2024). Порівняння нейронних мереж типу RNN ТА LSTM. Information Technology: Computer Science, Software Engineering and Cyber Security, (3), 97-107.
14. Чебан, К. В., Шпінарева, І. М., & Рудніченко, М. Д. (2023). Використання алгоритму XGBOOST для прогнозування часових рядів. Світ наукових досліджень. Випуск 23: матеріали Міжнародної конференції , 68-70.
15. Христина, Л. І. П., & Юрків, Х. (2023). Методи бустингового машинного навчання для нестационарних часових рядів. Measuring and computing devices in technological processes, (3), 19-30.
16. Pistunov, I., & Prykhodchenko, O. (2024). Прогнозування обсягу реалізації товарів торговельного підприємства. International Science Journal of Management, Economics & Finance, 3(3), 79-94.

17. Гудзовата, О. О., Костирко, В. І., & Артишук, І. В. (2019). Використання уніфікованої мови візуального моделювання UML (Unified Modeling Language) як інструменту підтримки проектування інформаційних систем. Підприємництво і торгівля, (24), 108-114.

18. Макоєдова, В. О. (2023). Моделювання функціональних вимог до інформаційної системи приймальної комісії за допомогою UML-діаграми прецедентів. Збірник тез доповідей Міжнародної наукової інтернет-конференції (випуск 75) 6-7 березня 2023 р. на сайті www.konferenciaonline.org.ua, 35-38.

19. Гороховатський В.О., Творошенко І.С. (2021). Методи інтелектуального аналізу та оброблення даних: навч. посібник. – Харків: ХНУРЕ. 92.

20. Тесленко П.О., Ворона М.В. (2024). Штучний інтелект в прогнозуванні попиту в ритейлі. Інформаційні технології: моделі, алгоритми, системи – 2024. V Всеукраїнська науково-практична інтернет конференція. URL : <https://itconf.nuos.edu.ua/2024/publications/artificial-intelligence-in-forecasting-demand-in-retail/>

21. Юрченко І. В. (2024). DataMining з використанням Python. Навчальний посібник.—Чернівці: ЧНУ. 143 с.

22. Верескун М. В. (2015). Методи оцінки ефективності впровадження інформаційних систем на промислових підприємствах. Теоретичні і практичні аспекти економіки та інтелектуальної власності, (1 (1)), 21-26.

23. Жадько К. С. (2014). Аспекти оцінки економічної ефективності інформаційних систем обліку та економічного аналізу. Науковий вісник Академії муніципального управління. Серія: Економіка, (1), 192-199.

24. Методичні вказівки для самостійної роботи з дисципліни «Безпека життєдіяльності та основи охорони праці» : навчально-методичний посібник / А. В. Руснак, І. І. Надточій. – Миколаїв : НУК, 2022. – 35 с.

25. Методичні вказівки до виконання розділів "Охорона праці та безпека в надзвичайних ситуаціях" й "Охорона навколишнього середовища" в дипломних

(магістерських) проектах, роботах / О. В. Щедролюєв, Г. В. Коновалова, А. М. Мозговий, В. В. Савельєв. – Миколаїв : НУК, 2017. – 24 с.

26. Крикавський Є., Мащак Н. (2015). Екологічний аспект транспортної логістики. Маркетингові технології в умовах глобалізації економіки України : тези доп. X міжнар. наук.-практ. конференції, Хмельницький : ХНУ, 83-86. URL : https://mr.khmnu.edu.ua/wp-content/uploads/sites/35/2021/06/Tezy_2015.pdf#page=83

ДОДАТОК А. ФРАГМЕНТ КОДУ ПРОГРАМИ

Завантаження та підготовка даних

```
import pandas as pd
```

```
import numpy as np
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.preprocessing import StandardScaler
```

Завантаження історичних даних

```
def load_data(file_path):
```

```
    data = pd.read_csv(file_path)
```

```
    return data
```

Очищення та підготовка даних

```
def clean_and_prepare_data(data):
```

```
    # Видалення пропущених значень
```

```
    data = data.dropna()
```

Перетворення категоріальних змінних

```
data = pd.get_dummies(data, drop_first=True)
```

Стандартизація даних

```
scaler = StandardScaler()
```

```
scaled_data = scaler.fit_transform(data)
```

```
return scaled_data
```

Завантаження даних

```
data = load_data('sales_data.csv')
```

Очищення та підготовка даних

```
prepared_data = clean_and_prepare_data(data)
```

Розділення на тренувальні та тестові дані

```
X = prepared_data.drop(columns=['target_column']) # 'target_column' - стовпець з  
прогнозованими даними
```

```
y = prepared_data['target_column']
```

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

Створення та навчання моделі прогнозування

```
from sklearn.linear_model import LinearRegression
```

```
from sklearn.ensemble import RandomForestRegressor
```

```
from sklearn.metrics import mean_squared_error, r2_score
```

Створення моделей

```
def train_models(X_train, y_train):
```

```
    # Лінійна регресія
```

```
    linear_model = LinearRegression()
```

```
    linear_model.fit(X_train, y_train)
```

```
    # Random Forest регресія
```

```
    rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
```

```
    rf_model.fit(X_train, y_train)
```

```
return linear_model, rf_model
```

```
# Навчання моделей
```

```
linear_model, rf_model = train_models(X_train, y_train)
```

```
# Оцінка моделей на тестових даних
```

```
def evaluate_model(model, X_test, y_test):
```

```
    predictions = model.predict(X_test)
```

```
    mse = mean_squared_error(y_test, predictions)
```

```
    r2 = r2_score(y_test, predictions)
```

```
    return mse, r2
```

```
linear_mse, linear_r2 = evaluate_model(linear_model, X_test, y_test)
```

```
rf_mse, rf_r2 = evaluate_model(rf_model, X_test, y_test)
```

```
print(f"Linear Regression - MSE: {linear_mse}, R2: {linear_r2}")
```

```
print(f"Random Forest - MSE: {rf_mse}, R2: {rf_r2}")
```

```
Прогнозування попиту на нові дані
```

```
# Прогнозування попиту
```

```
def predict_demand(model, new_data):
```

```
    prediction = model.predict(new_data)
```

```
    return prediction
```

```
# Завантажуємо нові дані для прогнозу
```

```
new_data = load_data('new_sales_data.csv')
```

```
new_data_prepared = clean_and_prepare_data(new_data)
```

```
# Прогнозування попиту за допомогою найкращої моделі (наприклад, Random Forest)
```

```
demand_predictions = predict_demand(rf_model, new_data_prepared)
```

```
# Виведення результатів прогнозу
```

```
print("Прогноз попиту на нові товари:", demand_predictions)
```

Генерація рекомендацій для управління запасами

```
def generate_inventory_recommendations(predictions, current_inventory, safety_stock):
```

```
    # Генерація рекомендацій щодо закупок на основі прогнозу попиту
```

```
    recommended_order = predictions - current_inventory + safety_stock
```

```
    recommended_order = np.maximum(recommended_order, 0) # Не можемо замовляти негативну кількість товару
```

```
    return recommended_order
```

```
# Поточні запаси і запас безпеки для кожного товару
```

```
current_inventory = np.array([100, 200, 50, 150]) # Приклад поточних запасів
```

```
safety_stock = np.array([20, 40, 10, 30]) # Приклад запасу безпеки
```

```
# Генерація рекомендацій
```

```
order_recommendations = generate_inventory_recommendations(demand_predictions, current_inventory, safety_stock)
```

```
print("Рекомендовані закупки для товарів:", order_recommendations)
```


Оцінка ефективності моделі прогнозування

```
def assess_forecasting_accuracy(model, X_test, y_test):
    predictions = model.predict(X_test)

    # Оцінка точності прогнозу за допомогою середньоквадратичної помилки
    (MSE) та коефіцієнта детермінації (R2)
    mse = mean_squared_error(y_test, predictions)
    r2 = r2_score(y_test, predictions)
    return mse, r2

# Оцінка ефективності моделі
mse, r2 = assess_forecasting_accuracy(rf_model, X_test, y_test)
print(f"Оцінка ефективності моделі - MSE: {mse}, R2: {r2}")
```

Оптимізація запасів

```
def optimize_inventory(recommendations, current_inventory, max_storage_capacity):
    # Оптимізація закупок з урахуванням максимальних можливих запасів на складі
    total_order = recommendations + current_inventory
    optimized_order = np.minimum(total_order, max_storage_capacity)
    return optimized_order

# Приклад максимальних можливих запасів на складі
max_storage_capacity = np.array([250, 300, 200, 400])

# Оптимізація закупок
optimized_inventory = optimize_inventory(order_recommendations, current_inventory,
max_storage_capacity)

print("Оптимізовані закупки:", optimized_inventory)
```