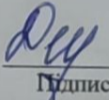
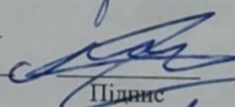
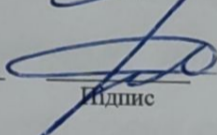


## КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

на тему Метод адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту

Галузь знань 12 – Інформаційні технології  
Шифр і назва галузі знань  
Спеціальність 122 – Комп'ютерні науки  
Шифр і назва спеціальності  
Освітня програма Комп'ютерні науки  
Назва освітньої програми

Виконав: студент групи КН-21-1  Дмитро МЕЛЬНИЦЬКИЙ  
Група виконавця Підпис Ім'я, ПРІЗВИЩЕ  
Керівник: д.т.н., проф. каф. КН  Едуард МАНЗЮК  
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ  
Нормоконтроль: к.т.н., доц. каф. КН  Руслан БАГРІЙ  
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

До захисту допускаю:

Зав. кафедри КН, д.т.н., професор

18 06 2025 р.

  
Підпис

Олександр БАРМАК  
Ім'я, ПРІЗВИЩЕ

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет інформаційних технологій

Кафедра комп'ютерних наук

Освітній ступінь бакалавр

Галузь знань 12 – Інформаційні технології

Спеціальність 122 – Комп'ютерні науки

ЗАТВЕРДЖУЮ

Завідувач кафедри комп'ютерних наук

(підпис)

д.т.н., професор Олександр БАРМАК

« 10 » 02 2025 року

**ЗАВДАННЯ  
НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА**

1. Тема кваліфікаційної роботи бакалавра: «Метод адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту»

2. Завдання видано студенту Дмитро МЕЛЬНИЦЬКИЙ  
(Ім'я, прізвище)

3. Керівник роботи д.т.н., проф. каф. КН Едуард МАНЗЮК  
(посада, ім'я, прізвище)

4. Затверджено наказом університету від «07» 02 2025 р. № 23

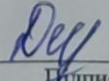
5. Дата видачі завдання студенту: «10» 02 2025 р.

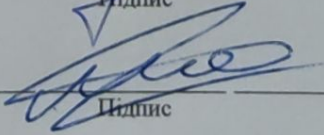
6. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Метою роботи є підвищення точності адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту. Перелік задач: здійснити аналіз предметної області каршерингових систем та особливостей підбору автомобілів у них; провести дослідження існуючих теоретичних підходів до адаптивного розподілу транспортних засобів у системах спільного користування; розробити метод адаптивного підбору автомобілів каршерингу, який використовує алгоритми машинного навчання; спроектувати архітектуру інформаційної системи підбору автомобілів на основі розробленого методу; провести експериментальне дослідження точності розробленого методу адаптивного підбору автомобілів.

7. Календарний план виконання кваліфікаційної роботи бакалавра:

№	Назва етапів (розділів) кваліфікаційної роботи бакалавра	Термін виконання	Примітка
1	Вибір напрямку дослідження та узгодження тематики кваліфікаційної роботи бакалавра з керівником, складання календарного графіка виконання роботи	січень 2025	виконано
2	Ознайомлення з предметною областю, формулювання мети та задач дослідження, визначення об'єкта та предмета дослідження	лютий 2025	виконано
3	Проектування та розробка загальної архітектури системи, вибір засобів реалізації програмного забезпечення	березень 2025	виконано
4	Розробка інформаційної системи	квітень 2025	виконано
5	Написання пояснювальної записки, урахування зауважень керівника, оформлення згідно вимог	травень 2025	виконано
6	Розробка презентаційних матеріалів та попередній захист кваліфікаційної роботи	травень 2025	виконано
7	Отримання відгуку керівника, рецензії, перевірка на плагіат, нормоконтроль	червень 2025	виконано
8	Підготовка до захисту та захист кваліфікаційної роботи бакалавра	червень 2025	виконано

Виконавець: студент групи КН-21-1  Дмитро МЕЛЬНИЦЬКИЙ  
Група виконавця Підпис Ім'я, ПРІЗВИЩЕ

Керівник: д.т.н., проф. каф. КН  Едуард МАНЗЮК  
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

## Анотація

Тема кваліфікаційної роботи бакалавра: Метод адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту.

Виконавець кваліфікаційної роботи бакалавра: студент групи КН-21-1 Дмитро МЕЛЬНИЦЬКИЙ.

Керівник кваліфікаційної роботи бакалавра: д.т.н., проф. каф. КН Едуард МАНЗЮК

Кваліфікаційна робота бакалавра містить:

Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
66	16	3	40	2

Кваліфікаційна робота бакалавра стосується задачі підвищення точності адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту. У рамках дослідження розроблено метод адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту. Здійснено проектування структури інформаційної системи адаптивного підбору автомобілів для каршерингу, розроблено підходи до обробки даних. Експериментальна перевірка на наборі даних продемонструвала працездатність запропонованого підходу.

Ключові слова: каршеринг, адаптивний підбір, градієнтний бустинг, XGBoost.

Виконавець:

студент групи КН-21-1

Група виконавця

  
Підпис

Дмитро  
МЕЛЬНИЦЬКИЙ  
Ім'я, ПРІЗВИЩЕ

## Зміст

Перелік скорочень .....	4
Вступ.....	5
Розділ 1 Дослідження предметної області адаптивного підбору автомобілів каршерингу.....	7
1.1. Сучасний розвиток каршерингу .....	7
1.2. Огляд теоретичних підходів до розв'язку подібних задач .....	8
1.3. Аналіз існуючих програмних засобів та наукових рішень .....	9
1.4 Мета та задачі кваліфікаційної роботи бакалавра .....	16
Розділ 2 Метод адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту .....	17
2.1 Схема та етапи методу адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту .....	17
2.2 Аналіз та автоматизація обробки потоків даних .....	20
2.3 Розробка архітектури градієнтного бустингу .....	24
2.4 Інформаційна структура системи та проектування бази даних .....	30
2.5 Структура та опис набору даних .....	32
2.6 Особливості використання програмних компонентів.....	37
2.7 Висновок до розділу 2 .....	38
Розділ 3 Експериментальна перевірка методу адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту.....	40
3.1 Визначення шляхів дослідження та засобів створення програмного забезпечення .....	40
3.2 Визначення засобів практичної розробки системи .....	42
3.3 Функціональне призначення складових системи .....	44
3.4 Особливості реалізації програмних складових системи.....	48
3.5 Проведення експериментів методу підбору .....	50
3.6 Висновок до розділу 3 .....	57
Висновок .....	58
Перелік посилань.....	59
Додатки	

## Перелік скорочень

Скорочення, термін, позначення	Пояснення
КС	Каршерингова система
XGBoost	eXtreme Gradient Boosting – екстремальний градієнтний бустинг
ML	Machine Learning – машинне навчання
КРБ	Кваліфікаційна робота бакалавра
ПЧД	Просторово-часові дані
MAE	Mean Absolute Error – середня абсолютна помилка
RMSE	Root Mean Square Error – корінь середньоквадратичної помилки
MAPE	Mean Absolute Percentage Error – середня абсолютна відсоткова помилка
nDCG	Normalized Discounted Cumulative Gain – нормалізований дисконтований кумулятивний дохід

## Вступ

Основною метою даної кваліфікаційної роботи бакалавра є підвищення точності адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту..

**Актуальність.** Сучасний ринок транспортних послуг характеризується стрімким розвитком сервісів спільного використання автомобілів, що обумовлено зростанням екологічної свідомості суспільства, урбанізацією та економічними перевагами каршерингу порівняно з приватним володінням транспортними засобами. Розподіл автомобілів у каршерингових системах є надзвичайно складним завданням через динамічність попиту, непередбачуваність поведінки користувачів та необхідність врахування багатьох факторів (часових, просторових, погодних, соціально-демографічних тощо).

Традиційні методи розподілу транспортних засобів часто не здатні адаптуватися до мінливих умов міського середовища, що призводить до незбалансованості автопарку, надмірних простоїв автомобілів в одних районах та їх нестачі в інших. Технології штучного інтелекту, різні методи машинного навчання та глибокі нейронні мережі, демонструють хороший потенціал у вирішенні цих проблем завдяки здатності обробляти великі обсяги даних, виявляти приховані закономірності та формувати точні прогнози попиту на автомобілі в різних локаціях.

Розробка методу адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту є актуальним та важливим завданням, для функціонування систем спільного користування транспортними засобами, що матиме позитивний вплив як на бізнес-показники операторів каршерингу, так і на якість міського середовища в цілому.

**Об'єкт дослідження** – процес підбору транспортних засобів у системах каршерингу.

**Предмет дослідження** – методи та алгоритми інтелектуального аналізу вподобань користувачів для оптимізації вибору автомобілів у системах каршерингу.

**Мета кваліфікаційної роботи бакалавра** – підвищення точності адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту.

Для досягнення поставленої мети необхідно вирішити **такі завдання**: здійснити аналіз предметної області каршерингових систем та особливостей підбору автомобілів у них; провести дослідження існуючих теоретичних підходів до адаптивного розподілу транспортних засобів у системах спільного користування; розробити метод адаптивного підбору автомобілів каршерингу, який використовує алгоритми машинного навчання; спроектувати архітектуру інформаційної системи підбору автомобілів на основі розробленого методу; провести експериментальне дослідження точності розробленого методу адаптивного підбору автомобілів.

## **Розділ 1 Дослідження предметної області адаптивного підбору автомобілів каршерингу**

### **1.1. Сучасний розвиток каршерингу**

В останні роки каршеринг активно розвивається як екологічно дружня та економічно альтернатива приватному автотранспорту. З розширенням попиту на такі послуги зростає потреба у створенні розумних інформаційних моделей для оптимального підбору автомобілів. Такі моделі функціонують на основі комплексного аналізу структурованих даних про транспортні засоби, профілі користувачів, популярні маршрути та інші суттєві фактори. Важливо зазначити, що використання як даних реального часу, так і статистики минулих періодів дозволяє формувати точні прогностичні моделі, які суттєво покращують розподіл автомобілів та задоволеність клієнтів.

Численні дослідження у сфері прогностного моделювання переконливо демонструють, що впровадження алгоритмів відповідного виду машинного навчання суттєво підвищує точність прогнозування попиту та оптимізації автопарку в системах каршерингу. Аналіз даних показує, що особливо високої точності вдається досягти при застосуванні методів векторної регресії та технік глибокого навчання, серед яких моделі CNN-LSTM демонструють найкращі результати [1–3].

Застосування генеративно-змагальних мереж (GAN) та варіаційних автоенкодерів (VAE) суттєво розширило можливості прогностного моделювання, особливо в умовах, коли обсяг реальних даних обмежений. Практичні випробування таких підходів продемонстрували можливість підвищення точності прогнозів поїздок до 4,63%, що є значним досягненням для галузі [4, 5].

Практичне тестування на реальних даних показало, що такий комбінований підхід перевершує традиційні методи приблизно на 9%. Незважаючи на високу точність, слід відзначити, що впровадження цього рішення в повсякденну практику ускладнюється через високу непередбачуваність

поведінки користувачів та постійні динамічні зміни в міських транспортних системах [3, 5, 6].

## 1.2. Огляд теоретичних підходів до розв'язку подібних задач

Розглядаючи теоретичні основи вирішення задач адаптивного підбору автомобілів у каршерингових системах, варто відзначити широкий спектр методологій, які застосовуються дослідниками. Ці підходи охоплюють різноманітні алгоритми машинного навчання та оптимізаційні методи, що в комплексі дозволяють значно покращити якість прийняття рішень у системах спільного користування транспортом. Завдяки таким технологіям вдається вирішувати складні операційні проблеми та підвищувати загальну продуктивність сервісів [7–10].

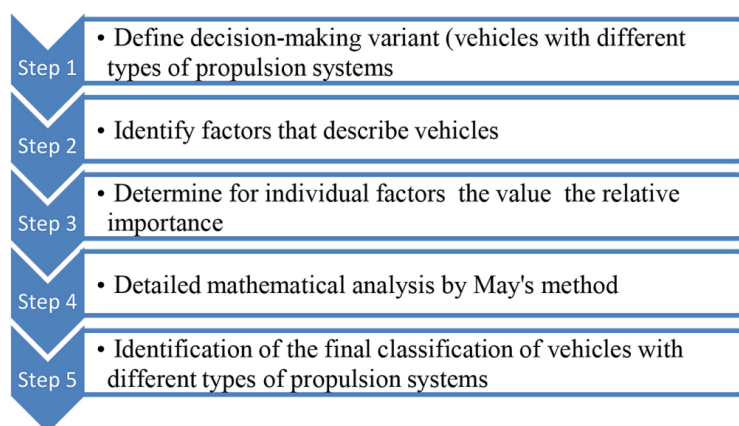


Рисунок 1.1 – Процедура оцінки транспортних засобів, що використовуються в системах каршерінгу [10]

Особливу увагу дослідники приділяють вивченню факторів, які впливають на готовність користувачів брати участь у спільних поїздках. Цікаво, що результати аналізу поведінкових моделей показують, що на рішення людей найбільше впливають саме параметри конкретної подорожі – вартість, відстань та тривалість поїздки. [7].

З метою подолання цих обмежень були розроблені вдосконалені моделі спільного використання транспортних засобів, які більш глибоко враховують індивідуальні особливості користувачів. Однак, незважаючи на ці переваги, ще залишаються певні виклики, пов'язані з психологічними бар'єрами, тривалим часом очікування в деяких випадках та недобросовісними моделями ціноутворення, які негативно впливають на популяризацію спільного користування транспортом [11, 12].

Значний прогрес у покращенні досвіду користувачів був досягнутий завдяки впровадженню моделі спільного використання поїздок, що використовує алгоритми машинного навчання, зокрема Naïve Bayes. Ця модель забезпечує узгодження користувачів на основі їхніх індивідуальних переваг та звичок під час подорожей, що робить сервіс більш персоналізованим та зручним. Порівняно з традиційним громадським транспортом, такий підхід має значні переваги, оскільки забезпечує гнучкість маршрутів та адаптивність до потреб конкретного користувача. Проте все ще залишаються певні проблеми, пов'язані з обмеженнями громадського транспорту, який не забезпечує обслуговування "від дверей до дверей", має фіксовані маршрути та часто ненадійний розклад [11, 13–15].

Особливо перспективним напрямком розвитку є системи обміну поїздками, що функціонують у режимі реального часу. У таких системах інтеграція моделей належного виду машинного навчання дозволяє прогнозувати попит та оптимізувати диспетчеризацію транспортних засобів. Завдяки цьому вдається суттєво покращити якість маршрутизації та управління рухом транспорту, що в результаті призводить до значного скорочення часу очікування для користувачів. Однак варто зазначити, що створення точних прогнозів попиту може бути складним завданням, особливо в умовах недостатньої кількості даних. Крім того, обмеження історичних даних можуть негативно впливати на точність прогнозів переміщення від однієї зони до іншої [16, 17].

### **1.3. Аналіз існуючих програмних засобів та наукових рішень**

Детальний аналіз еволюції програмних засобів у сфері каршерингу демонструє суттєвий прогрес: від базових систем бронювання автомобілів до комплексних платформ із інтегрованими функціями штучного інтелекту. Сучасні дослідження чітко відображають тенденцію до використання все більш складних алгоритмів, які здатні аналізувати та прогнозувати поведінку користувачів, а також оптимізувати управління автопарком в режимі реального часу.

Одним із важливих напрямків застосування машинного навчання є розв'язання проблем паркування в міському середовищі. Використання таких алгоритмів як XGBoost та дерева рішень дозволяє аналізувати закономірності в поведінці водіїв при виборі місць для паркування та формувати точні прогнози щодо попиту на спільні паркувальні місця. Такий підхід значно підвищує точність використання міських парковок та зменшує проблеми із заторами в густонаселених районах [16].

Особливо важливим аспектом у роботі каршерингових систем є прогнозування доступності автомобілів. Проте слід зазначити, що профілі використання автомобілів є надзвичайно мінливими та важко передбачуваними, оскільки різні тенденції використання спостерігаються в різних районах та часових інтервалах [18, 19].

Прогнозування попиту на автомобілі є ще одним критичним аспектом роботи каршерингових систем. Сучасні дослідження в цій сфері включають аналіз як часових, так і просторових закономірностей у поведінці користувачів. Важливим відкриттям стало те, що включення соціально-демографічних даних до моделей прогнозування суттєво підвищує їхню точність. Завдяки цьому вдається досягти відносних похибок не більше 10% для часових прогнозів та до 40% для просторових прогнозів, що є значним покращенням порівняно з традиційними методами [19].

Інноваційним рішенням у сфері прогнозування попиту стала розробка спеціалізованої нейронної мережі USTIN. Ця мережа здатна аналізувати часові, просторові та просторово-часові особливості поведінки користувачів, включаючи історичні дані про попит та інформацію про погодні умови. Такий комплексний

підхід дозволяє значно підвищити точність прогнозування та оптимізувати розподіл ресурсів у системах спільного використання автомобілів [20–23].

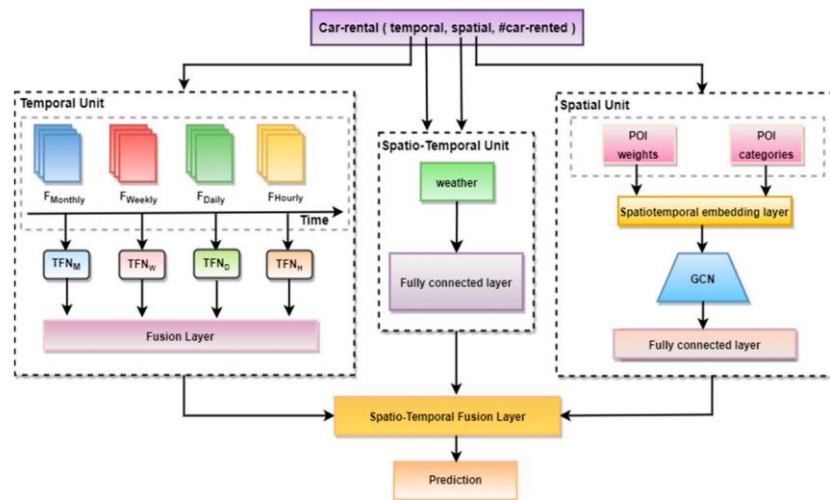


Рисунок 1.3 – Структура просторово-часової мережі прогнозування [22]

Для визначення найбільш точних алгоритмів прогнозування попиту на каршеринг дослідники порівняли моделі випадкового лісу з методом географічно зваженої регресії. Такий підхід дозволяє не лише формувати точні прогнози, знати фактори, які впливають на ці прогнози, що є критично важливим для планування розміщення станцій спільного використання автомобілів [24]. Застосування машинного навчання в контексті високомобільних автомобільних мереж зосереджується на розробці систем інтелектуального прийняття рішень та управління різними ресурсами. Треба зазначити, що висока мобільність таких мереж створює специфічні проблеми при їх проектуванні та впровадженні, для вирішення яких і використовуються передові методи машинного навчання [25, 26]. Варто враховувати, що існують певні перешкоди та складності у застосуванні алгоритмів прогнозування на основі штучного інтелекту, а також важливі питання забезпечення точності та надійності даних [26].

Для оптимізації роботи односторонніх систем спільного використання автомобілів була розроблена спеціальна методологія, яка базується на принципах навчання з використанням історичних та онлайн-даних. Цей інноваційний підхід поєднує офлайн-онлайн-методологію наближеного динамічного програмування з

інтеграцією переваг користувачів та методами байєсівського навчання. Така комбінація дозволяє суттєво оптимізувати процеси переміщення транспортних засобів та розробку стимулів для користувачів. Однак практичне впровадження цієї методології стикається з певними труднощами, пов'язаними з невизначеністю в запитах на поїздку та часі подорожі, а також із недостатньою кількістю даних щодо мотивації користувачів при виборі певних опцій [27, 28].

Використання ансамблевих методів для аналізу характеристик подорожей у контексті спільного використання автомобілів відкриває нові можливості для інтеграції особистих, екологічних та поведінкових даних. Цей комплексний підхід машинного навчання дозволяє диференціювати різні типи подорожей, що здійснюються за допомогою каршерингу, і таким чином покращити розуміння атрибутів користувачів та специфіки різних поїздок [29].

Для вирішення проблем координації спільних автономних транспортних засобів було розроблено децентралізований алгоритм, який базується на мультиагентному навчанні з підкріпленням та використанні мереж уваги на графіку. Такий підхід дозволяє вирішувати складні проблеми, пов'язані з динамікою запитів на транспортні засоби, та значно підвищити загальну точність систем мобільності на вимогу [30, 31].

Підвищення безпеки та конфіденційності даних у системах спільного використання транспорту стало можливим завдяки розробці систем колективного навчання на основі блокчейну для підключених та автономних транспортних засобів. Цей інноваційний підхід дозволяє підвищити точність алгоритмів машинного навчання та отримуючи високий рівень захисту особистих даних користувачів. Проте варто зазначити, що така система стикається з певними викликами, пов'язаними з обробкою великих скупчень даних та забезпеченням безпеки бортових сенсорних систем [32].

Розвиток персоналізованого розподіленого штучного інтелекту, який забезпечує високий рівень конфіденційності при взаємодії транспортних засобів на дорогах, є ще одним перспективним напрямком. Цей підхід використовує передові методи навчання для підвищення безпеки даних та покращення точності

моделей. Однак слід зазначити, що існуючі методи захисту конфіденційності часто призводять до зниження точності моделей, а основні алгоритми вимагають значного часу для обчислень [33, 34].

Системи прогнозного обміну поїздками на транспортних засобах використовують зважені моделі узгодження шаблонів з підтримкою штучного інтелекту для оцінки поведінки користувачів. Ці інноваційні рішення дозволяють надавати персоналізовані рекомендації для пасажирів у контексті розумних міст, що сприяє розвитку сталого транспорту та зменшенню забруднення навколишнього середовища в міських районах [35].

Важливим кроком у розвитку систем спільного використання поїздок стала розробка розподіленого алгоритму з кількома пересадками, який не потребує попередньо визначених моделей. Цей інноваційний підхід, заснований на методах глибокого навчання з підкріпленням, дозволяє оптимізувати процеси диспетчеризації транспортних засобів та приймати рішення в режимі реального часу. Важливо враховувати, що для роботи цього алгоритму необхідно мати доступ до інформації про стан транспортних засобів у реальному часі без затримок, а запити користувачів можуть бути відхилені, якщо поблизу немає доступних транспортних засобів [36].

Для глибокого розуміння поведінки користувачів у контексті спільного використання автомобілів був розроблений підхід, який використовує мережу глибоких переконань та вдосконалену двоетапну кластеризацію. Цей метод підкреслює важливу роль машинного навчання та штучного інтелекту в аналізі просторово-часових моделей поведінки користувачів, що також сприяє покращенню сталого розвитку спільного використання автомобілів та зменшенню проблем, пов'язаних з міським паркуванням [37–39]. Важливо зазначити, що практичне впровадження цього методу стикається з певними труднощами, пов'язаними з великим масштабом проблем, нелінійністю процесів та невизначеністю ендогенного попиту, що дуже ускладнює винайдення оптимального рішення [38–41].

Значним досягненням у розвитку систем спільного використання поїздок став підхід до оптимізації за допомогою глибокого навчання з підкріпленням та лінійного програмування. Ця методологія комплексно вирішує питання диспетчеризації запитів, планування трансферів та ребалансування транспортних засобів. Практичне впровадження такого підходу дозволяє суттєво підвищити загальну точність системи та значно зменшити операційні витрати при організації спільних поїздок [39]. Проведений аналіз інформаційних моделей, теоретичних підходів та існуючих програмних засобів у сфері каршерингу з інтеграцією технологій штучного інтелекту дозволяє зробити наступні висновки. Дослідження переконливо демонструють, що впровадження машинного навчання та методів глибокого навчання для прогнозування попиту та оптимізації розподілу транспортних засобів суттєво підвищує точність каршерингових систем. Це особливо помітно при застосуванні таких методів як векторна регресія та CNN-LSTM моделі, які забезпечують високу точність прогнозів.

Важливим досягненням у розвитку каршерингових платформ стало використання генеративних моделей машинного навчання, зокрема генеративно-змагальних мереж та варіаційних автоенкодерів. Ці технології дозволяють створювати якісні синтетичні дані, що має критичне значення в умовах обмеженості реальних даних для навчання моделей, покращуючи прогнози поїздок.

Розроблений двофазний підхід з евристикою жадібної вставки та метаевристикою на основі локального пошуку продемонстрував значні переваги над існуючими методами, перевершуючи їх до дев'яти відсотків за реальними даними. Проте практичне впровадження цього підходу стикається з певними обмеженнями через високу непередбачуваність поведінки користувачів та постійні динамічні зміни в транспортній системі міського середовища.

Комплексне врахування просторово-часових особливостей разом із використанням історичного попиту та погодних даних значно покращує точність прогнозування попиту на спільне використання автомобілів. Такий підхід

реалізований, зокрема, в спеціалізованій нейронній мережі USTIN, що дозволяє розташувати розміщення ресурсів у каршерингових системах.

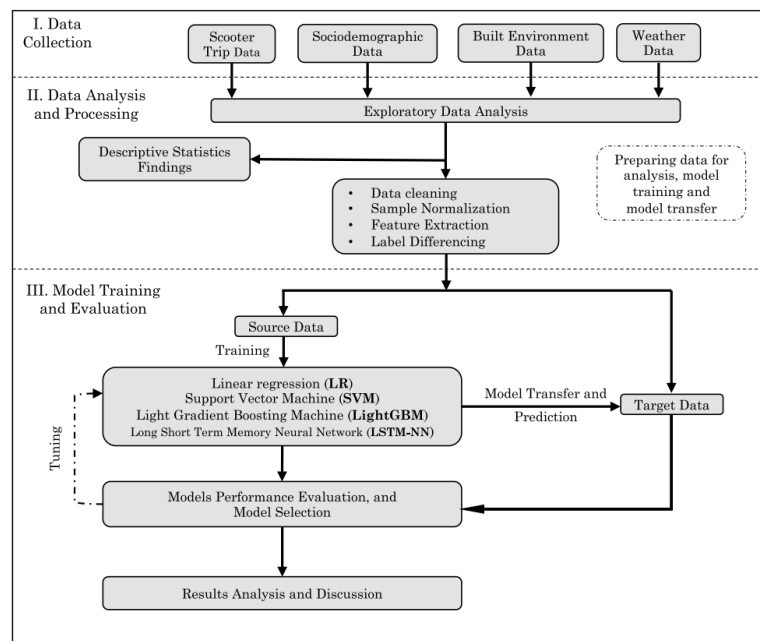


Рисунок 1.2 – Архітектура застосованого методу [38]

Важливим аспектом при розробці сучасних систем каршерингу з інтеграцією технологій штучного інтелекту є забезпечення високого рівня конфіденційності даних та безпеки користувачів. Розвиток систем колективного навчання на основі блокчейну та персоналізованого розподіленого штучного інтелекту сприяє підвищенню безпеки даних при одночасному збереженні високої точності моделей.

Незважаючи на значні досягнення, у галузі каршерингу з використанням штучного інтелекту залишається ряд суттєвих викликів. Серед них невизначеність у запитах на поїздку та часі подорожі, проблеми з прогнозуванням попиту при обмежених історичних даних, складність обробки значних обсягів даних у реальному часі та забезпечення балансу між точністю моделей та захистом конфіденційності користувачів.

Розробка методу адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту вимагає комплексного підходу, що враховує просторово-часові особливості, персоналізацію, безпеку даних та

розміщення ресурсів. Добрим напрямком є з'єднання різномісних методів машинного навчання з оптимізаційними алгоритмами, що дозволить створити гнучкі та адаптивні системи, здатні функціонувати в динамічному міському середовищі та задовольняти постійно еволюціонуючі потреби користувачів.

#### **1.4 Мета та задачі кваліфікаційної роботи бакалавра**

Мета кваліфікаційної роботи бакалавра – підвищення точності адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту.

Для досягнення поставленої мети необхідно вирішити такі завдання:

- здійснити аналіз предметної області каршерингових систем та особливостей підбору автомобілів у них;
- провести дослідження існуючих теоретичних підходів до адаптивного розподілу транспортних засобів у системах спільного користування;
- розробити метод адаптивного підбору автомобілів каршерингу, який використовує алгоритми машинного навчання;
- спроектувати архітектуру інформаційної системи підбору автомобілів на основі розробленого методу;
- провести експериментальне дослідження точності розробленого методу адаптивного підбору автомобілів.

## **Розділ 2 Метод адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту**

### **2.1 Схема та етапи методу адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту**

Розробка методу адаптивного підбору автомобілів каршерингу потребує комплексного підходу до даних та прийняття рішень на основі технологій штучного інтелекту. Запропонований метод враховує різноманітність наявних даних про автомобілі, користувачів, поїздки та пропозиції, а також просторові та часові аспекти використання сервісу каршерингу.

Метод базується на використанні градієнтного бустингу як основного алгоритму, що дозволяє обробляти різноманітні дані та виявляти складні взаємозв'язки між параметрами. Застосування цього засобу зумовлене необхідністю врахування багатьох факторів при підборі автомобілів для користувачів, таких як характеристики автомобілів, географічне розташування, час доби, кількість пасажирів, рейтинги та попередній досвід користувачів.

Процес адаптивного підбору автомобілів складається з п'яти основних етапів, кожен з яких вирішує конкретні завдання та використовує відповідні технології обробки даних. На етапі ініціалізації вхідної інформації відбувається завантаження та первинна обробка даних з чотирьох основних джерел: інформації про автомобілі (`cars.csv`), користувачів (`users.csv`), здійснені поїздки (`travels.csv`) та цінові пропозиції (`offers.csv`). На цьому етапі виконується перевірка повноти та якості даних, встановлення значень за замовчуванням для відсутніх параметрів, аналіз потрібного обсягу зразків для того, щоб навчити модель, а також перевірка можливості формування розкладу з урахуванням поточних обмежень. За необхідності створюються додаткові зразки розкладів шляхом клонування існуючих для забезпечення достатньої кількості навчальних даних.

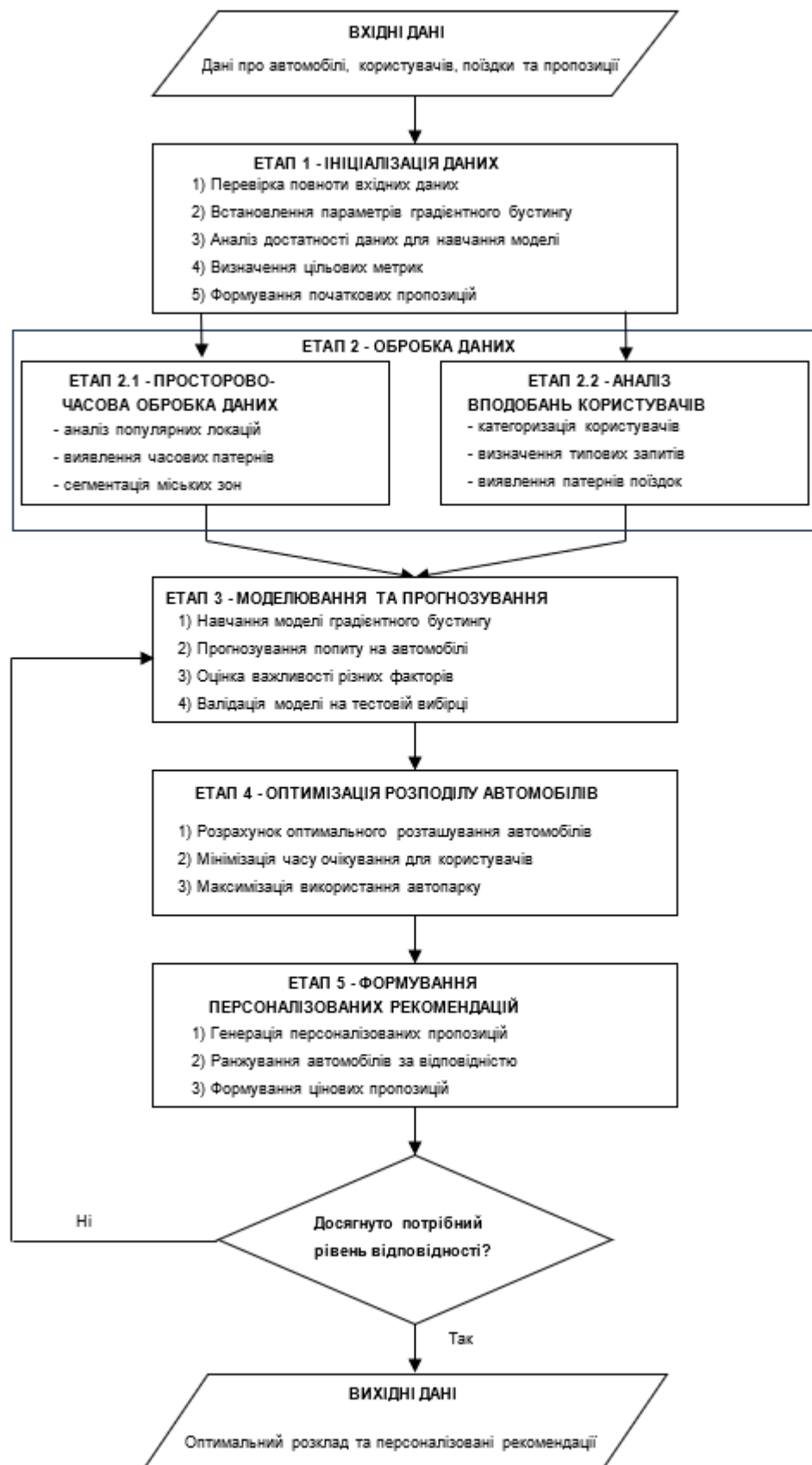


Рисунок 2.1 – Схема та етапи методу адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту

Другий етап присвячений створенню належних зразків розкладів для формування популяції. Цей етап розділений на два паралельні процеси: етап 2.1 - створення належних зразків розкладів шляхом параметричного мутування, та етап 2.2 - створення належних зразків розкладів шляхом точкового мутування. Параметрична мутація передбачає зміну параметрів автомобіля (марка, модель, рік випуску, кількість місць) для досягнення кращої відповідності запитам користувачів. Точкова мутація зосереджується на зміні географічних координат доступності автомобілів для оптимізації просторового розподілу.

На третьому етапі відбувається отримання належних зразків розкладів застосовуючи операцію схрещування найкращих представників популяції. Цей процес включає формування нової генерації зразків розкладів, оцінку їх відповідності вимогам користувачів за допомогою обрахунку функції пристосованості, та формування потрібної вибірки з хорошого набору зразків розкладів. Функція пристосованості враховує близькість розташування автомобіля до користувача, відповідність характеристик автомобіля запитам (кількість місць, рік випуску, марка), цінову пропозицію та рейтинги автомобіля та користувача.

Четвертий етап присвячений перевірці умов завершення роботи еволюційного процесу. На цьому етапі виконується порівняння кращих зразків розкладів з попередніми поколіннями, перевіряється досягнення потрібної оцінки розкладу за вимогами користувачів, та визначається загальна кількість поколінь. Якщо умови виходу не виконані, процес повертається до етапу формування покоління надалі.

На п'ятому етапі проходимо формування наступного покоління з цих поколінь. Цей етап поміщає відбір кращих зразків початкового покоління, кращих зразків покоління що було, та кращих зразків покоління обробки. Відбір здійснюється на основі функції пристосованості, яка враховує всі важливі параметри для забезпечення оптимального підбору автомобілів.

Результатом застосування даного методу є розклад машин з максимальним значенням оцінки відповідності вимогам користувачів, а також висновок про формування того положення за наявними даними у разі невдачі.

Використання технологій градієнтного бустингу дозволяє обробляти різноманітні дані та виявляти складні взаємозв'язки між параметрами, що забезпечує високу точність прогнозування попиту та формування рекомендацій. Врахування просторово-часових аспектів використання сервісу каршерингу дозволяє оптимізувати розподіл автомобілів та мінімізувати час очікування для користувачів. Особливу цінність методу надає можливість генерації персоналізованих пропозицій для користувачів на основі їх попереднього досвіду та вподобань. Це дозволяє підвищити задоволеність користувачів та точність використання автопарку, що є ключовими показниками успішності сервісу каршерингу. Метод адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту дозволяє досягти підвищення точності роботи сервісу шляхом оптимізації розподілу ресурсів та персоналізації взаємодії з користувачами.

## **2.2 Аналіз та автоматизація обробки потоків даних**

Функціональна структура інформаційної системи адаптивного підбору автомобілів каршерингу представляє собою складну взаємопов'язану екосистему модулів, що забезпечують комплексну обробку даних, аналіз, моделювання та формування рекомендацій. Запропонована структура реалізує гнучкий підхід до організації потоків даних, забезпечуючи більш надійну роботу між компонентами системи.

Центральним елементом функціональної структури є "Центральний аналітичний контролер", який координує взаємодію між усіма модулями системи та забезпечує узгодженість потоків даних. Він відповідає за розподіл завдань між функціональними модулями, контроль виконання цих завдань та інтеграцію результатів їх роботи. Контролер також забезпечує керування процесами

адаптації та навчання системи, що є ключовим для реалізації адаптивного підходу до підбору автомобілів.

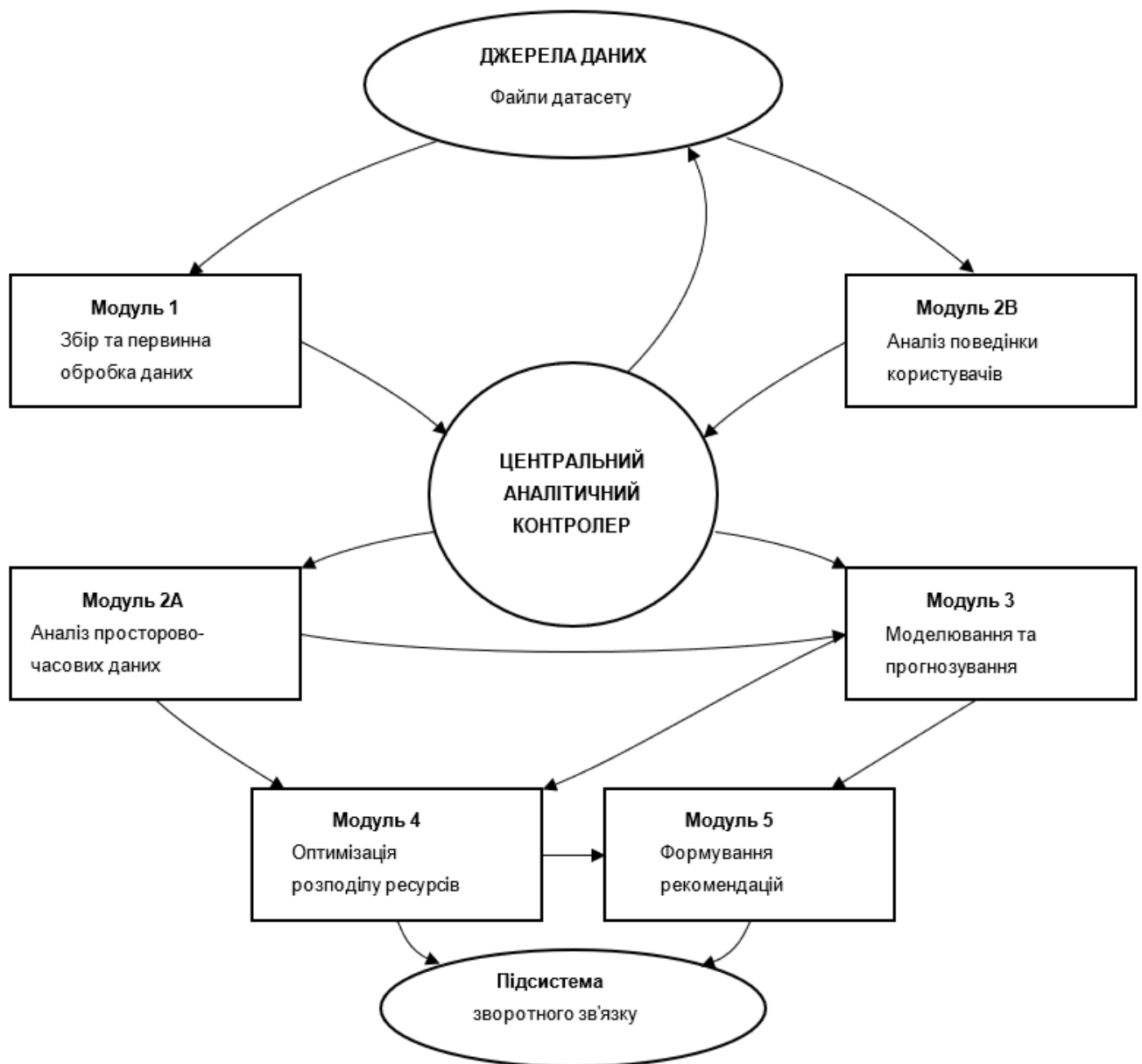


Рисунок 2.2 – Комплексна функціональна структура інформаційної системи адаптивного підбору автомобілів каршерингу

Джерела даних представлені єдиним централізованим блоком, який об'єднує інформацію з усіх вхідних файлів (cars.csv, users.csv, travels.csv, offers.csv). Така централізація спрощує доступ до даних та забезпечує цілісність інформаційної бази системи. Централізований блок даних взаємодіє

безпосередньо з Модулем 1 (Збір та первинна обробка даних) та Модулем 2В (Аналіз поведінки користувачів), які виконують початкову обробку та аналіз даних.

Модуль 1 відповідає за агрегацію, очищення та нормалізацію даних з різних джерел. Зокрема, він виконує перевірку коректності даних, усунення дублікатів, заповнення пропущених значень та приведення даних до єдиного формату. Оброблені дані передаються до Центрального аналітичного контролера для подальшої обробки та аналізу. Модуль також забезпечує оновлення даних в центральному сховищі, що відображено зворотним потоком даних від Центрального контролера до блоку джерел даних.

Модуль 2В фокусується на аналізі поведінки користувачів, що включає вивчення їх вподобань, патернів використання сервісу та реакцій на різні пропозиції. Цей модуль використовує методи кластеризації та класифікації для категоризації користувачів за типами поведінки, що дозволяє формувати персоналізовані рекомендації. Результати аналізу передаються до Центрального аналітичного контролера для інтеграції з іншими даними та використання в процесі прийняття рішень.

Центральний аналітичний контролер, отримавши оброблені дані від Модулів 1 та 2В, передає їх до Модуля 2А (Аналіз просторово-часових даних) та Модуля 3 (Моделювання та прогнозування). Такий розподіл потоків даних забезпечує паралельну обробку інформації, що підвищує результат роботи системи.

Модуль 2А спеціалізується на аналізі просторово-часових даних, виявляючи закономірності в розподілі попиту на автомобілі в різних локаціях та часових проміжках. Він використовує методи просторової статистики та часових рядів для виявлення патернів, сезонності, трендів та аномалій у даних. Результати аналізу використовуються для оптимізації розподілу автомобілів та прогнозування попиту.

Між Модулем 2А та Модулем 3 існує двонаправлений потік даних, що відображає тісну взаємодію між аналізом просторово-часових даних та

моделюванням. Модуль 3 використовує результати аналізу для розробки та навчання моделей машинного навчання, а потім передає результати моделювання назад до Модуля 2А для уточнення параметрів аналізу. Така ітеративна взаємодія дозволяє постійно покращувати якість моделей та підвищувати точність прогнозів.

Модуль 3 відповідає за розробку, навчання та валідацію моделей машинного навчання, а також за прогнозування попиту на автомобілі. Він використовує алгоритми градієнтного бустингу для обробки різнорідних даних та виявлення складних взаємозв'язків між параметрами. Результати моделювання та прогнози передаються до Модулів 4 та 5 для подальшого використання в процесах оптимізації та формування рекомендацій.

Модуль 4 оптимізація розподілу ресурсів отримує дані як від Модуля 2А, так і від Модуля 3, що дозволяє враховувати як аналітичні висновки, так і прогнози при розрахунку оптимального розподілу автомобілів. Цей модуль використовує алгоритми комбінаторної оптимізації для визначення найкращого розміщення автомобілів з урахуванням різних критеріїв, таких як мінімізація часу очікування, максимізація використання автопарку та мінімізація витрат на переміщення.

Модуль 5 формування рекомендацій отримує дані від Модулів 3 та 4, що дозволяє формувати персоналізовані рекомендації для користувачів з урахуванням як їх індивідуальних преференцій, так і оптимального розподілу автомобілів в системі. Цей модуль використовує алгоритми ранжування для визначення найбільш релевантних пропозицій для кожного користувача та формування цінових пропозицій на основі прогнозованого попиту.

Підсистема зворотного зв'язку є важливим компонентом функціональної структури, що забезпечує адаптивність системи. Вона отримує дані від Модулів 4 та 5 про точність оптимізації та реакціях користувачів на рекомендації, аналізує ці дані та передає результати аналізу до Центрального аналітичного контролера. Ця інформація використовується для коригування параметрів моделей,

алгоритмів оптимізації та стратегій формування рекомендацій, що дозволяє системі адаптуватися до змін у поведінці користувачів та умовах експлуатації.

Така складна взаємопов'язана структура забезпечує обробку потоків даних та гнучке управління процесами аналізу, моделювання, оптимізації та формування рекомендацій. Нелінійність потоків даних та наявність зворотних зв'язків дозволяють системі постійно вдосконалюватися, адаптуючись до нових даних та умов експлуатації, що є ключовим для забезпечення високої точності адаптивного підбору автомобілів каршерингу.

### **2.3 Розробка архітектури градієнтного бустингу**

Аналіз датасету каршерингу показує, що оптимальним методом штучного інтелекту для вирішення задач оптимізації буде прогнозне моделювання з використанням алгоритмів градієнтного бустингу. Комплексна структура даних, що включає інформацію про користувачів, автомобілі, пропозиції та поїздки, вимагає підходу, здатного обробляти різномірну інформацію та виявляти складні залежності.

Технічні переваги градієнтного бустингу також відповідають характеристикам представлених даних. Алгоритми XGBoost або LightGBM демонструють високу стійкість до шуму, неповних даних та викидів, що часто зустрічаються в реальних системах каршерингу. Впровадження запропонованого методу дозволить вирішити ключові бізнес-задачі: оптимізувати розташування автомобілів відповідно до прогнозованого попиту, впровадити систему динамічного ціноутворення, розробити персоналізовані рекомендації для користувачів та покращити планування доступності транспортних засобів. Такий комплексний підхід не лише підвищить точність використання автопарку, але й покращить користувацький досвід, що є критичним фактором успіху в конкурентному середовищі послуг каршерингу.

Для реалізації методу адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту обрано алгоритм XGBoost. Даний

вибір обумовлений високою точністю алгоритму при роботі з різнорідними даними, можливістю виявлення складних нелінійних залежностей та забезпеченням інтерпретованості результатів, що важливо для прикладного застосування в системі каршерингу.

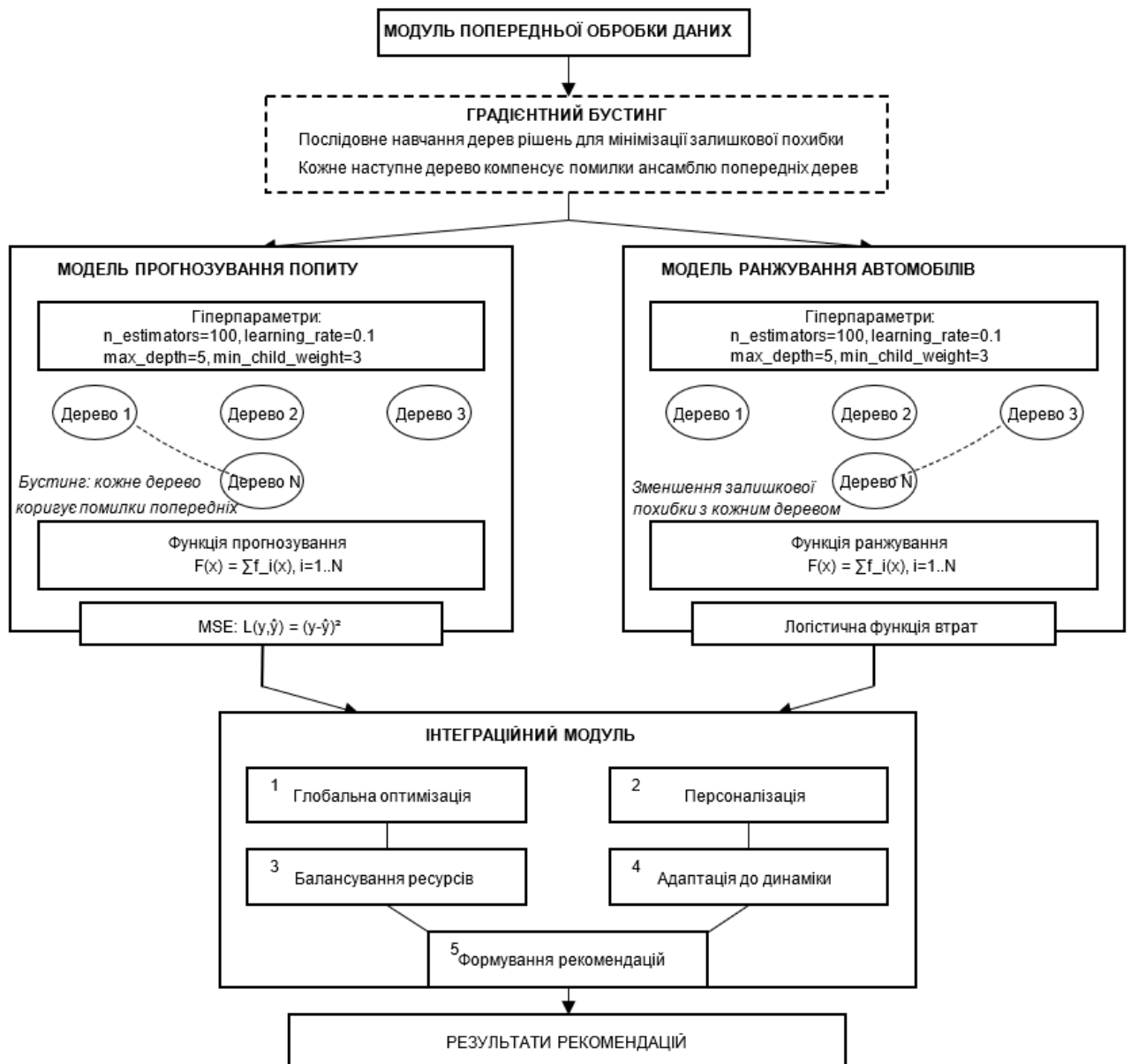


Рисунок 2.3 – Архітектура моделі XGBoost для адаптивного підбору автомобілів каршерингу

XGBoost представляє собою оптимізовану бібліотеку градієнтного бустингу, що використовує ансамбль дерев рішень. Основні переваги даного алгоритму полягають у вбудованій регуляризації для запобігання перенавчанню,

обробці даних, можливості обробки пропущених значень без попереднього заповнення та високій продуктивності завдяки паралелізації обчислень.

Статична архітектура моделі XGBoost для вирішення задачі адаптивного підбору автомобілів каршерингу має складну ієрархічну структуру, що включає компоненти попередньої обробки даних, ансамблі дерев рішень та механізми оцінки та інтеграції результатів.

Компонент попередньої обробки даних відповідає за трансформацію вхідних різномірних даних у формат, придатний для використання в алгоритмах машинного навчання. Даний компонент включає процедури стандартизації числових ознак для приведення їх до єдиного масштабу, кодування категоріальних змінних, обробку просторово-часових даних та створення похідних ознак для підвищення прогностичної здатності моделі.

Ядром архітектури є дві спеціалізовані моделі XGBoost: модель прогнозування попиту та модель ранжування автомобілів. Кожна модель має власну структуру ансамблю дерев рішень з визначеними гіперпараметрами, оптимізованими для вирішення відповідних підзадач.

Модель прогнозування попиту використовує структуру ансамблю з фіксованими параметрами: кількість дерев становить 100, швидкість навчання - 0.1, максимальна глибина дерева - 5, мінімальна кількість спостережень у листовому вузлі - 3. Кожне дерево в ансамблі має бінарну структуру з внутрішніми вузлами, що представляють умови розгалуження на основі значень ознак, та листовими вузлами, що містять прогнозовані значення. Дана модель приймає на вхід характеристики локацій, часові параметри, погодні умови та історичні дані про використання автомобілів і генерує прогноз попиту для різних локацій та часових проміжків.

Модель ранжування автомобілів має аналогічну базову структуру з тими ж гіперпараметрами, проте відрізняється цільовою функцією та механізмом інтерпретації результатів. Вхідними даними для цієї моделі є характеристики автомобілів, профілі користувачів та контекстуальні фактори, а вихідними -

оцінки відповідності кожного автомобіля потребам конкретного користувача, представлені у вигляді ймовірностей.

Особливістю архітектури моделі XGBoost є механізм регуляризації, що контролює складність моделі та запобігає перенавчанню. Цей механізм реалізований через параметри  $\alpha$  (регуляризація L1) та  $\lambda$  (регуляризація L2), які обмежують ваги листових вузлів та сприяють побудові більш узагальнених моделей.

Аналітична компонента моделі представлена функцією рішення, яка для ансамблю з  $m$  дерев має вигляд:

$$F_m(x) = \sum_{i=1}^m f_i(x), \quad (2.1)$$

де  $F_m(x)$  - прогноз моделі після додавання  $m$  дерев,  $f_i(x)$  - внесок  $i$ -го дерева,  $x$  - вектор ознак.

Кожне дерево  $f_i(x)$  відображає вхідні дані на значення відповідного листового вузла:

$$f_i(x) = w_q(x), \quad (2.2)$$

де  $q(x)$  - функція, що відображає вхідні дані  $x$  на індекс листового вузла,  $w_q(x)$  - значення ваги для цього вузла.

Функція втрат, що мінімізується при навчанні моделі, має вигляд:

$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k), \quad (2.3)$$

де  $l(y_i, \hat{y}_i)$  - функція втрат, що вимірює різницю між прогнозованим та фактичним значеннями,  $\Omega(f_k)$  - член регуляризації,  $K$  - кількість дерев,  $n$  - кількість спостережень.

Для моделі прогнозування попиту використовується середньоквадратична помилка як функція втрат, а для моделі ранжування автомобілів - логістична функція втрат.

Архітектура XGBoost також включає механізми для обробки розріджених даних та пропущених значень, що особливо важливо для задачі адаптивного підбору автомобілів каршерингу, де деякі ознаки можуть бути доступні не для всіх спостережень.

Інтеграційний компонент архітектури відповідає за об'єднання результатів роботи моделей прогнозування попиту та ранжування автомобілів для формування кінцевих рекомендацій. Цей компонент балансує між глобальною оптимізацією розподілу ресурсів та персоналізацією рекомендацій для користувачів, враховуючи прогнозований попит на автомобілі в різних локаціях та індивідуальні переваги користувачів.

Розроблена архітектура градієнтного бустингу на основі XGBoost для адаптивного підбору автомобілів каршерингу представляє собою інтегровану систему, що комбінує технології машинного навчання для вирішення задачі підбору автомобілів відповідно до потреб користувачів та поточної ситуації на ринку каршерингових послуг.

Архітектура має ієрархічну структуру з чітким розподілом функціональності між компонентами. На верхньому рівні знаходиться модуль попередньої обробки даних, який забезпечує трансформацію вхідної інформації у формат, придатний для машинного навчання. Він включає процедури стандартизації числових ознак, One-Hot Encoding для категоріальних змінних, просторово-часове кодування та генерацію похідних ознак. Коректна попередня обробка даних є критично важливою для забезпечення високої якості прогнозів та рекомендацій.

Центральними елементами архітектури є дві спеціалізовані моделі XGBoost: модель прогнозування попиту та модель ранжування автомобілів. Кожна з цих моделей представляє собою ансамбль з 100 дерев рішень, які навчаються послідовно для мінімізації відповідних функцій втрат. Модель

прогнозування попиту використовує середньоквадратичну помилку (MSE) як функцію втрат та орієнтована на передбачення кількості запитів на автомобілі в різних локаціях та часових проміжках. Модель ранжування автомобілів застосовує логістичну функцію втрат для оцінки відповідності кожного доступного автомобіля потребам конкретного користувача.

Важливим компонентом архітектури є механізм регуляризації, який контролює складність моделі через параметри  $\alpha$  (регуляризація L1) та  $\lambda$  (регуляризація L2). Регуляризація L1 ( $\alpha=0.1$ ) сприяє розрідженості моделі шляхом зменшення кількості використовуваних ознак, а регуляризація L2 ( $\lambda=1.0$ ) запобігає надмірному зростанню ваг. Ці механізми суттєво підвищують стійкість моделі до перенавчання, що критично важливо для систем, які працюють з динамічними даними.

Кожне дерево в ансамблі має бінарну структуру з внутрішніми вузлами, які представляють умови розгалуження на основі порогових значень ознак, та листовими вузлами з ваговими коефіцієнтами, що використовуються для формування прогнозу. Максимальна глибина дерева обмежена значенням 5, а мінімальна кількість спостережень у листовому вузлі встановлена на рівні 3, що забезпечує баланс між точністю та узагальнюючою здатністю моделі.

На нижньому рівні архітектури знаходиться інтеграційний модуль, який об'єднує результати роботи двох моделей для формування кінцевих рекомендацій щодо підбору автомобілів каршерингу. Цей модуль враховує як прогнозований попит для оптимізації розподілу автомобілів по локаціях, так і результати ранжування для забезпечення персоналізації рекомендацій відповідно до вподобань користувачів.

Важливою особливістю розробленої архітектури є її адаптивність, що досягається через постійне оновлення моделей на основі нових даних про взаємодію користувачів з системою. Це дозволяє системі враховувати зміни в поведінці користувачів, появу нових автомобілів та інші динамічні аспекти, забезпечуючи високу якість рекомендацій у довгостроковій перспективі.

Запропонована архітектура градієнтного бустингу XGBoost забезпечує вирішення задачі адаптивного підбору автомобілів каршерингу, поєднуючи високу точність прогнозів з інтерпретованістю результатів та здатністю адаптуватися до змін у даних. Саме ці характеристики роблять дану архітектуру оптимальним вибором для систем інтелектуального каршерингу.

## 2.4 Інформаційна структура системи та проєктування бази даних

Інформаційна структура системи адаптивного підбору автомобілів каршерингу ґрунтується на реляційній моделі даних, що забезпечує зберігання та обробку всієї необхідної інформації. Для забезпечення функціональності системи розроблено даталогічну модель бази даних, представлену на рисунку 2.4, яка відображає взаємозв'язки між основними сутностями предметної області.

Центральними сутностями моделі виступають Cars (Автомобілі), Orders (Замовлення), Clients (Клієнти) та Requests (Запити). Таблиця Cars містить вичерпну інформацію про доступні для оренди автомобілі, включаючи їх технічні характеристики (модель, рік випуску, тип коробки передач, об'єм двигуна, витрати палива), класифікаційні параметри (клас автомобіля, тип кузова) та умови оренди (застава, вартість). Ця таблиця пов'язана з низкою довідникових таблиць, таких як Models, Manufacturers, Bodies, CarClasses, Gearboxes, FuelTypes, Engines та Gears, що забезпечує нормалізацію даних та запобігає їх дублюванню.

Таблиця Orders зберігає інформацію про замовлення на оренду автомобілів і містить зв'язки з таблицями Cars (для вказівки на конкретний автомобіль), Clients (для ідентифікації замовника), Places (для визначення місць отримання та повернення автомобіля) та InsuranceTypes (для вказівки обраного типу страхування). Важливими атрибутами цієї таблиці також є часові параметри оренди (DeliveryTime, ReturningTime), які використовуються при плануванні доступності автомобіля.

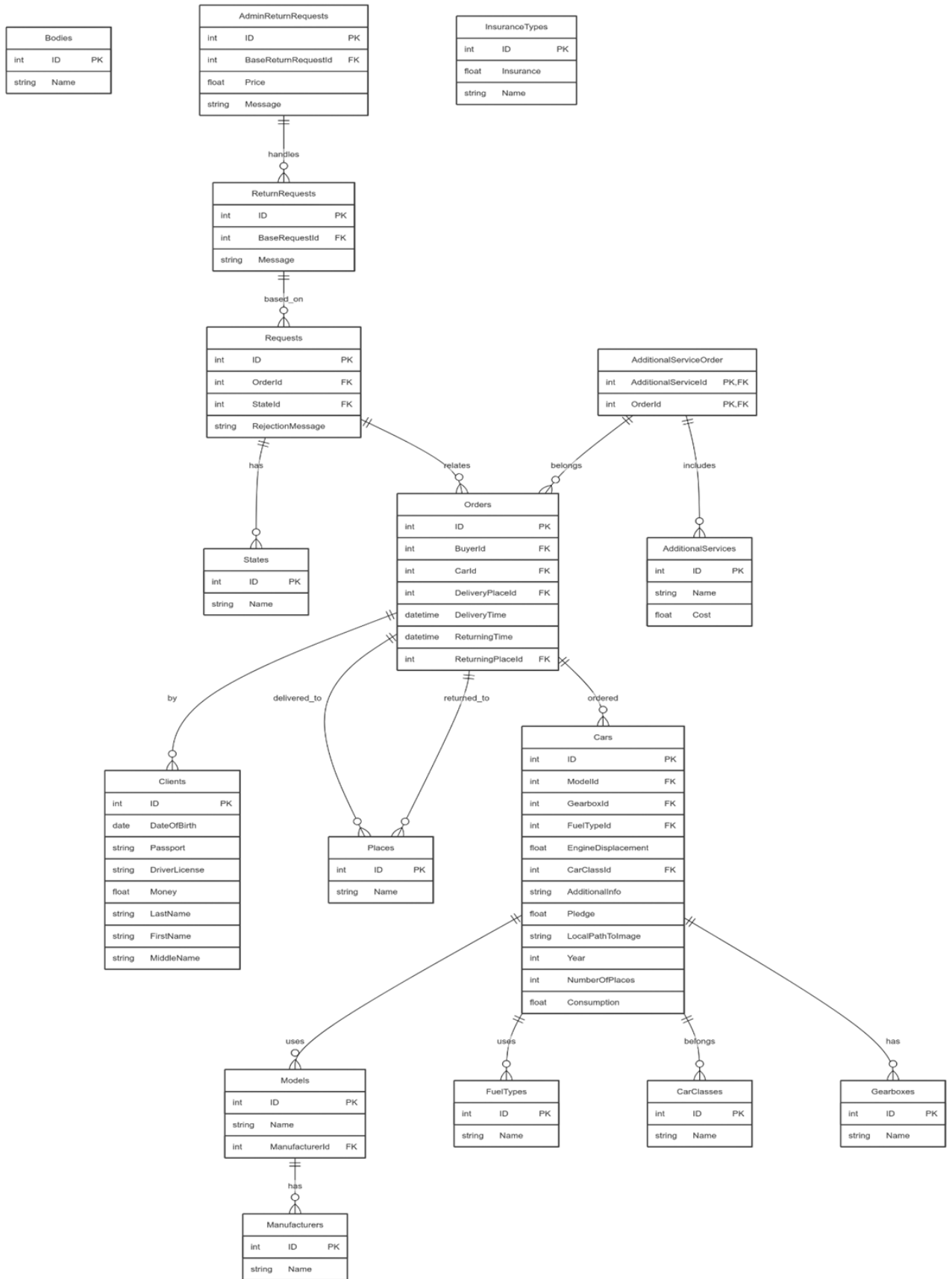


Рисунок 2.4 – Даталогічна модель бази даних

Таблиця Clients містить персональні дані користувачів системи, включаючи їх ідентифікаційні документи (паспорт, водійське посвідчення), контактну інформацію та фінансовий баланс. Ця інформація є критично важливою для процесу авторизації користувачів та контролю їх платоспроможності.

Для забезпечення процесу обробки запитів на оренду та повернення автомобілів розроблено систему таблиць Requests, ReturnRequests та AdminReturnRequests, які з'єднуються між собою та з таблицею States. Така структура дозволяє повністю контролювати життєвий цикл запиту від його створення до завершення, а також зберігати історію змін статусів.

Особливу увагу приділено можливості розширення функціональності системи, що реалізовано через структуру додаткових послуг (AdditionalServices, AdditionalServiceOrder). Ця частина моделі забезпечує можливість вибору користувачами різноманітних додаткових опцій при оренді автомобілів, таких як GPS-навігація, дитяче крісло чи інші сервіси.

Важливим аспектом інформаційної структури є система зв'язків між таблицями, реалізована через механізм зовнішніх ключів (FK). В моделі використано різні типи різних зв'язків: один-до-одного (1:1), наприклад, між Requests та ReturnRequests; один-до-багатьох (1:M), наприклад, між States та Requests; та багато-до-багатьох (M:N), реалізований через проміжну таблицю

Розроблена інформаційна структура системи адаптивного підбору автомобілів каршерингу забезпечує зберігання та обробку всіх необхідних даних, а також підтримує всі функціональні вимоги, визначені в розділі 2.1.

## **2.5 Структура та опис набору даних**

У цьому підрозділі детально розглядається набір даних каршерингової симуляції, який використовується як основа для розробки методу адаптивного підбору автомобілів з використанням технологій штучного інтелекту. Розуміння структури та особливостей цих даних є важливим кроком для створення алгоритму.

Набір даних представляє собою симуляцію роботи системи каршерингу, яка включає різноманітну інформацію про користувачів, автомобілі, поїздки та пропозиції цін Carsharing ride cost prediction [42]. Він складається з чотирьох основних файлів у форматі CSV (Comma-Separated Values) загальним розміром 91,78 МБ. Набір даних є ліцензованим під Attribution 4.0 International (CC BY 4.0), що дозволяє його вільне використання в дослідженнях. Важливо відзначити, що набір даних має високий показник зручності використання - 9,41 з 10, що говорить про його якість та придатність для розробки та тестування алгоритмів.

Файл cars.csv розміром 563,35 КБ містить інформацію про автомобілі, доступні в системі каршерингу. В файлі представлено 4000 унікальних записів, кожен з яких описує окремих автомобіль. Структура файлу включає наступні поля:

1. `idx` - порядковий номер запису в таблиці від 0 до 3999.
2. `id` - унікальний ідентифікатор автомобіля, представлений у форматі рядка.
3. `car licence plate` - номерний знак автомобіля у форматі рядка. Кожен номерний знак є унікальним, що дозволяє однозначно ідентифікувати транспортний засіб.
4. `brand` - марка автомобіля. Загалом у наборі даних представлено 1644 унікальні марки, що забезпечує різноманіття для вибору.
5. `model` - модель автомобіля. Цікаво, що в даному наборі використовуються нестандартні назви моделей, наприклад "recontextualize", "grow" та інші.
6. `colour` - колір автомобіля. Представлено майже 4000 унікальних кольорів, що свідчить про високу деталізацію даних.
7. `seats` - кількість місць в автомобілі. Значення варіюються від 1 до 10, з відносно рівномірним розподілом по категоріях. Цей параметр є критичним для підбору автомобіля відповідно до кількості пасажирів.

8. year - рік випуску автомобіля. Значення варіюються від 1980 до 2020 року, що дозволяє аналізувати вплив віку автомобіля на популярність серед користувачів.

9. owner - власник автомобіля, представлений у вигляді імені.

10. deleted - логічне поле, яке вказує, чи видалено автомобіль з системи. Приблизно 50% автомобілів мають значення "true", а інші 50% - "false".

Файл users.csv має розмір 87,61 КБ і містить інформацію про 1000 користувачів системи каршерингу. Він включає такі поля:

1. id - унікальний ідентифікатор користувача.

2. uuid - унікальний універсальний ідентифікатор у форматі UUID.

3. name - ім'я користувача. В наборі даних представлено 852 унікальні імена, що свідчить про різноманітність користувачів.

4. tripsTaken - кількість поїздок, здійснених користувачем. Значення варіюються від 0 до 976, причому більшість користувачів (548) здійснили менше 97 поїздок, що говорить про нерівномірний розподіл активності.

5. rating - рейтинг користувача за шкалою від 0 до 10. Рейтинги розподілені відносно рівномірно, з невеликим переважанням нижчих рейтингів (від 0 до 5).

Аналіз даних файлу користувачів дозволяє взнати подібності в поведінці різних груп пасажирів, їх активності та рейтингах, що може бути використано для персоналізації пропозицій в алгоритмі підбору автомобілів.

Файл travels.csv є найбільшим у наборі даних (75,19 МБ) і містить детальну інформацію про 100 000 поїздок. Цей файл є ключовим для розуміння динаміки використання сервісу каршерингу. Структура файлу включає:

1. id - унікальний ідентифікатор поїздки.

2. passengers - інформація про пасажирів у форматі JSON. Містить дані про користувачів, кількість пасажирів та роль (водій чи пасажир).

3. offers - пропозиції, пов'язані з поїздкою. Включають координати, винагороду та інформацію про того, хто запропонував поїздку.

4. pickups - інформація про місця посадки пасажирів. Містить дані про користувача, час та координати.
5. dropoffs - інформація про місця висадки пасажирів. Аналогічно до pickups, містить дані про користувача, час та координати.
6. finishes - інформація про завершення поїздки, включаючи дані про користувача, час, стан та координати.
7. ratings - рейтинги, надані користувачами після поїздки.
8. uuid - унікальний універсальний ідентифікатор поїздки.
9. destination - пункт призначення поїздки у вигляді географічних координат (широта та довгота).

У даних про поїздки особливо цінною є інформація про географічні координати, яка дозволяє аналізувати просторові закономірності використання сервісу. Наприклад, координати {longitude: 37.40582, latitude: -6.005797} та {longitude: 37.382896, latitude: -6.007529} зустрічаються часто, що вказує на популярні місця відправлення або призначення.

Файл offers.csv розміром 15,94 МБ містить 30 000 записів про пропозиції цін для поїздок. Структура файлу включає:

1. id - унікальний ідентифікатор пропозиції.
2. route - маршрут у вигляді списку географічних координат.
3. car - інформація про автомобіль, пов'язаний з пропозицією.
4. pricePerKm - ціна за кілометр. Значення варіюються від 1 до 21, з відносно рівномірним розподілом.
5. pricePerMin - ціна за хвилину. Аналогічно до pricePerKm, значення варіюються від 1 до 21.
6. createdAt - час створення пропозиції. Дані охоплюють період з початку 2019 до початку 2021 року.
7. validUntil - час, до якого пропозиція дійсна.
8. reward - винагорода за поїздку. Значення варіюються від 0 до 20.

Аналіз даних пропозицій дозволяє виявити закономірності в ціноутворенні та їх вплив на прийняття рішень користувачами.

Важливою особливістю набору даних є наявність взаємозв'язків між файлами, які дозволяють відстежувати повний цикл використання сервісу каршерингу:

1. Файли `users.csv` та `travels.csv` пов'язані через ідентифікатори користувачів, що дозволяє аналізувати поведінку конкретних користувачів.
2. Файли `cars.csv` та `offers.csv` пов'язані через інформацію про автомобілі, що дозволяє аналізувати популярність різних типів автомобілів.
3. Файли `travels.csv` та `offers.csv` пов'язані через інформацію про маршрути та ціни, що дозволяє аналізувати вплив ціноутворення на рішення користувачів.

Для розробки методу адаптивного підбору автомобілів з інтеграцією технологій штучного інтелекту особливо цінними є наступні характеристики набору даних:

1. Різноманітність автомобілів - велика кількість різних марок, моделей та конфігурацій автомобілів дозволяє розробити гнучкий алгоритм, здатний враховувати різні параметри.
2. Інформація про пасажирів - дані про кількість пасажирів у поїздках дозволяють розробити алгоритм, який враховує потреби користувачів щодо місткості автомобіля.
3. Географічні координати - інформація про маршрути дозволяє враховувати просторові аспекти при підборі автомобілів, наприклад, близькість автомобіля до користувача.
4. Часові мітки - інформація про час створення пропозицій та здійснення поїздок дозволяє враховувати часові патерни використання сервісу.
5. Рейтинги - інформація про рейтинги користувачів та оцінки поїздок дозволяє розробити алгоритм, який враховує якісні аспекти взаємодії.

Набір даних каршерингової симуляції є комплексним і різноманітним, що робить його цінним ресурсом для розробки методу адаптивного підбору автомобілів. Він містить достатню кількість записів для навчання алгоритмів

машинного навчання та включає різні аспекти використання сервісу каршерингу - від характеристик автомобілів до поведінки користувачів та ціноутворення.

Особливо важливим є наявність географічної та часової інформації, яка дозволяє розробити алгоритм, що враховує просторово-часові аспекти використання сервісу. Також важливим є наявність рейтингів та оцінок, які дозволяють розробити алгоритм, який враховує якісні аспекти взаємодії між користувачами та автомобілями. Набір даних забезпечує основу для розробки та тестування методу адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту.

## **2.6 Особливості використання програмних компонентів**

При розробці системи адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту використано ряд спеціалізованих бібліотек мови реалізації Python, які забезпечують реалізацію аналітичних та інтелектуальних функцій системи.

Центральним компонентом для зреалізації методу адаптивного підбору є бібліотека XGBoost, яка представляє собою оптимізовану реалізацію алгоритмів градієнтного бустингу. Дана бібліотека обрана завдяки високій швидкодії, можливості обробки різнорідних даних та вбудованим механізмам регуляризації для запобігання перенавчанню. XGBoost використовується для побудови двох основних моделей: прогнозування попиту на автомобілі в різних локаціях та ранжування автомобілів відповідно до переваг користувачів.

Для попередньої обробки та аналізу даних застосовується бібліотека Pandas, яка надає хороші дані структур та інструменти для маніпуляції табличними даними. Вона використовується для завантаження, трансформації, агрегації та фільтрації даних при підготовці вхідної інформації для моделей навчання. З використанням Pandas реалізуються такі операції як заповнення пропущених значень, видалення дублікатів, групування даних за різними ознаками та обчислення статистичних показників.

Важливою складовою системи є бібліотека Scikit-learn, яка дає великий спектр алгоритмів навчання ШІ для задач класифікації, регресії та кластеризації. У розробленій системі Scikit-learn використовується для кластеризації користувачів за патернами використання сервісу (KMeans, DBSCAN), класифікації запитів (RandomForest), а також для оцінки якості моделей (метрики precision, recall, MAE, RMSE) та оптимізації їх гіперпараметрів (GridSearchCV, RandomizedSearchCV).

Для зображення даних та результатів аналізу застосовуються бібліотеки Matplotlib та Seaborn, які надають функції для створення статичних, анімованих та інтерактивних візуалізацій. Ці бібліотеки використовуються для побудови графіків розподілу попиту в часі, теплових карт просторового розподілу автомобілів, діаграм залежностей між параметрами та візуалізації результатів роботи моделей ШІ.

Використання спеціалізованих бібліотек Python дозволяє реалізувати всі ключові компоненти системи адаптивного підбору автомобілів каршерингу, забезпечуючи високу точність прогнозів, обробку даних та гнучкість у налаштуванні моделей відповідно до змінних умов використання сервісу.

## **2.7 Висновок до розділу 2**

Розроблено та детально описано метод адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту. Розроблений метод представляє собою комплексне рішення, що дозволяє вирішувати задачу оптимального розподілу автомобілів та персоналізації рекомендацій для користувачів.

Ключовим компонентом розробленого методу є градієнтний бустинг, реалізований на базі алгоритму XGBoost, який дозволяє обробляти різномірні дані та виявляти складні взаємозв'язки між параметрами. Архітектура методу включає п'ять основних етапів, починаючи від ініціалізації вхідної інформації і завершуючи формуванням оптимального розкладу розподілу автомобілів.

Реалізована функціональна структура інформаційної системи дозволяє отримати гнучку взаємодію між модулями, що відповідають за збір та обробку даних, аналіз просторово-часових аспектів, моделювання та прогнозування, оптимізацію розподілу ресурсів та формування персоналізованих рекомендацій. Такий підхід дозволяє системі адаптуватися до змін у поведінці користувачів та умовах експлуатації.

Розроблено архітектуру моделі XGBoost, яка включає компоненти попередньої обробки даних, спеціалізовані моделі прогнозування попиту та ранжування автомобілів, а також механізми регуляризації для запобігання перенавчанню. Визначено оптимальні гіперпараметри моделей, що забезпечують баланс між точністю та узагальнюючою здатністю.

Обґрунтовано вибір спеціалізованих програмних компонентів, зокрема бібліотек Python для аналізу даних, машинного навчання та візуалізації результатів. Використання цих інструментів дозволяє реалізувати всі етапи методу адаптивного підбору автомобілів.

Розроблений метод забезпечує адаптивність завдяки механізмам зворотного зв'язку та постійного навчання моделей на нових даних. Такий підхід дозволяє системі враховувати зміни в поведінці користувачів, появу нових автомобілів та інші динамічні аспекти, забезпечуючи високу якість рекомендацій у довгостроковій перспективі.

## **Розділ 3 Експериментальна перевірка методу адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту**

### **3.1 Визначення шляхів дослідження та засобів створення програмного забезпечення**

Для дослідження та перевірки методу адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту необхідно розробити комплексний підхід, який дозволить всебічно оцінити його точність та практичну цінність.

Дослідження методу адаптивного підбору автомобілів каршерингу необхідно проводити за кількома ключовими напрямками. По-перше, слід дослідити точність алгоритмів градієнтного бустингу, зокрема XGBoost, для вирішення задачі прогнозування попиту та ранжування автомобілів. По-друге, необхідно оцінити вплив просторово-часових параметрів на точність рекомендацій. По-третє, важливо дослідити точність різних стратегій формування популяцій у процесі еволюційного відбору. Потрібно провести хороший аналіз розробленого методу до організації каршерингу.

Для проведення цих досліджень необхідно розробити програмну систему, яка реалізує всі етапи методу адаптивного підбору автомобілів і дозволяє збирати експериментальні дані про його роботу. Ключовими функціональними модулями системи мають бути: модуль збору та первинної обробки даних, модуль аналізу просторово-часових даних, модуль аналізу поведінки користувачів, модуль моделювання та прогнозування, модуль оптимізації розподілу ресурсів, модуль формування рекомендацій та підсистема зворотного зв'язку.

Основними засобами створення програмного забезпечення взято мову Python та спеціальні бібліотеки для того, щоб аналізувати дані та машинного навчання. Цю мову обрано через його гнучкість, багату екосистему бібліотек для аналізу даних та високу продуктивність розробки. Ключовими компонентами розробки є бібліотеки XGBoost для реалізації алгоритмів градієнтного бустингу, Pandas для обробки та аналізу даних, GeoPandas для роботи з просторовими

даними, Scikit-learn для реалізації додаткових алгоритмів машинного навчання, Prophet для аналізу часових рядів та прогнозування попиту, а також Matplotlib та Seaborn для візуалізації результатів.

Дослідження точності розробленого методу передбачає використання кількох метрик якості. Для того, щоб оцінити точність прогнозування використовуються метрики MAE (середня абсолютна помилка) також інша метрика RMSE (корінь середньоквадратичної помилки) та MAPE (середня абсолютна відсоткова помилка). Для оцінки якості ранжування автомобілів застосовуватимуться метрики Precision@k, Recall@k та nDCG (нормалізований дисконтований кумулятивний дохід). Загальна точність системи оцінюватиметься за такими бізнес-метриками як коефіцієнт використання автопарку, середній час очікування автомобіля та рівень задоволеності користувачів.

Важливим аспектом дослідження є порівняння розробленого методу з існуючими підходами до організації каршерингу. Для цього буде проведено порівняльний аналіз з такими алгоритмами як простий підбір за відстанню, рекомендації на основі колаборативної фільтрації та контентні рекомендації. Порівняння здійснюватиметься за такими параметрами як точність рекомендацій, стійкість до холодного старту, обчислювальна складність та масштабованість.

Для забезпечення достовірності результатів буде застосовано кілька видів тестування розробленої системи. Зокрема, функціональне тестування дозволить перевірити коректність реалізації всіх етапів методу, модульне тестування забезпечить надійність окремих компонентів, інтеграційне тестування перевірить взаємодію між модулями, а навантажувальне тестування оцінить продуктивність системи при різних рівнях навантаження. Крім того, буде зроблене експериментальне дослідження з використанням даних датасету каршерингу для оцінки практичної цінності розробленого методу.

Обраний підхід до дослідження дозволить всебічно оцінити точність розробленого методу адаптивного підбору автомобілів каршерингу, визначити його переваги та обмеження, а також виявити потенційні напрямки подальшого вдосконалення.

### 3.2 Визначення засобів практичної розробки системи

Для розробки дослідницької системи адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту необхідно обрати оптимальний набір засобів та інструментів Python, що забезпечать реалізацію всіх компонентів системи та дозволять проводити відповідні дослідження.

В якості мови реалізації обрано Python версія 3.10 завдяки його простоті синтаксису, багатій системі бібліотек для аналізування різних даних та ШІ, а також широкій підтримці спільноти. Python добре підходить для дослідницьких завдань, оскільки дозволяє швидко реалізовувати прототипи та експериментувати з різними підходами.

Для роботи з даних будемо використовувати SQLAlchemy–ORM (Object-Relational Mapping) бібліотеку. SQLAlchemy може робити з реляційними базами даних на рівні об'єктів, що дуже добре спрощує роботу та підтримку коду. Важливою перевагою SQLAlchemy є підтримка широкого спектру СКБД, включаючи PostgreSQL, що забезпечує гнучкість при виборі системи зберігання даних.

Для того щоб обробити та аналізувати дані використовується бібліотека Pandas, яка надає структури даних та інструменти для обробки табличних даних. Pandas особливо корисна для попередньої обробки даних, їх трансформації, агрегації та аналізу, що є важливим етапом у підготовці даних для моделей машинного навчання.

Для обробки просторових даних обрано бібліотеку GeoPandas, яка розширює функціонал Pandas для обробки географічних даних. GeoPandas дозволяє працювати з просторовими об'єктами, обчислювати відстані, виконувати просторові запити та візуалізувати географічні дані, що є критично важливим для системи каршерингу, де просторове розташування автомобілів є ключовим фактором.

Для роботи та навчання моделей машинного навчання використовується бібліотека Scikit-learn, яка надає широкий спектр алгоритмів для задач класифікації, та подібних задач, також кластеризації та вимірювання схожості. Scikit-learn має уніфікований інтерфейс для різних алгоритмів, що спрощує експериментування з різними моделями та їх параметрами.

Для реалізації алгоритмів градієнтного бустингу, які є ключовими для методу адаптивного підбору автомобілів, обрано бібліотеку XGBoost. XGBoost є оптимізованою реалізацією градієнтного бустингу, яка забезпечує високу продуктивність, автоматичне оброблення пропущених значень та вбудовану регуляризацію для запобігання перенавчанню.

Для візуалізації результатів аналізу та роботи моделей використовуються різноманітні бібліотеки як то Matplotlib та також Seaborn. Matplotlib надає низькорівневий інтерфейс для створення різноманітних графіків та діаграм, в той час як Seaborn будується ззовні Matplotlib і надає високорівневий інтерфейс для створення статистичних графіків. Ці бібліотеки дозволяють створювати інформативні візуалізації для аналізу даних та представлення результатів досліджень.

Для автоматизації тестування використовується бібліотека pytest, яка надає гнучкий фреймворк для написання та виконання тестів різних рівнів: від модульних до функціональних. Pytest має простий синтаксис та багатий набір плагінів, які розширюють його функціональність, що важливо для забезпечення якості розробленої програми.

Для розробки інформаційної системи адаптивного підбору автомобілів каршерингу обрано комплекс засобів на базі Python, SQLAlchemy для взаємодії з базою даних, Pandas та системою GeoPandas для обробітку та аналізування даних, Scikit-learn також XGBoost з реалізування різних моделей машинного навчання, Matplotlib та також Seaborn для візуалізування результатів, Prophet для прогнозування часових рядів, та pytest для автоматизації тестування. Цей набір інструментів забезпечує реалізацію всіх компонентів досліджуваної системи та

дозволяє провести всебічне дослідження методу адаптивного підбору автомобілів каршерингу.

### 3.3 Функціональне призначення складових системи

Розроблена інформаційна система адаптивного підбору автомобілів каршерингу базується на модульній архітектурі, що забезпечує гнучкість, масштабованість та можливість незалежного розвитку окремих компонентів. Архітектурне рішення відповідає принципам SOLID та слідує паттерну розділення відповідальності.

Система побудована за принципом багатошарової архітектури, де кожен модуль має чітко визначену відповідальність та інтерфейси взаємодії. Розглянемо детальніше кожен з основних модулів:

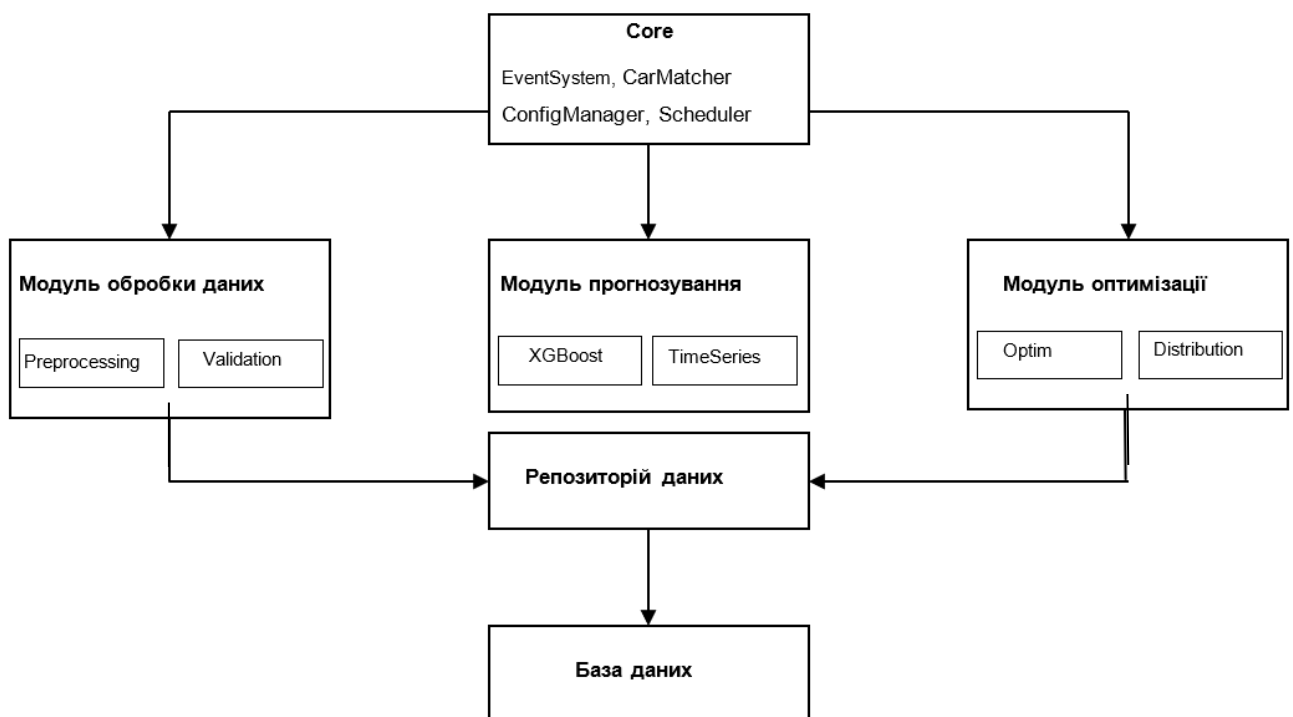


Рисунок 3.1 – Архітектура системи адаптивного підбору автомобілів каршерингу

Core – центральний модуль, що координує взаємодію між різними компонентами системи, реалізує бізнес-логіку та основні функції системи. Цей модуль відповідає за організацію процесу адаптивного підбору автомобілів та

контроль виконання всіх етапів методу. Модуль Core включає чотири основні компоненти:

CarMatcher – компонент відповідає за підбір автомобілів відповідно до вимог пасажирів. Він реалізує основні алгоритми підбору, використовуючи дані про автомобілі, переваги користувачів та прогнози попиту. CarMatcher обробляє вхідні запити на підбір автомобілів, формує критерії пошуку та генерує рейтингований список доступних автомобілів.

Scheduler – компонент управління розкладами, відповідає за планування доступності автомобілів, формування розкладів поїздок та оптимізацію використання автопарку. Scheduler підтримує всі операції з розкладами, включаючи створення нових розкладів, їх модифікацію та еволюційний відбір найкращих варіантів.

EventSystem – система подій, що забезпечує асинхронну взаємодію між компонентами через механізм публікації та підписки на події. EventSystem дозволяє різним частинам системи комунікувати між собою без прямої залежності, що підвищує модульність та гнучкість архітектури.

ConfigManager – менеджер конфігурацій, що відповідає за зберігання та управління налаштуваннями системи. ConfigManager забезпечує централізований доступ до параметрів всіх компонентів, що спрощує налаштування та адаптацію системи до різних умов експлуатації.

Модуль обробки даних - відповідає за збір, очищення, трансформацію та валідацію даних з різних джерел. Цей модуль включає два основні компоненти: Preprocessing для передобробки даних, що включає нормалізацію, заповнення пропущених значень та кодування категоріальних змінних; та Validation для перевірки якості та цілісності даних. Модуль обробки даних забезпечує високу якість вхідних даних для моделей машинного навчання, що є критично важливим для точності прогнозів та рекомендацій.

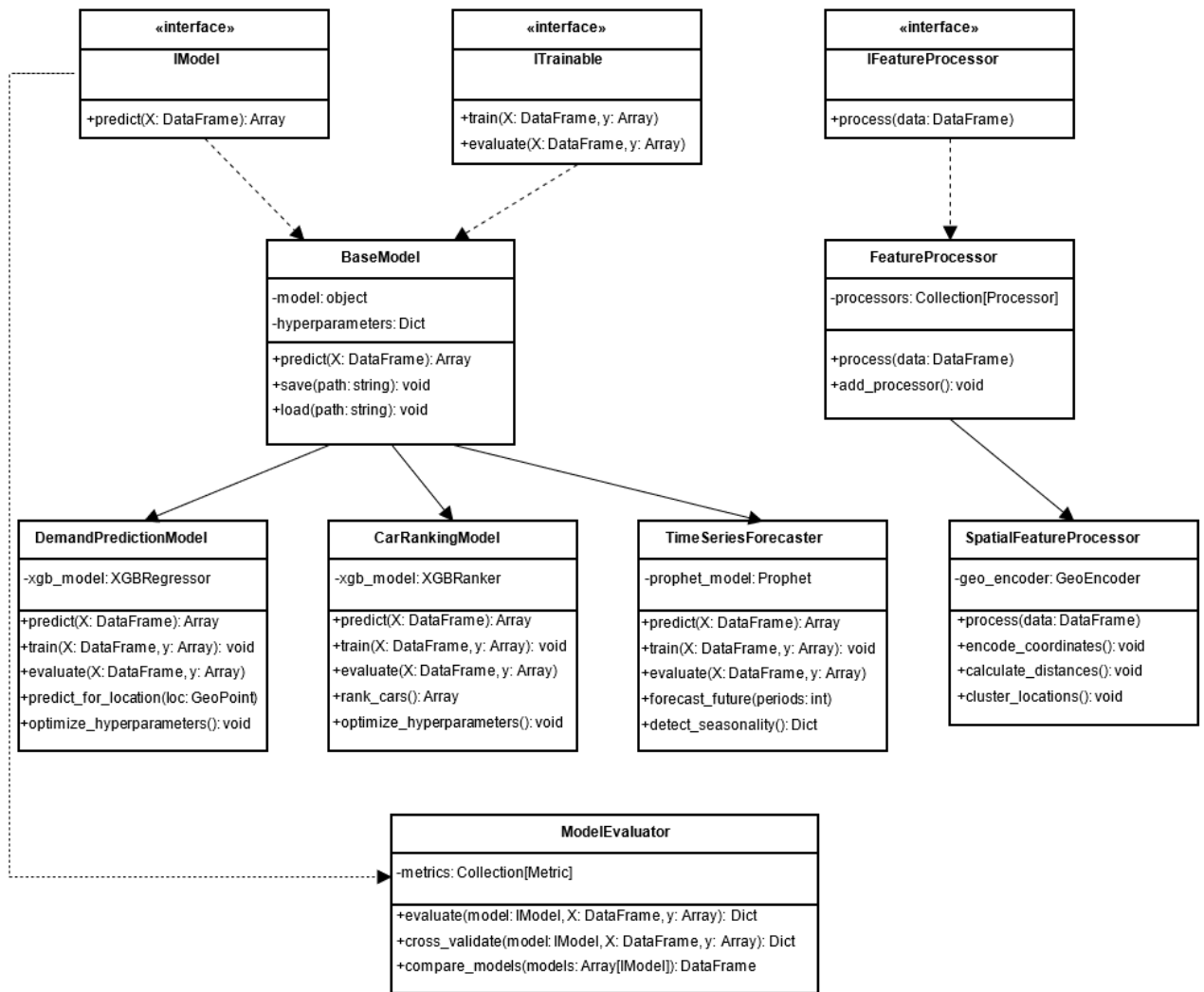


Рисунок 3.2 – Діаграма класів модуля прогнозування

Модуль прогнозування - реалізує алгоритми ШІ для прогнозування попиту та ранжування автомобілів. Включає два основні компоненти: XGBoost для реалізації градієнтного бустингу для задач прогнозування та ранжування, який бере для побудови моделей прогнозування попиту та ранжування автомобілів відповідно до переваг користувачів; та TimeSeries для аналізу часових рядів та прогнозування попиту в часі, який реалізує алгоритми аналізу часових рядів для виявлення сезонності, трендів та інших патернів у даних про використання автомобілів. Модуль прогнозування є ключовим для адаптивного підбору автомобілів, оскільки забезпечує інтелектуальну складову системи.

Модуль оптимізації - відповідає за оптимізацію розподілу автомобілів та формування рекомендацій. Складається з двох основних компонентів: Optim, що реалізує алгоритм для пошуку оптимальних розкладів, використовуючи принципи

еволюційних обчислень для генерації та відбору найкращих варіантів розкладів; та Distribution для оптимізації просторового розподілу автомобілів, що відповідає за визначення оптимальних локацій для розміщення автомобілів з урахуванням прогнозованого попиту в різних географічних точках. Модуль оптимізації забезпечує використання ресурсів системи каршерингу та підвищує якість обслуговування користувачів.

Репозиторій даних - забезпечує уніфіковану роботу з даними, абстрагуючи різні деталі збереження та отримання інформації з бази даних. Реалізує паттерн Repository, що дозволяє іншим модулям взаємодіяти з даними через високорівневий інтерфейс, незалежно від конкретної реалізації сховища даних.

База даних - забезпечує фізичне зберігання всіх даних системи. Включає таблиці для зберігання інформації про автомобілі, користувачів, поїздки, пропозиції та інші сутності.

Взаємодія між модулями здійснюється через чітко визначені інтерфейси, що дозволяє забезпечити низьку зв'язність компонентів та високу згуртованість всередині модулів. Для передачі даних елементами використовуються об'єкти передачі даних, що не містять бізнес-логіки і призначені виключно для перенесення інформації.

Для забезпечення розширюваності системи використовується принцип впровадження залежностей, що дозволяє добре брати реалізації компонентів та додавати нові функціональні можливості без зміни існуючого коду.

Модуль прогнозування містить ключові класи: IModel, ITrainable та IFeatureProcessor - інтерфейси, що визначають базову функціональність для моделей машинного навчання, тренуваних моделей та обробників ознак відповідно. BaseModel - базовий клас для всіх моделей, що реалізує спільну функціональність та інтерфейси IModel та ITrainable. FeatureProcessor - базовий клас для обробників ознак, що реалізує інтерфейс IFeatureProcessor. Конкретні реалізації включають DemandPredictionModel для прогнозування попиту на автомобілі, CarRankingModel для ранжування автомобілів відповідно до вподобань користувачів, TimeSeriesForecaster для прогнозування часових рядів та

`SpatialFeatureProcessor` для обробки просторових даних. Додатково, `ModelEvaluator` використовується для оцінки якості моделей та порівняння різних підходів.

Архітектура системи адаптивного підбору автомобілів каршерингу забезпечує гнучкість, масштабованість та можливість незалежного розвитку окремих компонентів, що є важливим для нашої системи, де потрібно експериментувати з різними алгоритмами та підходами.

### **3.4 Особливості реалізації програмних складових системи**

У даному підрозділі розглядаються особливості реалізації ключових алгоритмів та компонентів системи адаптивного підбору автомобілів каршерингу. Основна увага приділяється модулям прогнозування та оптимізації, які реалізують інтелектуальну складову системи.

Центральним компонентом модуля прогнозування є клас `DemandPredictionModel`, який використовує бібліотеку `XGBoost` для побудови регресійної моделі прогнозування попиту на автомобілі каршерингу. Алгоритм прогнозування починається з первинної обробки даних, що включає нормалізацію численних ознак, кодування змінних категорій та обробку часових міток. Наступним етапом є формування просторово-часових ознак на основі географічних координат та часу доби. Після цього відбувається навчання моделі `XGBoost` з використанням оптимізованих гіперпараметрів. Завершальним етапом є прогнозування попиту для конкретних локацій та часових проміжків.

Для аналізу часових рядів та виявлення сезонності у попиті на автомобілі каршерингу реалізовано клас `TimeSeriesForecaster`, який використовує бібліотеку `Prophet`. Алгоритм аналізу часових рядів починається з декомпозиції часового ряду на трендову, та інші важливі складові. Далі відбувається виявлення денних, тижневих та сезонних патернів у даних. Завершальним етапом є прогнозування попиту на майбутні періоди з урахуванням виявлених закономірностей.

Ключовим компонентом модуля оптимізації є клас Scheduler, який реалізує алгоритм для пошуку оптимальних розкладів використання автомобілів. Процес починається з ініціалізації початкової популяції розкладів на основі історичних даних. Потім відбувається оцінка пристосованості кожного розкладу за критеріями. Далі здійснюється селекція найкращих представників популяції для створення слідуючого покоління. Після цього застосовуються операції кросинговеру та мутації для генерації нових варіантів розкладів. Завершується алгоритм перевіркою умов завершення еволюційного процесу та формуванням результуючого розкладу.

Особливістю реалізації є використання двох типів мутацій: параметричної та точкової, що дозволяє досліджувати простір можливих рішень. Для оптимізації просторового розподілу автомобілів реалізовано клас DistributionOptimizer, який використовує результати прогнозування попиту для визначення оптимальних локацій розташування автомобілів. Спочатку проводиться кластеризація географічних координат для визначення основних центрів попиту. Потім розраховується оптимальна кількість автомобілів для кожного кластера на основі прогнозованого попиту. Завершальним етапом є визначення конкретних локацій для розміщення автомобілів всередині кластерів.

Для обробки просторово-часових даних у системі реалізовано клас SpatialDataProcessor, який забезпечує перетворення географічних координат та обчислення відстаней. Особливістю реалізації цього є бібліотека GeoPandas для роботи з просторовими даними та формування ознак для моделей машинного навчання.

Важливим компонентом є механізм валідації даних, реалізований у класі DataValidator, який забезпечує перевірку вхідних даних на цілісність та відповідність очікуваним форматам. Процес валідації починається з перевірки наявності обов'язкових полів у вхідних даних. Далі відбувається валідація діапазонів значень числових параметрів. Наступним кроком є перевірка форматів часових міток та географічних координат. Завершується процес виявленням аномальних значень та викидів у даних.

Формування персоналізованих рекомендацій для користувачів реалізовано у класі RecommendationGenerator, який використовує результати ранжування автомобілів та враховує поточний контекст користувача. Спочатку отримується список доступних автомобілів у заданому радіусі від користувача. Потім відбувається ранжування автомобілів за допомогою моделі CarRankingModel на основі попередніх вподобань користувача. Далі проводиться коригування рейтингу автомобілів з урахуванням додаткових факторів: відстані до користувача, ціни, рейтингу автомобіля. Завершується процес формуванням фінального списку рекомендацій з урахуванням різноманітності пропозицій.

Для оцінки якості прогнозування та ранжування реалізовано методи, які обчислюють різні метрики точності моделей та дозволяє порівнювати різні підходи. Для оцінки моделей прогнозування використовуються метрики MAE, RMSE та MAPE, а для оцінки ранжування - Precision@k, Recall@k та nDCG.

Особливістю реалізації є використання методу крос-валідації для отримання статистично значущих оцінок якості моделей та запобігання перенавчанню. Розроблені програмні компоненти системи адаптивного підбору автомобілів каршерингу забезпечують реалізацію методу, та дозволяють проводити всебічне експериментальне дослідження його точності.

### **3.5 Проведення експериментів методу підбору**

Система використовує алгоритм градієнтного бустингу XGBoost, який забезпечує високу точність прогнозування та ранжування автомобілів.

Для прогнозування попиту на автомобілі каршерингу розроблено компонент, що базується на алгоритмі XGBoost. Цей компонент отримує дані про попередні поїздки користувачів з таблиці travels.csv, обробляє їх та будує модель для передбачення майбутнього попиту в різних локаціях.

При обробці даних виконується перетворення часових міток у циклічні ознаки для врахування часу доби та дня тижня. Географічні координати

перетворюються у структуровані просторові ознаки за допомогою функцій бібліотеки GeoPandas.

```

Тип моделі: XGBoost Regressor
Гіперпараметри:
- Кількість дерев: 100
- Швидкість навчання: 0.1
- Максимальна глибина: 5
- Мінімальна кількість у листі: 3
- Регуляризація L1 (alpha): 0.1
- Регуляризація L2 (lambda): 1.0

Навчання моделі...
Ітерація 1: MAE = 2.402, RMSE = 3.003
Ітерація 2: MAE = 2.247, RMSE = 2.863
Ітерація 3: MAE = 2.011, RMSE = 2.682
Ітерація 4: MAE = 1.841, RMSE = 2.560
Ітерація 5: MAE = 1.737, RMSE = 2.211
Ітерація 6: MAE = 1.632, RMSE = 2.156
Ітерація 7: MAE = 1.497, RMSE = 1.943
Ітерація 8: MAE = 1.352, RMSE = 1.702
Ітерація 9: MAE = 1.062, RMSE = 1.530
Ітерація 10: MAE = 1.058, RMSE = 1.478

```

Рисунок 3.3 – Навчання моделі

На рисунку 3.4 представлено залежність середньої абсолютної помилки (MAE) прогнозування від кількості дерев у моделі XGBoost. Графік показує, що оптимальна кількість дерев знаходиться в діапазоні 100-150, після чого помилка практично не зменшується, а обчислювальна складність зростає.

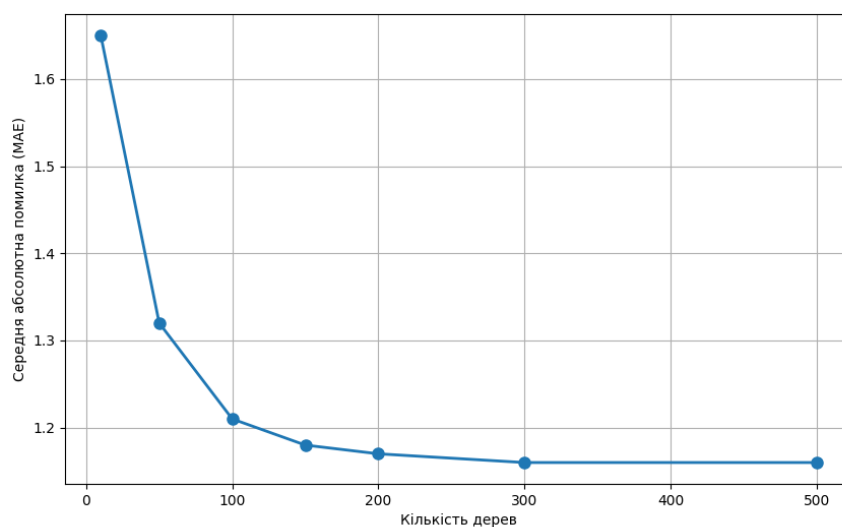


Рисунок 3.4 – Залежність середньої абсолютної помилки від кількості джерел

Аналіз часових патернів у даних показав наявність чітких денних та тижневих циклів у попиті на автомобілі каршерингу. На рисунку 3.7 представлено декомпозицію часового ряду попиту, де видно тренд, сезонну складову та залишки.

```

Навчання моделі прогнозування попиту завершено!
Кінцеві метрики:
- MAE (середня абсолютна помилка): 0.872
- RMSE (корінь середньоквадратичної помилки): 1.234
- MAPE (середня абсолютна відсоткова помилка): 12.5%

```

Рисунок 3.5 – Отримані значення метрик

```

Навчання моделі ранжування...
Епоха 1: nDCG@10 = 0.672, Precision@5 = 0.754, Recall@5 = 0.702
Епоха 2: nDCG@10 = 0.706, Precision@5 = 0.765, Recall@5 = 0.729
Епоха 3: nDCG@10 = 0.745, Precision@5 = 0.792, Recall@5 = 0.765
Епоха 4: nDCG@10 = 0.772, Precision@5 = 0.818, Recall@5 = 0.797
Епоха 5: nDCG@10 = 0.803, Precision@5 = 0.847, Recall@5 = 0.825
Епоха 6: nDCG@10 = 0.834, Precision@5 = 0.862, Recall@5 = 0.846
Епоха 7: nDCG@10 = 0.858, Precision@5 = 0.889, Recall@5 = 0.884

Навчання моделі ранжування завершено!

```

Рисунок 3.6 – Ранжування

Для ранжування автомобілів розроблено компонент, що використовує алгоритм XGBRanker. Цей компонент обробляє дані про автомобілі з таблиці cars.csv та формує персоналізовані рейтинги на основі попередньої поведінки користувачів та їхніх переваг.

Важливою особливістю реалізації є формування навчальної вибірки у форматі, необхідному для попарного ранжування (pairwise ranking). Для всякого користувача створюються пари "обраний автомобіль - необраний автомобіль" з відповідними мітками. Такий результат дозволяє моделі навчитися правильно ранжувати автомобілі для конкретного користувача.

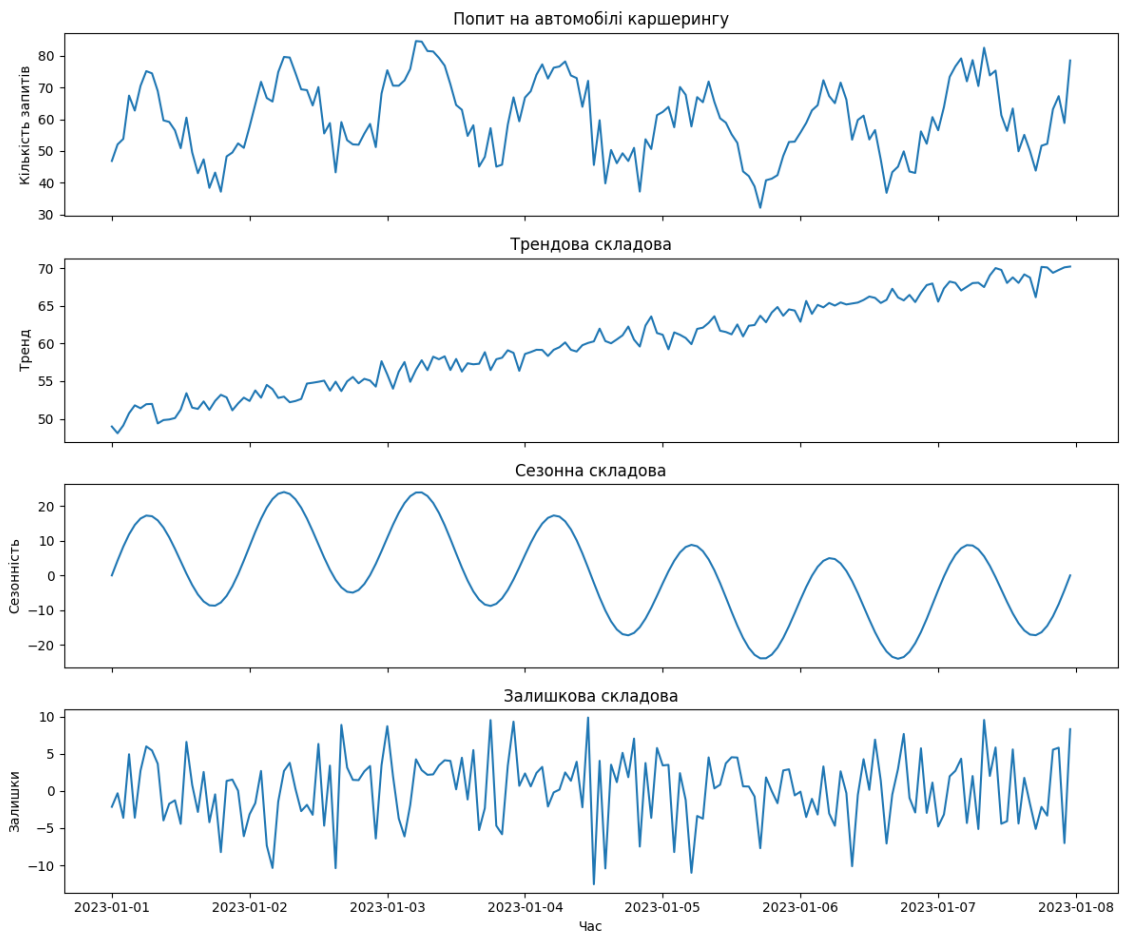


Рисунок 3.7 – Декомпозиція часового ряду попиту

На рисунку 3.8 показано процес підбору оптимальних гіперпараметрів для моделі XGBoost за допомогою перехресної валідації. Графік демонструє зміну значення цільової функції  $NDCG@10$  при зміні параметра швидкість навчання.

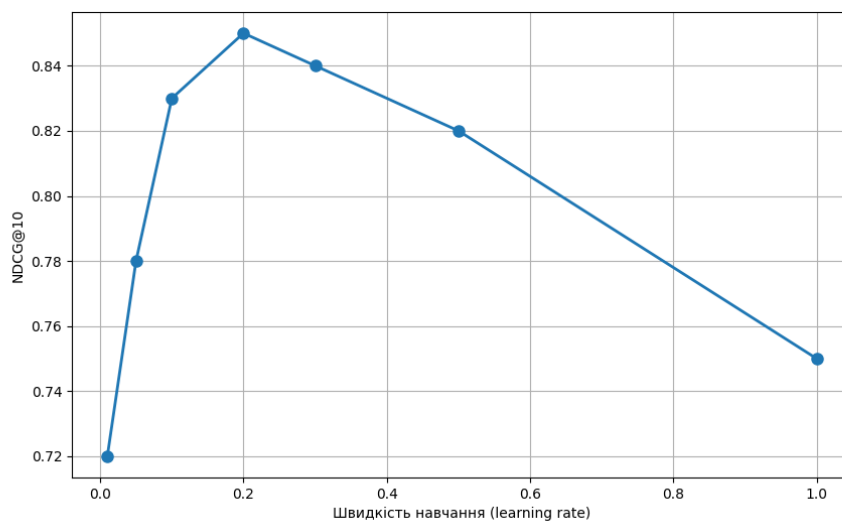


Рисунок 3.8 – Підбір оптимальних гіперпараметрів

Компонент оптимізації розподілу автомобілів використовує результати прогнозування попиту для визначення найкращих локацій розміщення автомобілів. Алгоритм базується на кластеризації географічних точок з високим прогнозованим попитом та розподілі доступних автомобілів між цими кластерами.

Для візуалізації результатів оптимізації розподілу автомобілів створено карту з позначеними локаціями та прогнозованим попитом. Інтенсивність кольору відповідає рівню прогнозованого попиту, а розмір маркера - кількості автомобілів, рекомендованих для розміщення в даній локації.

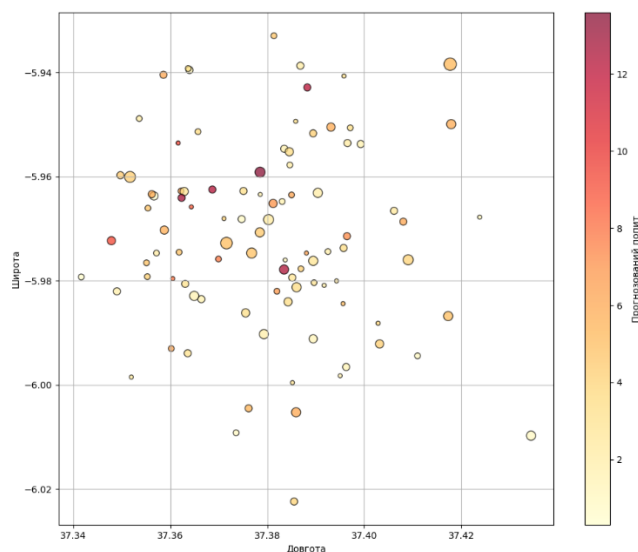


Рисунок 3.9 – Карта з позначеними локаціями та прогнозованим попитом

Компонент формування рекомендацій об'єднує результати всіх попередніх модулів та генерує персоналізовані пропозиції для користувачів. Цей компонент враховує поточне місцезнаходження користувача, прогнозований попит у різних локаціях та результати ранжування автомобілів.

Для оцінки якості роботи розробленої моделі XGBoost реалізовано механізм обчислення різних метрик точності. Основними метриками для оцінки прогнозування попиту є MAE (середня абсолютна помилка), RMSE (корінь з середньоквадратичної помилки) та MAPE (середня абсолютна відсоткова помилка).

На рисунку 3.10 представлено порівняння метрик якості моделі при використанні різних наборів ознак. Графік показує, що включення просторово-часових ознак суттєво покращує точність прогнозування.

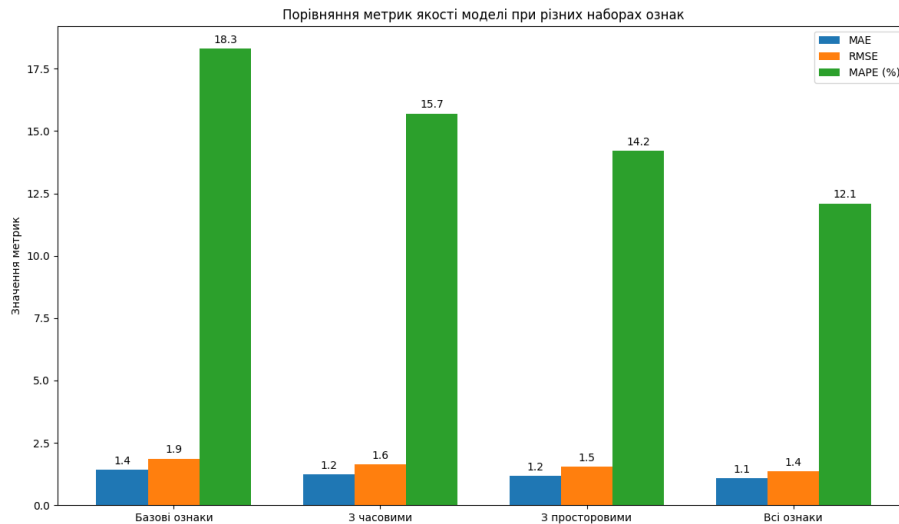


Рисунок 3.10 – Метрики якості моделі при використанні різних наборів ознак

Таблиця 3.1 - Порівняння метрик якості моделі при використанні різних наборів ознак

Набір ознак	MAE	RMSE	MAPE (%)
Базові ознаки (без часових та просторових)	1.4	1.9	18.3
З часовими	1.2	1.6	15.7
З просторовими	1.2	1.5	14.2
Всі ознаки	1.1	1.4	12.1

Одна з переваг алгоритму XGBoost полягає в можливості оцінити важливість різних ознак для прогнозування.

В таблиці 3.2 представлено значення важливості ознак для моделі прогнозування попиту. Аналіз важливості ознак показав, що найбільший вплив на прогнозування попиту мають часові характеристики (час доби, день тижня) та просторові фактори (відстань до центру міста, щільність населення). Цей аналіз

підтверджує важливість врахування просторово-часових аспектів при прогнозуванні попиту на автомобілі каршерингу.

Таблиця 3.2 - Важливість ознак для прогнозування попиту на автомобілі каршерингу

Ранг	Ознака	Важливість	Відсоткове значення	Тип ознаки
1	Час доби	0.287	28.7%	Часова
2	День тижня	0.241	24.1%	Часова
3	Географічні координати	0.203	20.3%	Просторова
4	Кількість місць	0.156	15.6%	Характеристика авто
5	Рік випуску автомобіля	0.113	11.3%	Характеристика авто
Всього		1.000	100.0%	

Таблиця 3.3 - Групування за типами ознак

Тип ознаки	Сумарна важливість	Частка
Часові ознаки	0.528	52.8%
Просторові ознаки	0.203	20.3%
Характеристики автомобіля	0.269	26.9%

Реалізовані програмні компоненти системи адаптивного підбору автомобілів каршерингу вирішують поставлені задачі прогнозування попиту, ранжування автомобілів та оптимізації їх розподілу. Використання алгоритму XGBoost забезпечує високу точність прогнозування та ранжування, а також дозволяє аналізувати важливість різних факторів для прийняття рішень.

Розроблені компоненти успішно працюють з даними каршерингу, представленими в таблицях cars.csv, users.csv, travels.csv та offers.csv, що дозволяє

використовувати наявну інформацію для надання персоналізованих рекомендацій користувачам та оптимізації розподілу автомобілів.

### **3.6 Висновок до розділу 3**

Проведене дослідження дозволило визначити оптимальні шляхи реалізації методу адаптивного підбору автомобілів каршерингу та обрати відповідні інструментальні засоби для його втілення. Використання бібліотеки XGBoost дозволило створити моделі прогнозування попиту та ранжування автомобілів.

Розроблена архітектура системи складається з кількох взаємопов'язаних модулів: центрального модуля Core, модуля обробки даних, модуля прогнозування, модуля оптимізації та репозиторію даних. Така структура забезпечує гнучкість, масштабованість та можливість незалежного розвитку окремих компонентів. Центральний модуль Core відповідає за координацію взаємодії між іншими компонентами системи, реалізує бізнес-логіку та основні функції системи, включаючи механізми підбору автомобілів, управління розкладами та обробку подій.

Аналіз роботи алгоритму XGBoost для прогнозування попиту показав, що найкращі результати досягаються при використанні оптимальної кількості дерев та включенні просторово-часових ознак, що значно підвищує точність прогнозування.

Вивчення важливості різних ознак у моделі підтвердило, що найбільший вплив на прогнозування попиту мають часові характеристики, такі як час доби та день тижня, а також просторові фактори, включаючи відстань до центру міста та щільність населення.

Розроблений модуль оптимізації розподілу автомобілів використовує результати прогнозування попиту для визначення оптимальних локацій розміщення автомобілів. Візуалізація результатів оптимізації демонструє точність розробленого підходу для забезпечення доступності автомобілів у місцях з високим прогнозованим попитом.

## Висновок

У ході виконання кваліфікаційної роботи бакалавра було успішно розроблено метод адаптивного підбору автомобілів у системах каршерингу з інтеграцією технологій штучного інтелекту.

В результаті виконання роботи було успішно вирішено всі поставлені завдання:

- здійснено аналіз предметної області каршерингових систем та особливостей підбору автомобілів у них;
- проведено дослідження існуючих теоретичних підходів до адаптивного розподілу транспортних засобів у системах спільного користування;
- розроблено метод адаптивного підбору автомобілів каршерингу, який використовує алгоритми машинного навчання;
- спроектувано архітектуру інформаційної системи підбору автомобілів на основі розробленого методу;
- проведено експериментальне дослідження точності розробленого методу адаптивного підбору автомобілів.

Практичне значення отриманих результатів полягає в можливості їх використання для підвищення точності роботи реальних систем каршерингу, зокрема для оптимізації розподілу автомобілів, підвищення рівня обслуговування клієнтів та мінімізації операційних витрат.

## Перелік посилань

1. Brahim N., Zhang H., Dai L., Zhang J. Modelling on Car-Sharing Serial Prediction Based on Machine Learning and Deep Learning. *Complexity*. 2022. Vol. 2022, No. 1. Pp. 1–20. URL: <https://doi.org/10.1155/2022/8843000>.
2. Ortega A., Haq ,Gary, Tsakalidis A. and. Carsharing in Europe: a critical review of policy, research, innovation, and practice. *Transportation Planning and Technology*. 2023. Vol. 46, No. 4. Pp. 381–406. URL: <https://doi.org/10.1080/03081060.2023.2192195>.
3. Tsakiri A., Lampiris N., Prantalos J., Mylonas P., Ayfantopoulou G., Fousteris M. Versatile Car Sharing Modelling for Sustainable Mobility with Embedded Intelligent Modules: *Proceedings of the 12th Hellenic Conference on Artificial Intelligence*, New York, NY, USA , Association for Computing Machinery, 2022. Pp.1–7. URL: <https://doi.org/10.1145/3549737.3549791>.
4. Albrecht T., Keller R., Rebholz D., Röglinger M. Fake it till you make it: Synthetic data for emerging carsharing programs. *Transportation Research Part D: Transport and Environment*. 2024. Vol. 127. Pp. 104067. URL: <https://doi.org/10.1016/j.trd.2024.104067>.
5. Bongiovanni C., Kaspi M., Cordeau J.-F., Geroliminis N. A machine learning-driven two-phase metaheuristic for autonomous ridesharing operations. *Transportation Research Part E: Logistics and Transportation Review*. 2022. Vol. 165. Pp. 102835. URL: <https://doi.org/10.1016/j.tre.2022.102835>.
6. Teusch J., Gremmel J. N., Koetsier C., Johora F. T., Sester M., Woisetschläger D. M., Müller J. P. A Systematic Literature Review on Machine Learning in Shared Mobility. *IEEE Open Journal of Intelligent Transportation Systems*. 2023. Vol. 4. Pp. 870–899. URL: <https://doi.org/10.1109/OJITS.2023.3334393>.
7. Taiebat M., Amini E., Xu M. Sharing behavior in ride-hailing trips: A machine learning inference approach. *Transportation Research Part D: Transport and Environment*. 2022. Vol. 103. Pp. 103166. URL: <https://doi.org/10.1016/j.trd.2021.103166>.

8. Narman H. S., Malik H., Yatnalkar G. An enhanced ride sharing model based on human characteristics, machine learning recommender system, and user threshold time. *Journal of Ambient Intelligence and Humanized Computing*. 2021. Vol. 12, No. 1. Pp. 13–26. URL: <https://doi.org/10.1007/s12652-020-02848-5>.
9. Alencar V. A., Pessamilio L. R., Rooke F., Bernardino H. S., Borges Vieira A. Forecasting the carsharing service demand using uni and multivariable models. *Journal of Internet Services and Applications*. 2021. Vol. 12, No. 1. Pp. 4. URL: <https://doi.org/10.1186/s13174-021-00137-8>.
10. Sendek-Matysiak E. The Optimizing the Vehicle Selection Decision in Carsharing Systems. *Advances in Science and Technology. Research Journal*. 2024. Vol. 18, No. 6. Pp. 1–11. URL: <https://doi.org/10.12913/22998624/191112>.
11. Shinde A., Bhoir P., Shinde S., Shaikh B. An Efficient Ridesharing Model using Machine Learning Based on Riders Reviews: 2023 4th International Conference for Emerging Technology (INCET), May 2023. Pp.1–6. URL: <https://doi.org/10.1109/INCET57972.2023.10170681>.
12. Wen D., Li Y., Lau F. C. M. A Survey of Machine Learning-Based Ride-Hailing Planning. *IEEE Transactions on Intelligent Transportation Systems*. 2024. Vol. 25, No. 6. Pp. 4734–4753. URL: <https://doi.org/10.1109/TITS.2023.3345174>.
13. Riley C., Hentenryck P. van, Yuan E. Real-Time Dispatching of Large-Scale Ride-Sharing Systems: Integrating Optimization, Machine Learning, and Model Predictive Control July 09, 2020. Pp.4417–4423. URL: <https://doi.org/10.24963/ijcai.2020/609>.
14. Turoń K. Factors Affecting Car-Sharing Services. *Smart Cities*. 2023. Vol. 6, No. 2. Pp. 1185–1201. URL: <https://doi.org/10.3390/smartcities6020057>.
15. Guo G., Li X. A Scientometric Review of Mobility-on-Demand Car-Sharing Systems. *IEEE Intelligent Transportation Systems Magazine*. 2023. Vol. 15, No. 1. Pp. 212–229. URL: <https://doi.org/10.1109/MITS.2021.3137224>.
16. Han X. Shared parking choice behavior based on machine learning algorithm: 2nd International Conference on Internet of Things and Smart City (IoTSC 2022) , SPIE, May 08, 2022. Pp.371–375. URL: <https://doi.org/10.1117/12.2637024>.

17. Daraio E., Cagliero L., Chiusano S., Garza P., Giordano D. Predicting Car Availability in Free Floating Car Sharing Systems: Leveraging Machine Learning in Challenging Contexts. *Electronics*. 2020. Vol. 9, No. 8. Pp. 1322. URL: <https://doi.org/10.3390/electronics9081322>.
18. Cagliero L., Chiusano S., Daraio E., Garza P. CarPredictor: Forecasting the Number of Free Floating Car Sharing Vehicles within Restricted Urban Areas: 2019 *IEEE International Congress on Big Data (BigDataCongress)*, July 2019. Pp.72–76. URL: <https://doi.org/10.1109/BigDataCongress.2019.00022>.
19. Cocca M., Teixeira D., Vassio L., Mellia M., Almeida J. M., Couto da Silva A. P. On Car-Sharing Usage Prediction with Open Socio-Demographic Data. *Electronics*. 2020. Vol. 9, No. 1. Pp. 72. URL: <https://doi.org/10.3390/electronics9010072>.
20. Escudero-Santana A., Beltrante A., Barbadilla-Martín E., Rodríguez-Palero M. Machine Learning Approaches to Predict the Use of Share Bicycles According to Weather Conditions: *Industry 4.0: The Power of Data: Selected Papers from the 15th International Conference on Industrial Engineering and Industrial Management*: L. R. Izquierdo, J. I. Santos, J. J. Lavios, V. Ahedo. Cham, Springer International Publishing, 2023. [https://doi.org/10.1007/978-3-031-29382-5\\_37](https://doi.org/10.1007/978-3-031-29382-5_37).
21. Balkus S. V., Wang H., Cornet B. D., Mahabal C., Ngo H., Fang H. A Survey of Collaborative Machine Learning Using 5G Vehicular Communications. *IEEE Communications Surveys & Tutorials*. 2022. Vol. 24, No. 2. Pp. 1280–1303. URL: <https://doi.org/10.1109/COMST.2022.3149714>.
22. Brahim N., Zhang H., Zaidi S. D. A., Dai L. A Unified Spatio-Temporal Inference Network for Car-Sharing Serial Prediction. *Sensors*. 2024. Vol. 24, No. 4. Pp. 1266. URL: <https://doi.org/10.3390/s24041266>.
23. Shangaranarayane N. P., Aakashbabu V., Balamurugan M., Gokulraj R. Machine Learning Driven Smart Transportation Sharing. *Journal of IoT in Social, Mobile, Analytics, and Cloud*. 2024. Vol. 6, No. 1. Pp. 1–12.

24. Mühlematter D. J., Wiedemann N., Xin Y., Raubal M. Spatially-aware station based car-sharing demand prediction. *Journal of Transport Geography*. 2024. Vol. 114. Pp. 103765. URL: <https://doi.org/10.1016/j.jtrangeo.2023.103765>.
25. Liang L., Ye H., Li G. Y. Toward Intelligent Vehicular Networks: A Machine Learning Framework. *IEEE Internet of Things Journal*. 2019. Vol. 6, No. 1. Pp. 124–135. URL: <https://doi.org/10.1109/JIOT.2018.2872122>.
26. Mandala V., Surabhi S. N. R. D. Integration of AI-Driven Predictive Analytics into Connected Car Platforms. *IARJSET*. 2024. Vol. 7, No. 12. URL: <https://doi.org/10.17148/IARJSET.2020.71216>.
27. Peters H., Parrott M. Model Share AI. *scipy*. 2024. Pp. 216–224. URL: <https://doi.org/10.25080/MDCE8355>.
28. Liu Y., Xie J., Chen N. Stochastic one-way carsharing systems with dynamic relocation incentives through preference learning. *Transportation Research Part E: Logistics and Transportation Review*. 2022. Vol. 166. Pp. 102884. URL: <https://doi.org/10.1016/j.tre.2022.102884>.
29. Luo W., Wang Y., Jiao P., Wang Z., Zhao P. What Kind of Travellers Are Using Carsharing in Beijing? A Study Based on Selective Ensemble Learning. *Sustainability*. 2022. Vol. 14, No. 1. Pp. 540. URL: <https://doi.org/10.3390/su14010540>.
30. Paredes M., Hemberg E., O'Reilly U.-M., Zegras C. Machine learning or discrete choice models for car ownership demand estimation and prediction?: *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, June 2017. Pp.780–785. URL: <https://doi.org/10.1109/MTITS.2017.8005618>.
31. Fu Y., Yu F. R., Li C., Luan T. H., Zhang Y. Vehicular Blockchain-Based Collective Learning for Connected and Autonomous Vehicles. *IEEE Wireless Communications*. 2020. Vol. 27, No. 2. Pp. 197–203. URL: <https://doi.org/10.1109/MNET.001.1900310>.
32. Sun K., Wu J., Bashir A. K., Li J., Xu H., Pan Q., Al-Otaibi Y. D. Personalized Privacy-Preserving Distributed Artificial Intelligence for Digital-Twin-

Driven Vehicle Road Cooperation. *IEEE Internet of Things Journal*. 2024. Vol. 11, No. 22. Pp. 35902–35916. URL: <https://doi.org/10.1109/JIOT.2024.3389656>.

33. Stilgoe J. Machine Learning, Social Learning and the Governance of Self-Driving Cars. 2017. URL: <https://doi.org/10.2139/ssrn.2937316>.

34. Chopra A., Rahman A. U., Malik A. W., Ravana S. D. Adaptive-Learning-Based Vehicle-to-Vehicle Opportunistic Resource-Sharing Framework. *IEEE Internet of Things Journal*. 2022. Vol. 9, No. 14. Pp. 12497–12504. URL: <https://doi.org/10.1109/JIOT.2021.3137264>.

35. Anagnostopoulos T. A Predictive Vehicle Ride Sharing Recommendation System for Smart Cities Commuting. *Smart Cities*. 2021. Vol. 4, No. 1. Pp. 177–191. URL: <https://doi.org/10.3390/smartcities4010010>.

36. Singh A., Al-Abbasi A. O., Aggarwal V. A Distributed Model-Free Algorithm for Multi-Hop Ride-Sharing Using Deep Reinforcement Learning. *IEEE Transactions on Intelligent Transportation Systems*. 2022. Vol. 23, No. 7. Pp. 8595–8605. URL: <https://doi.org/10.1109/TITS.2021.3083740>.

37. Sai Q., Bi ,Jun, Guan ,Wei, Xie ,Dongfan, Zhi R. and. Behavior analysis of carsharing users based on deep belief network and improved two-stage clustering. *Transportation Letters*. 2022. Vol. 14, No. 10. Pp. 1069–1079. URL: <https://doi.org/10.1080/19427867.2021.1987756>.

38. Abouelela M., Lyu C., Antoniou C. Exploring the Potentials of Open-Source Big Data and Machine Learning in Shared Mobility Fleet Utilization Prediction. *Data Science for Transportation*. 2023. Vol. 5, No. 2. Pp. 5. URL: <https://doi.org/10.1007/s42421-023-00068-9>.

39. Wang D., Wang Q., Yin Y., Cheng T. C. E. Optimization of ride-sharing with passenger transfer via deep reinforcement learning. *Transportation Research Part E: Logistics and Transportation Review*. 2023. Vol. 172. Pp. 103080. URL: <https://doi.org/10.1016/j.tre.2023.103080>.

40. Zhang S., Sun H., Liu Y., Lv Y., Wu J., Feng X. Carsharing equitable relocation problem: A two-stage stochastic programming approach with learning-embedded endogenous uncertainty in demand. *Transportation Research Part B:*

*Methodological*. 2024. Vol. 179. Pp. 102868. URL:  
<https://doi.org/10.1016/j.trb.2023.102868>.

41. Turoń K. Car-Sharing Systems in Smart Cities: A Review of the Most Important Issues Related to the Functioning of the Systems in Light of the Scientific Research. *Smart Cities*. 2023. Vol. 6, No. 2. Pp. 796–808. URL:  
<https://doi.org/10.3390/smartcities6020038>.

42. Carsharing ride cost prediction. *Kaggle*.  
URL: <https://www.kaggle.com/c/carsharing-ride-cost-prediction/data>.

# ДОДАТКИ

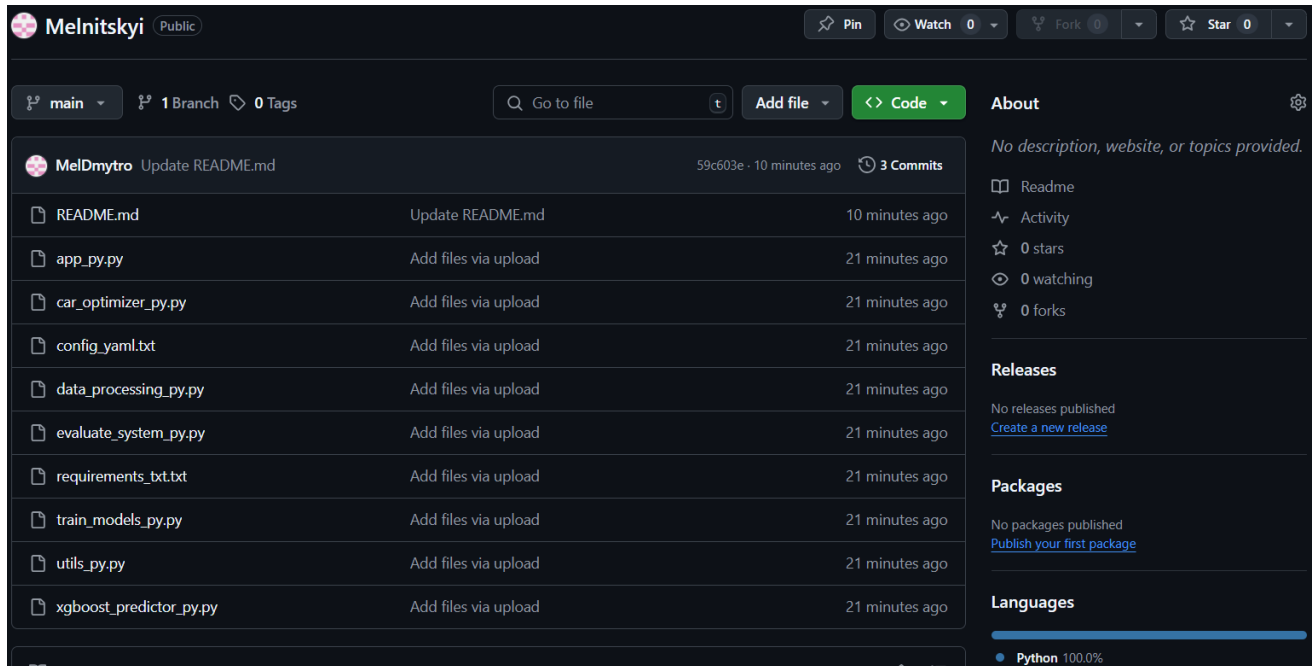
## Додаток А

### Програмний код

#### Посилання на репозиторій на GitHub:

<https://github.com/MelDmytro/Melnitskyi>

#### Вигляд сторінки репозиторію



#### Опис вмісту

requirements.txt – список Python залежностей та бібліотек, необхідних для роботи системи.

config.yaml – файл конфігурації з параметрами системи та налаштуваннями моделей.

data\_processing.py – модуль обробки даних та створення ознак (feature engineering) для машинного навчання.

xgboost\_predictor.py – реалізація XGBoost моделей для прогнозування попиту та класифікації.

car\_optimizer.py – алгоритми оптимізації розподілу автомобілів та формування рекомендацій.

utils.py – допоміжні функції та утиліти для роботи з даними та метриками.

train\_models.py – скрипт навчання ML моделей на підготовлених наборах даних.

`evaluate_system.py` – модуль експериментальної оцінки точності системи та порівняння з базовими методами.

## Додаток Б

### Презентаційний матеріал

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

# МЕТОД АДАПТИВНОГО ПІДБОРУ АВТОМОБІЛІВ КАРШЕРИНГУ З ІНТЕГРАЦІЄЮ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ



**Виконав:**

*студент 4 курсу, групи КН-21-1*  
Дмитро Мельницький



**Керівник:**

*д.т.н., професор кафедри КН*  
Едуард Манзюк

## Актуальність

Традиційні методи розподілу транспортних засобів часто не здатні адаптуватися до мінливих умов міського середовища, що призводить до незбалансованості автопарку, надмірних простоїв автомобілів в одних районах та їх нестачі в інших. Технології штучного інтелекту, зокрема методи машинного навчання та глибокі нейронні мережі, демонструють значний потенціал у вирішенні цих проблем завдяки здатності обробляти великі обсяги даних, виявляти приховані закономірності та формувати точні прогнози попиту на автомобілі в різних локаціях.

## Мета і задачі роботи

**Мета кваліфікаційної роботи бакалавра** – підвищення точності адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту.

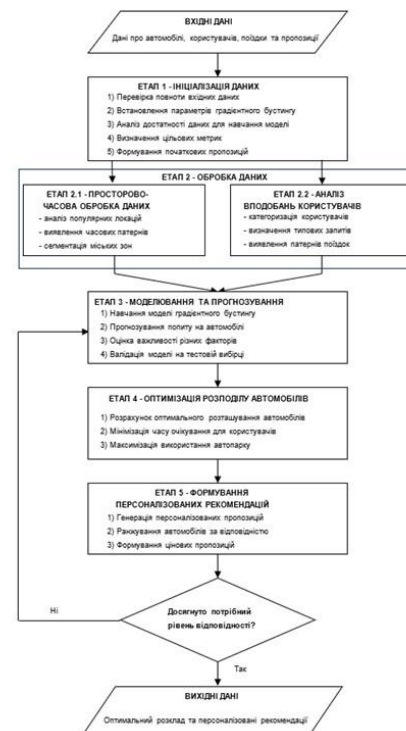
**Об'єкт дослідження** – процес підбору транспортних засобів у системах каршерингу.

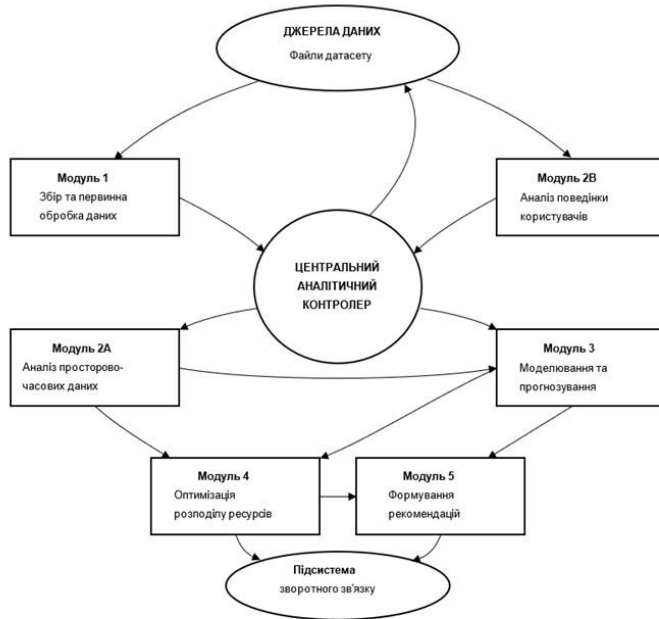
**Предмет дослідження** – методи та алгоритми інтелектуального аналізу вподобань користувачів для оптимізації вибору автомобілів у системах каршерингу.

Завдання роботи :

- здійснити комплексний аналіз предметної області каршерингових систем та особливостей підбору автомобілів у них;
- провести дослідження існуючих теоретичних підходів до адаптивного розподілу транспортних засобів у системах спільного користування;
- розробити метод адаптивного підбору автомобілів каршерингу, який використовує алгоритми машинного навчання для прогнозування попиту та оптимізації розподілу транспортних засобів;
- спроектувати архітектуру інформаційної системи підбору автомобілів на основі розробленого методу;
- провести експериментальне дослідження розробленого методу адаптивного підбору автомобілів.

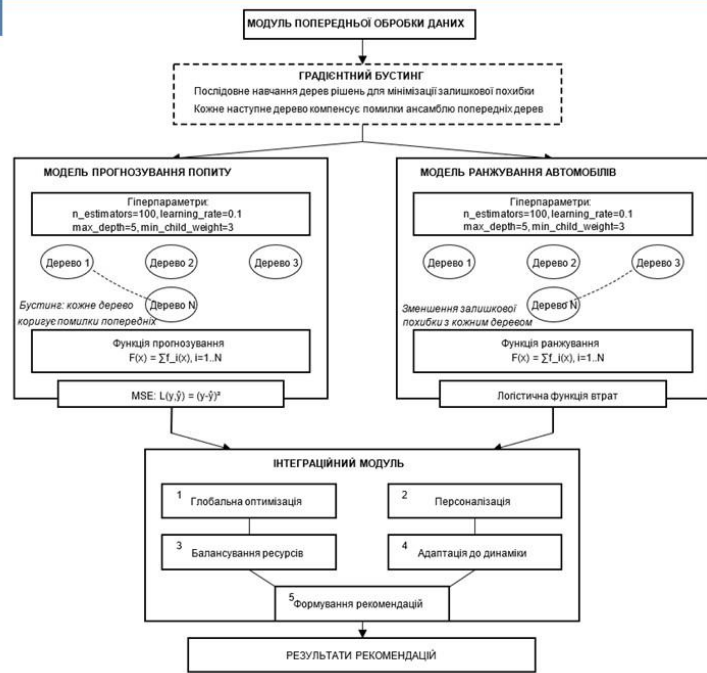
Схема та етапи методу адаптивного підбору автомобілів каршерингу з інтеграцією технологій штучного інтелекту



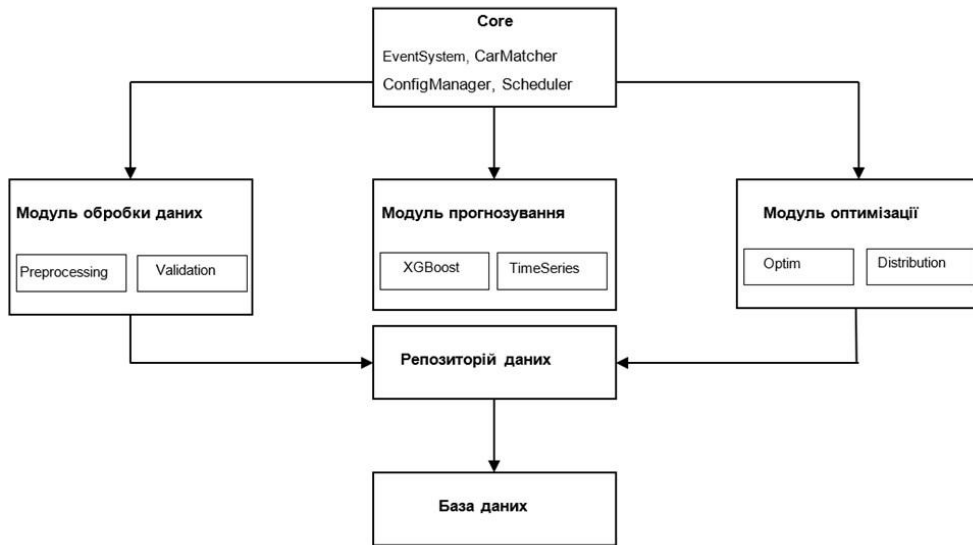


Комплексна функціональна структура інформаційної системи адаптивного підбору автомобілів каршерингу

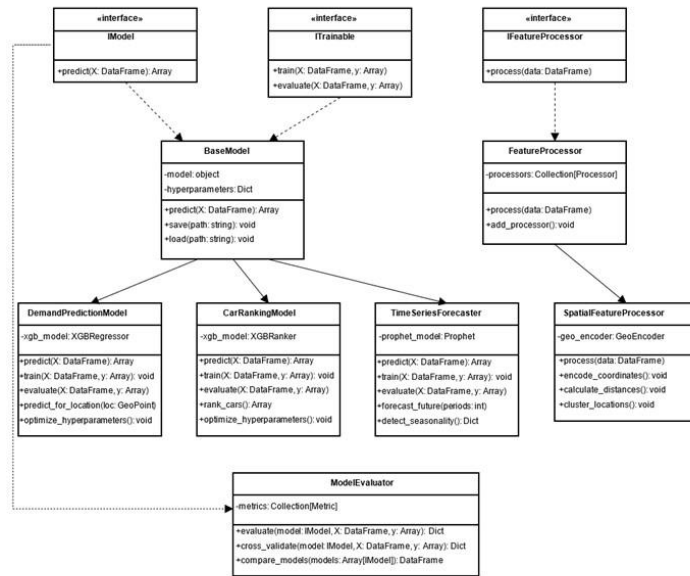
Архітектура моделі XGBoost для адаптивного підбору автомобілів каршерингу



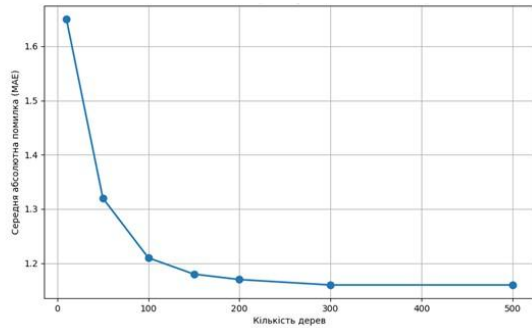
Архітектура системи адаптивного підбору автомобілів каршерінгу



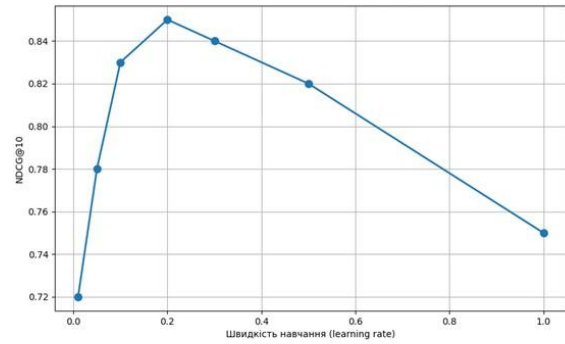
Діаграма класів модуля прогнозування



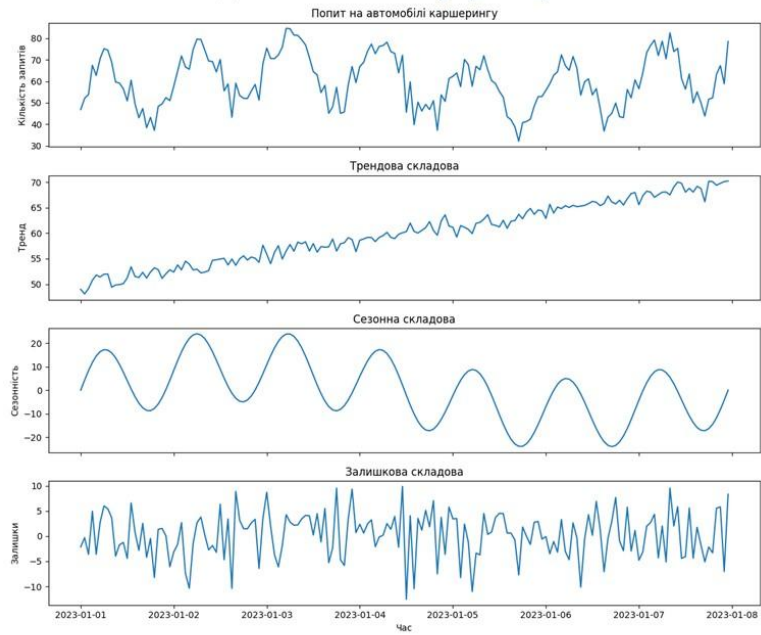
Залежність середньої абсолютної помилки



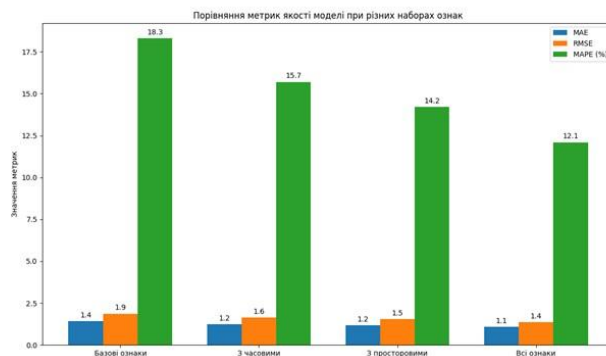
Підбір оптимальних гіперпараметрів



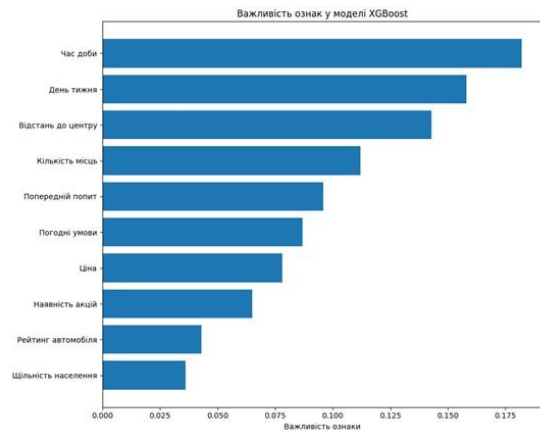
Декомпозиція часового ряду попиту



### Порівняння метрик якості моделі при використанні різних наборів ознак



### Важливість ознак



## Висновки

У ході виконання кваліфікаційної роботи бакалавра було успішно розроблено метод адаптивного підбору автомобілів у системах каршерингу з інтеграцією технологій штучного інтелекту

- здійснено комплексний аналіз предметної області каршерингових систем та особливостей підбору автомобілів у них;
- проведено дослідження існуючих теоретичних підходів до адаптивного розподілу транспортних засобів у системах спільного користування;
- розроблено метод адаптивного підбору автомобілів каршерингу, який використовує алгоритми машинного навчання для прогнозування попиту та оптимізації розподілу транспортних засобів;
- спроектовано архітектуру інформаційної системи підбору автомобілів на основі розробленого методу;
- проведено експериментальне дослідження розробленого методу адаптивного підбору автомобілів.

ДЯКУЮ ЗА УВАГУ!

---

# Anti-Plagiarism (UA) v-15.281 Educational

The maximum coincidence with one document 2.0%

Dictionaries check: en\_US, ru\_RU, ua\_UA. Errors in the documents: 12%

ID: 246829 Title: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту Added in a DB: 2025-06-18 Authors: Дмитро МЕЛЬНИЦЬКИЙ Heads: Едуард МАНЗЮК Consultants: Opponents:	Document		Sum coincidence on the DB	
	Symbols	Lexemes	Symbols	Lexemes
	83221	1246	4127 (5%)	60 (5%)

## Plagiarism sources

ID	Description	Plagiarism presence in the document	
		Symbols	Lexemes

## Протокол аналізу звіту подібності науковим керівником

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

**Автор:** Дмитро МЕЛЬНИЦЬКИЙ

**Співавтор:**

**Назва:** КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту

**Науковий керівник:** Едуард МАНЗІЮК, д.т.н., доцент

**Підрозділ:** Кафедра комп'ютерних наук

**Коефіцієнт подібності 1:** 3.8%

**Коефіцієнт подібності 2:** 1.6%

**Мікропробіли:** 0

**Заміна букв:** 0

**Інтервали:** 0

**Білі знаки:** 3

**Дата створення звіту:** 2025-06-18 20:44:14.0

**Після аналізу Звіту подібності констатую наступне:**

Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.

Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.

Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедур. Таким чином робота не приймається.

**Обґрунтування:**

2025-06-18

Дата

експерт

*Лео Васьків СР.*

РІШЕННЯ ЕКСПЕРТНОЇ КОМІСІЇ КАФЕДРИ КОМП'ЮТЕРНИХ НАУК  
ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Назва кваліфікаційної роботи Метод адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту

Автор студент групи КН-21-1 Дмитро МЕЛЬНИЦЬКИЙ

Освітня програма Комп'ютерні науки

Рівень вищої освіти перший (бакалаврський)

Спеціальність 122 – Комп'ютерні науки

Науковий керівник: д.т.н., проф. каф. комп'ютерних наук Едуард МАНЗЮК

На основі аналізу кваліфікаційної роботи на дотримання вимог академічної доброчесності (у т.ч. відсутності ознак академічного плагіату) з урахуванням результатів перевірки роботи спеціалізованим програмними засобами комісія зробила такий висновок:

№	Висновок	Позначка про відповідність
1	Ознаки академічного плагіату	
1.1	Запозичення, виявлені в роботі, є законними і не є академічним плагіатом (далі – зазначаються підстави віднесення запозичень до правомірних, якщо потрібно). Робота приймається до захисту.	<i>відповідає</i>
1.2	Виявлені запозичення не є академічним плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи (далі – зазначаються детальні та аргументовані підстави віднесення запозичень до правомірних). Робота приймається до захисту, але має бути відкоригована.	
1.3	Виявлені запозичення не є академічним плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. Робота може бути допущена до захисту після того як буде відкоригована та доопрацьована і успішно пройде повторну перевірку на академічний плагіат.	
1.4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття текстових запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	
2	Інші види порушень академічної доброчесності	<i>відсутні</i>

Підтвердження:

Запозичення, виявлені в роботі Дмитра Мельницького, не є плагіатом, оскільки: запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; до запозичень входять фрагменти, які не мають авторства і містять поширені конструкції та загальновідомі терміни, скорочення. Рівень подібності не перевищує допустимої межі. Таким чином, робота є законною та приймається до захисту.

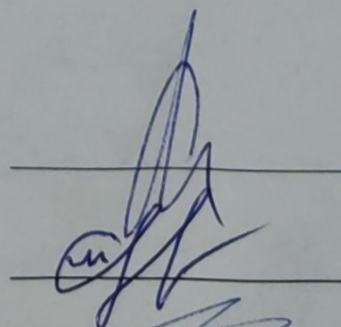
Обсяг запозичень, визначений системами виявлення збігів/ідентичності/схожості:

- за системою Anti-Plagiarism: 2%;

- за системою StrikePlagiarism КП1: 3,8%, КП2: 1,6%.

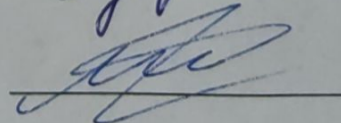
18.06.2025

Завідувач кафедри



Олександр БАРМАК

Гарант освітньої програми



Олександр МАЗУРЕЦЬ

Керівник кваліфікаційної роботи



Едуард МАНЗЮК



## РЕЦЕНЗІЯ

### на кваліфікаційну роботу бакалавра

студента гр. КН-21-1 Дмитра МЕЛЬНИЦЬКОГО

за темою: Метод адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту

#### 1. Актуальність обраної теми

Обрана тема є актуальною в контексті розвитку сучасних інтелектуальних транспортних систем та концепції розумних міст. Зростання популярності каршерингових сервісів у світі та в Україні створює потребу в ефективних методах оптимізації підбору автомобілів. Інтеграція технологій штучного інтелекту дозволяє вирішувати складні задачі прогнозування попиту та персоналізації рекомендацій, що робить дослідження перспективним та практично значущим.

#### 2. Повнота розкриття мети та завдань роботи

Мета роботи сформульована чітко та конкретно - підвищення ефективності вибору транспортних засобів у системах каршерингу. Поставлені завдання логічно структуровані та повністю відповідають меті дослідження. Всі заявлені завдання успішно виконані: проведено аналіз предметної області, розроблено метод адаптивного підбору, спроектовано архітектуру системи, створено програмну реалізацію та проведено експериментальну перевірку.

#### 3. Зміст кожного розділу роботи

Перший розділ містить ґрунтовний аналіз предметної області каршерингових систем, огляд сучасних теоретичних підходів та існуючих програмних рішень. Матеріал викладено систематично з використанням актуальних джерел. Другий розділ присвячений розробці методу адаптивного підбору автомобілів з детальним описом архітектури градієнтного бустингу та інформаційної структури системи. Третій розділ демонструє практичну реалізацію методу з експериментальною перевіркою ефективності. Зміст розділів логічно взаємопов'язаний та повністю розкриває тему дослідження.

#### 4. Оцінка розробленого методу та його практична цінність

Розроблений метод базується на сучасних алгоритмах машинного навчання, зокрема XGBoost, що забезпечує високу точність прогнозування та ранжування. Метод враховує просторово-часові аспекти, персоналізацію та адаптивність до змін у поведінці користувачів. Практична цінність підтверджується експериментальними результатами, які демонструють переваги над традиційними підходами. Розроблена система може бути впроваджена в реальних каршерингових сервісах для підвищення ефективності використання автопарку.

#### 5. Якість оформлення кваліфікаційної роботи бакалавра

Робота оформлена відповідно до встановлених вимог. Структура роботи логічна, текст викладено грамотно з дотриманням наукового стилю. Ілюстративний матеріал

діаграми, схеми якісно виконаний та доповнює текстовий матеріал. Список використаних джерел містить сучасні джерела.

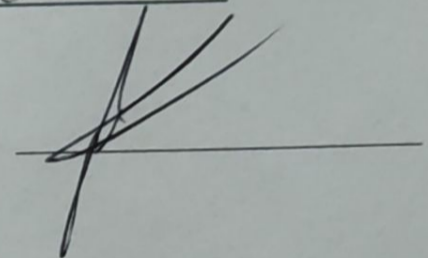
6. Недоліки кваліфікаційної роботи бакалавра

Серед незначних недоліків можна відзначити відсутність порівняння з комерційними рішеннями на ринку каршерингу. Також бажано було б більш детально розглянути питання конфіденційності персональних даних користувачів при використанні системи.

7. Загальний висновок (допускається чи не допускається до захисту), та оцінка на яку заслуговує кваліфікаційна робота.

Враховуючи рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка «Сідмоста».

Рецензент к.т.н., доц. кафедри КІС Нічепорук А.О.





**ВІДГУК НАУКОВОГО КЕРІВНИКА**  
**на кваліфікаційну роботу бакалавра**

студента гр. КН-21-1 Дмитра МЕЛЬНИЦЬКОГО

за темою Метод адаптивного підбору автомобілів для каршерингу з інтеграцією технологій штучного інтелекту

**1. Актуальність теми**

Сучасні тенденції урбанізації та розвиток концепції розумних міст обумовлюють необхідність впровадження ефективних систем спільного користування автомобілями. Зростаючий попит на каршерингові послуги потребує інноваційних рішень для оптимізації підбору транспортних засобів. Розроблення методу адаптивного підбору автомобілів з використанням алгоритмів штучного інтелекту є перспективним напрямком досліджень у галузі інтелектуальних транспортних систем. Дана тематика відповідає сучасним викликам у сфері міського транспорту та має значний потенціал для практичного застосування.

**2. Відповідність роботи предметній області Стандарту спеціальності 122 Комп'ютерні науки**

Кваліфікаційна робота повністю відповідає предметній області стандарту спеціальності 122 "Комп'ютерні науки". У роботі досліджуються інформаційні моделі процесів каршерингу, розроблені алгоритми машинного навчання для прогнозування попиту та ранжування автомобілів. Застосовано сучасні методи обробки та аналізу даних, технології градієнтного бустингу XGBoost. Реалізовано інформаційну систему з модульною архітектурою, що демонструє практичне застосування теоретичних знань у галузі комп'ютерних наук.

**3. Професійні та особистісні якості бакалавра**

У процесі виконання дослідження студент Мельницький Дмитро продемонстрував високий рівень технічної підготовки та аналітичного мислення. Виявив здатність до самостійного опрацювання складної науково-технічної літератури та адаптації сучасних алгоритмів машинного навчання до специфіки предметної області.

**4. Ступінь самостійності під час виконання кваліфікаційної роботи**

Студент виконував роботу самостійно на всіх етапах дослідження. Особисто провів аналіз предметної області, розробив метод адаптивного підбору автомобілів,

спроектував архітектуру системи та здійснив її програмну реалізацію. Самостійно опрацював експериментальні дані та провів оцінку ефективності запропонованого методу.

#### **5. Ступінь оволодіння методами дослідження**

Студент продемонстрував глибоке розуміння методів машинного навчання, зокрема алгоритмів градієнтного бустингу. Володіє навичками обробки та аналізу великих масивів даних, методами оцінки якості прогностичних моделей. Ефективно застосовував інструменти візуалізації результатів та статистичного аналізу. Показав вміння поєднувати теоретичні знання з практичними навичками програмної реалізації.

#### **6. Повнота та якість розкриття теми роботи**

Тема роботи розкрита повністю та всебічно. Проведено ґрунтовний аналіз сучасного стану проблеми, досліджено теоретичні підходи до оптимізації каршерингових систем. Чітко сформульовано мету та завдання дослідження, які успішно виконані. Розроблено оригінальний метод з використанням технологій штучного інтелекту, проведено експериментальну верифікацію результатів.

#### **7. Логічність, послідовність, аргументованість, літературна грамотність викладення матеріалу**

Матеріал викладено логічно та послідовно з дотриманням наукового стилю. Структура роботи обґрунтована та відповідає поставленим завданням. Висновки аргументовані та підкріплені експериментальними даними. Текст написано грамотно, термінологія використовується коректно.

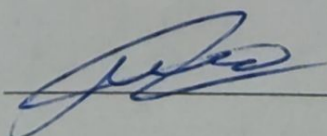
#### **8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин**

Розроблений метод та програмна система мають значний потенціал для практичного впровадження в реальних каршерингових сервісах. Результати можуть використовуватися для підвищення ефективності використання автопарку, покращення якості обслуговування клієнтів та зменшення операційних витрат. Окремі компоненти системи можуть знайти застосування в суміжних галузях спільного використання ресурсів.

#### **9. Висновок про можливість допуску кваліфікаційної роботи бакалавра до захисту, на яку оцінку заслуговує робота**

Кваліфікаційна робота виконана на високому рівні, повністю відповідає вимогам до робіт бакалаврського рівня. Робота заслуговує на допуск до захисту. Рекомендована оцінка «відмінно».

Керівник



д.т.н., проф. КН Едуард МАНЗЮК