МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ПРИРОДОКОРИСТУВАННЯ
ФАКУЛЬТЕТ МЕХАНІКИ, ЕНЕРГЕТИКИ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА

другого (магістерського) рівня вищої освіти

на тему: **«Прогнозування вартості вживаних автомобілів із вибором програмної платформи автоматичного створення моделі машинного навчання»**

Виконав: студент групи Іт-61

Спеціальності 126 «Інформаційні системи та технології»
(шифр і назва)
Досяк Михайло Орестович
(Прізвище та ініціали)

Керівник: к.ф-м.н., в.о. доцента Чухрай Л.В.
(Прізвище та ініціали)

Рецензент: к.т.н., доцент Тимочко В.О.
(Прізвище та ініціали)

ДУБЛЯНИ-2024

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ПРИРОДОКОРИСТУВАННЯ
ФАКУЛЬТЕТ МЕХАНІКИ, ЕНЕРГЕТИКИ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Другий (магістерський) рівень вищої освіти
Спеціальність 126 «Інформаційні системи та технології»

«ЗАТВЕРДЖУЮ»

Завідувач кафедри_____

д.т.н., проф. А.М. Тригуба
«____» _____2023 р.

ЗАВДАННЯ

на кваліфікаційну роботу студенту

Досяку Михайлу Орестовичу

1. Тема роботи: «Прогнозування вартості вживаних автомобілів із вибором програмної платформи автоматичного створення моделі машинного навчання»
Керівник роботи Чухрай Любомир Володимирович, в.о. доцента
затверджені наказом по університету від 28.04.2023 року № 133/к-с.

2. Строк подання студентом роботи 10.01.2024 р.

3. Вихідні дані до роботи: *дані оптимізації портфелів проектів розвитку громад; програмні платформи автоматичного створення моделі машинного навчання.*

4. Зміст розрахунково-пояснювальної записки (перелік питань, які необхідно розробити)

Вступ.

1. Аналіз наявних програмних платформ для автоматичного створення моделей машинного навчання та завдання кваліфікаційної роботи.

2. Вибір програмної платформи та методів автоматичного створення моделі машинного навчання для прогнозування вартості вживаних автомобілів.

3. Результати підготовки даних та автоматичного створення моделі машинного навчання для прогнозування вартості вживаних автомобілів.

4. Охорона праці та безпека у надзвичайних ситуаціях.

5. Визначення економічної ефективності.

Висновки та пропозиції.

Список використаної літератури.

5. Перелік ілюстраційного матеріалу (з точним зазначенням обов'язкових слайдів): *аналіз наявних програмних платформ для автоматичного створення моделей машинного навчання та завдання кваліфікаційної роботи; вибір програмної платформи та методів автоматичного створення моделі машинного навчання для прогнозування вартості вживаних автомобілів; результати підготовки даних та автоматичного створення моделі машинного навчання для прогнозування вартості вживаних автомобілів; результати порівняння запропонованої моделі із існуючими аналогами; економічна ефективність.*

6. Консультанти з розділів:

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1, 2, 3, 5	*Чухрай Л.В. в.о. доцента кафедри інформаційних технологій*		
4	*Городецький І.М., доцент кафедри фізики, інженерної механіки та безпеки виробництва*		

7. Дата видачі завдання 28 квітня 2023 р.

Календарний план

№ з/п	Назва етапів кваліфікаційної роботи	Терміни виконання етапів роботи	Примітка
1	*Написання першого розділу*	*28.04-20.05.23*	
2	*Виконання другого розділу та аркушів ілюстраційного матеріалу до нього*	*21.05-14.08.23*	
3.	*Виконання третього розділу та аркушів ілюстраційного матеріалу до нього*	*15.08-10.11.23*	
4.	*Написання розділу «Охорона праці та безпека у надзвичайних ситуаціях»*	*11.11-30.11.23*	
5.	*Оцінення ефективності запропонованої системи*	*01.12-20.12.23*	
6.	*Завершення оформлення розрахунково-пояснювальної записки та аркушів ілюстраційного матеріалу*	*21-31.12.23*	
7.	*Завершення роботи вцілому*	*01-10.01.24*	

Студент _____ Досяк М.О.
(підпис)

Керівник роботи _____ Чухрай Л.В.
(підпис)

УДК 004.85:004.93.5 (075.8)

Прогнозування вартості вживаних автомобілів із вибором програмної платформи автоматичного створення моделі машинного навчання.

Досяк М.О. Кафедра інформаційних технологій – Дубляни, ЛНУП, 2024.

Кваліфікаційна робота: 73 с. текст. част., 24 рис., 5 табл., 15 арк. ілюстраційного матеріалу, 51 джерело.

Виконано аналіз наявних програмних платформ для автоматичного створення моделей машинного навчання та завдання кваліфікаційної роботи. Наведено особливості прогнозування вартості вживаних автомобілів. Проведено аналіз стану використання автоматичного створення моделей машинного навчання. Сформульовано завдання кваліфікаційної роботи.

Здійснено вибір програмної платформи та методів автоматичного створення моделі машинного навчання для прогнозування вартості вживаних автомобілів. Обгрунтовано доцільність та переваги AutoML. Проаналізовано бібліотеку машинного навчання PYCARET, програмний пакет для автоматизованого машинного навчання Sklearn, бібліотеку машинного навчання MLBOX, інструмент оптимізації конвеєрів TPOT, платформу H2O, бібліотеку Auto-Keras, платформу Datarobot. Представлено методи створення моделі машинного навчання для прогнозування вартості вживаних автомобілів.

Подано результати підготовки даних та автоматичного створення моделі машинного навчання для прогнозування вартості вживаних автомобілів. Створено модель машинного навчання для прогнозування вартості вживаних автомобілів. Подано результати обґрунтування моделі машинного навчання для прогнозування вартості вживаних автомобілів. Подано результати порівняння запропонованої моделі із існуючими аналогами.

Розроблено заходи із охорони праці. Визначено ефективність від використання моделі.

ЗМІСТ

ВСТУП 7
РОЗДІЛ 1. АНАЛІЗ НАЯВНИХ ПРОГРАМНИХ ПЛАТФОРМ ДЛЯ АВТОМАТИЧНОГО СТВОРЕННЯ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ТА ЗАВДАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ 8
1.1. Особливості прогнозування вартості вживаних автомобілів 8
1.2. Використання технологій машинного навчання для прогнозування вартості вживаних автомобілів 9
1.3. Формування наборів даних для прогнозування вартості вживаних автомобілів 14
1.4. Аналіз стану використання автоматичного створення моделей машинного навчання 18
1.5. Завдання кваліфікаційної роботи 20
РОЗДІЛ 2. ВИБІР ПРОГРАМНОЇ ПЛАТФОРМИ ТА МЕТОДІВ АВТОМАТИЧНОГО СТВОРЕННЯ МОДЕЛІ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ВАРТОСТІ ВЖИВАНИХ АВТОМОБІЛІВ 22
2.1. Доцільність та переваги automl 22
2.2. Вибір програмної платформи для прогнозування вартості вживаних автомобілів 23
2.2.1. Бібліотека машинного навчання PYCARET 23
2.2.2. Програмний пакет для автоматизованого машинного навчання sklearn 24
2.2.3. Бібліотека машинного навчання MLBOX 24
2.2.4. ІНСТРУМЕНТ оптимізації конвеєрів TPOT 25
2.2.5. Платформа H2o 26
2.2.6. Бібліотека Auto-Keras 27
2.2.7. Платформа Datarobot 28
2.3. Методи створення моделі машинного навчання для прогнозування вартості вживаних автомобілів 29

РОЗДІЛ 3. РЕЗУЛЬТАТИ ПІДГОТОВКИ ДАНИХ ТА АВТОМАТИЧНОГО СТВОРЕННЯ МОДЕЛІ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ВАРТОСТІ ВЖИВАНИХ АВТОМОБІЛІВ 32
3.1. Підготовка даних для прогнозування вартості вживаних автомобілів 32
3.2. Аналіз даних для прогнозування вартості вживаних автомобілів 37
3.3. Створення моделі машинного навчання для прогнозування вартості вживаних автомобілів 41
3.4. Результати обґрунтування моделі машинного навчання для прогнозування вартості вживаних автомобілів 43
3.5. Результати порівняння запропонованої моделі із існуючими аналогами 49
РОЗДІЛ 4. ОХОРОНА ПРАЦІ ТА БЕЗПЕКА У НАДЗВИЧАЙНИХ СИТУАЦІЯХ 53
4.1. Аналіз шкідливих чинників під час роботи за комп'ютером 53
4.2. Розробка заходів щодо покращення умов праці під час роботи за комп'ютером 54
4.3. Розробка логічно-імітаційної моделі процесу виникнення травм під час роботи за комп'ютером 55
4.4. Розробка заходів щодо безпеки у надзвичайних ситуаціях 58
РОЗДІЛ 5. ВИЗНАЧЕННЯ ЕКОНОМІЧНОЇ ЕФЕКТИВНОСТІ ВІД ВИКОРИСТАННЯ РОЗРОБЛЕНОЇ МОДЕЛІ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ВАРТОСТІ ВЖИВАНИХ АВТОМОБІЛІВ 59
ВИСНОВКИ І ПРОПОЗИЦІЇ 62
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 68

ВСТУП

У сучасному світі, насиченому великим обсягом інформації та швидкими технологічними змінами, важливо мати ефективні інструменти для прогнозування різних явищ. Один із таких важливих аспектів – визначення вартості вживаних автомобілів. Ця проблема є актуальною для споживачів, продавців та експертів у сфері автомобільного ринку.

Машинне навчання визнано одним із потужних інструментів для розв’язання завдань прогнозування, а відповідний вибір програмної платформи для створення моделей машинного навчання стає ключовим етапом у розробці ефективних прогностичних систем. У цьому контексті необхідні дослідження та вдосконалення методів прогнозування вартості вживаних автомобілів з використанням сучасних програмних платформ.

Наша кваліфікаційна робота має на меті вивчення та аналіз різних програмних платформ для автоматичного створення моделей машинного навчання та їх застосування для ефективного прогнозування вартості вживаних автомобілів. Результати цього дослідження можуть сприяти вдосконаленню методів оцінки автомобільного ринку та забезпеченню інформаційних рішень у сфері купівлі та продажу автотранспорту.

Об’єктом дослідження є алгоритми та моделі машинного навчання для прогнозування вартості вживаних автомобілів.

Предмет дослідження є вплив програмних платформ для автоматичного створення моделей машинного навчання на якість прогнозування вартості вживаних автомобілів.

РОЗДІЛ 1.
АНАЛІЗ НАЯВНИХ ПРОГРАМНИХ ПЛАТФОРМ ДЛЯ АВТОМАТИЧНОГО СТВОРЕННЯ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ТА ЗАВДАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ

1.1. Особливості прогнозування вартості вживаних автомобілів

Прогнозування вартості перепродажу автомобіля – це процес для оцінки ціни старого автомобіля. Уживані автомобілі користуються високим попитом, але процес оцінки їхньої вартості має недоліки, що може призвести до нереалістичних цін на них.

У окремих наукових працях пропонується комплексний підхід машинного навчання XG Boost (рис. 1.1) [26], який дає можливість врахувати важливі характеристики автомобіля та прогнозує вартість на їх основі.

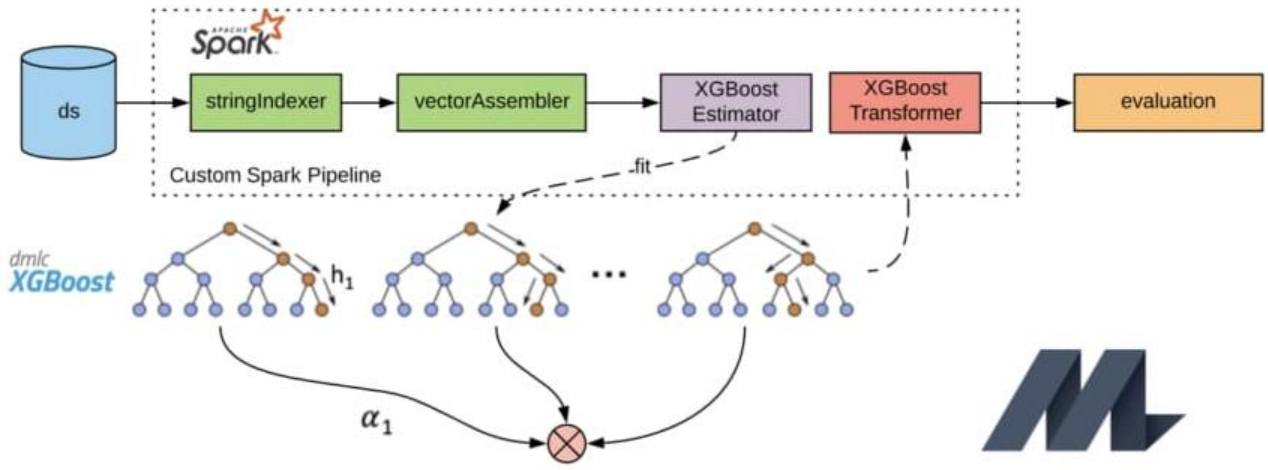

Рисунок 1.1 – Комплексний підхід машинного навчання із використанням XG Boost [26]

Існуючі системи використовують кілька посилених градієнтом дерев, що мінімізує переобладнання та усуває нерівності в прогнозах. Крім того, обчислюється та порівнюється точність для різних алгоритмів машинного

навчання, таких як лінійна регресія, регресія ласо, випадковий ліс, KNN і CART.

Згідно з результатами експериментів, основними факторами, які вплинули на ціни автомобіля, є рік випуску, відстань, тип палива та трансмісія. Запропоновані підходи дають багатообіцяючі результати у прогнозуванні вартості перепродажу автомобіля та досягає точності 93,73%.

1.2. Використання технологій машинного навчання для прогнозування вартості вживаних автомобілів

Машинне навчання – це галузь штучного інтелекту, яка допомагає створювати аналітичні моделі. Вона просуває ідею мінімального втручання людини, щоб дозволити системам навчатися самостійно та відповідно коригувати дії. Тут комп'ютери можуть навчатися, визначаючи та спостерігаючи закономірності в наданих даних, і приймають рішення на основі прикладів, наданих на етапі навчання.

Прогнозування вартості перепродажу автомобіля – це метод оцінки ціни старого автомобіля для цілей продажу та купівлі за розумними цінами [27; 28].

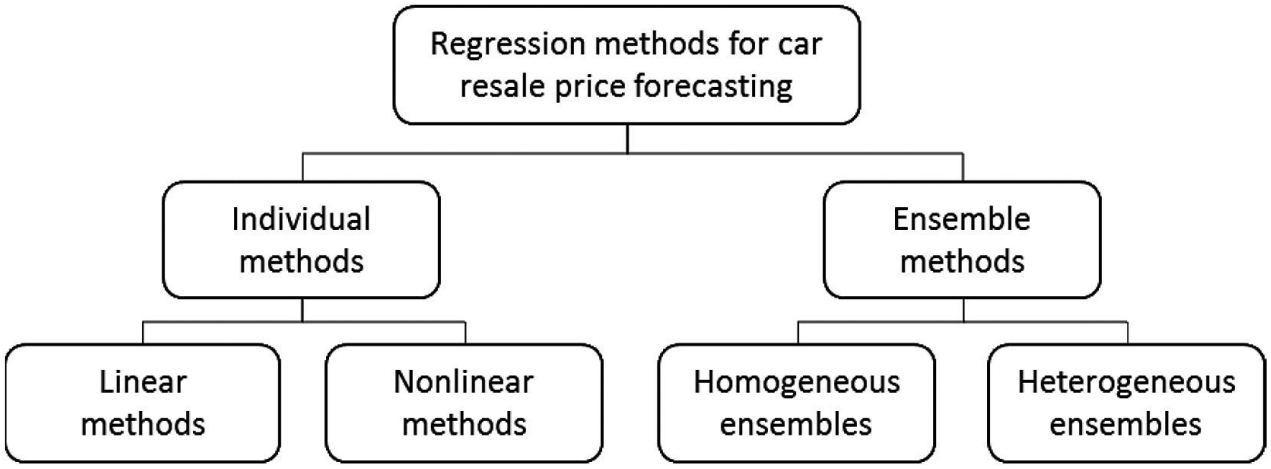

Рисунок 1.2 – Таксономія методів регресії, використаних у прогнозуванні вартості вживаних автомобілів [27]

Оцінити ціну вживаного автомобіля – непросте завдання, оскільки воно включає багато факторів, які слід брати до уваги, і багато з них значною мірою впливають на ціну автомобіля [29; 33].

Виходячи з представлених на даний момент досліджень, найкраще рішення для прогнозування отримано на основі фундаментального аналізу та технічного аналізу з аналізом настроїв і моделями глибокого навчання. Методи ансамблю забезпечують особливо багатообіцяючі результати прогнозування. Алгоритми, які використовуються для прогнозування з урахуванням наукових статей, наведені на рис. 1.3.

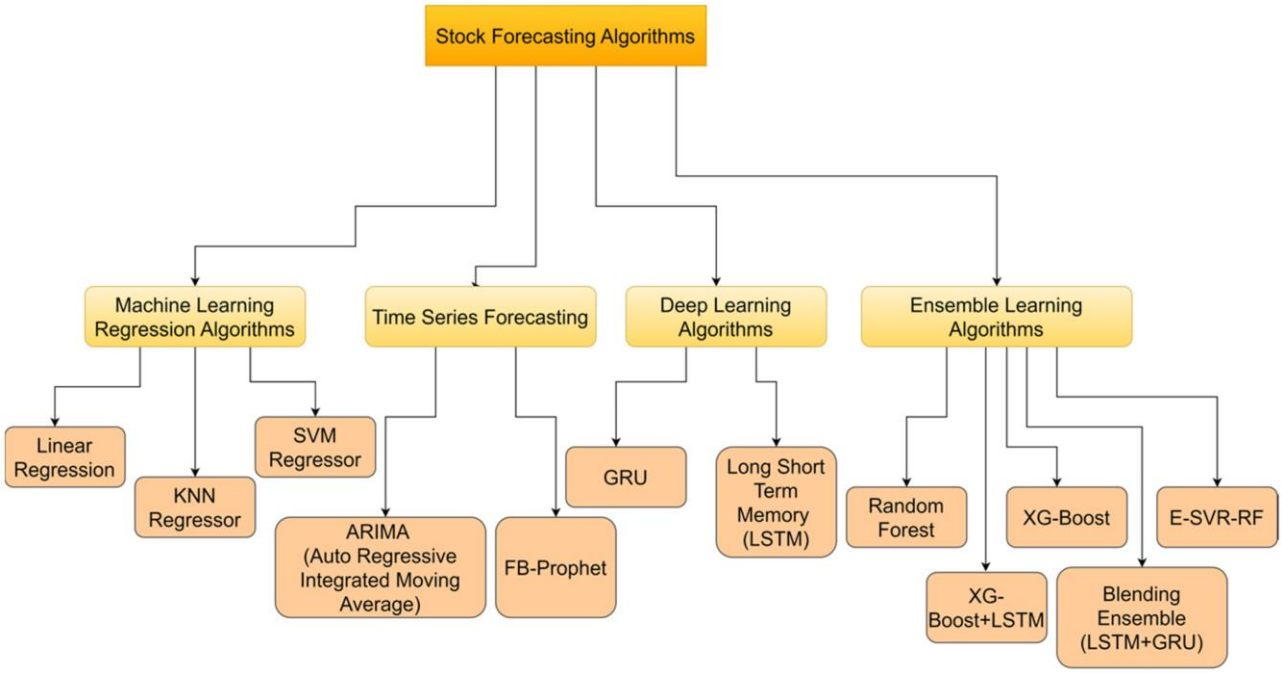

Рисунок 1.3 – Алгоритми машинного навчання для вирішення задач прогнозування

Таким чином, різноманітні алгоритми прогнозування на основі технологій машинного навчання більш ефективно підходять для цієї теми. У роботі [38] вказується, що прогнозування ціни на транспортний засіб було більш популярним у дослідницькій галузі, і воно потребує переважної роботи та інформації про експертів у цій конкретній галузі. Авторами цієї роботи визначається кількість різних атрибутів, а також виконується точний прогноз. Щоб визначити ціну вживаних транспортних засобів, було розроблено чітко

визначену модель за допомогою трьох методів машинного навчання, таких як штучна нейронна мережа, опорна векторна машина та випадковий ліс. Ці методи використовувалися не для окремих елементів, а для всієї групи елементів даних. Цю групу даних було взято з веб-порталу та використано для прогнозу. Дані зібрані за допомогою веб-парсингу, написаного мовою програмування PHP. Щоб отримати найкращий результат із наданого набору даних, порівнювали різні алгоритми машинного навчання різної продуктивності. Остаточна модель прогнозування була написана на програмі Java (рис. 1.4).

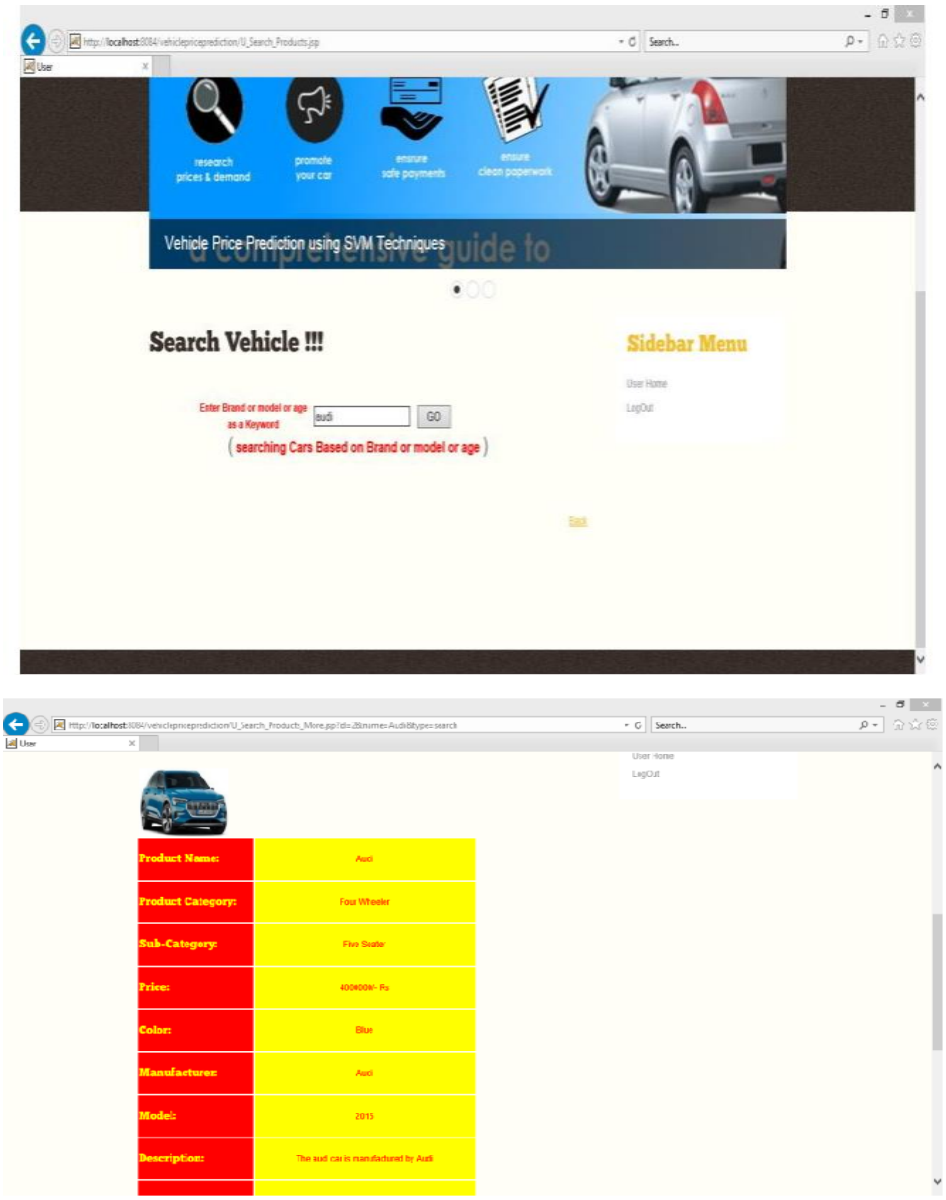

Рисунок 1.4 – Інформаційна система прогнозу ціни вживаних транспортних засобів на основі алгоритмів машинного навчання [38]

Декілька найважливіших факторів, які враховують інші дослідники: бренд, модель, ціна, вік автомобіля, пробіг, тип продавця, тип палива, тип трансмісії, пробіг, двигун, максимальна потужність, кількість місць тощо [39].

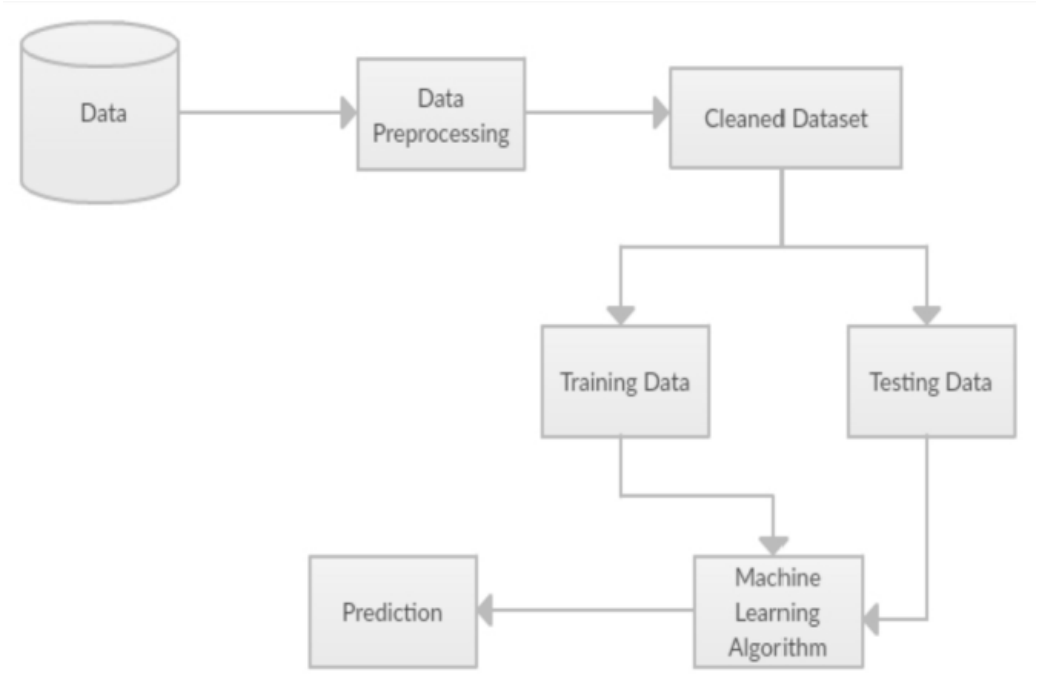

Рисунок 1.5 – Прогнозування вартості перепродажу автомобіля за допомогою алгоритму лінійної регресії [39]

У цій науковій праці [39] досліджується потенціал методів прогнозування для покращення прийняття рішень на ринку вживаних автомобілів. Зокрема, автори зосередилися на розробці моделі ансамблю для визначення точних цін на вживаний автомобіль (рис. 1.6).

Щоб визначити найточнішу модель для цього дослідження, порівнюється продуктивність різних алгоритмів машинного навчання, таких як лінійна регресія, хребтова регресія, ласо-регресія, дерево рішень, випадковий ліс і запропонована нами модель.

```
[28]: plt.plot(x_train,y_train,"r.",
              x_train,hypothesis(a,x_train,b),"b",
              x_train,hypothesis(final_a,x_train,final_b),"g")

[28]: [<matplotlib.lines.Line2D at 0x16c4083e160>,
       <matplotlib.lines.Line2D at 0x16c4083e278>,
       <matplotlib.lines.Line2D at 0x16c4083eac8>]
```

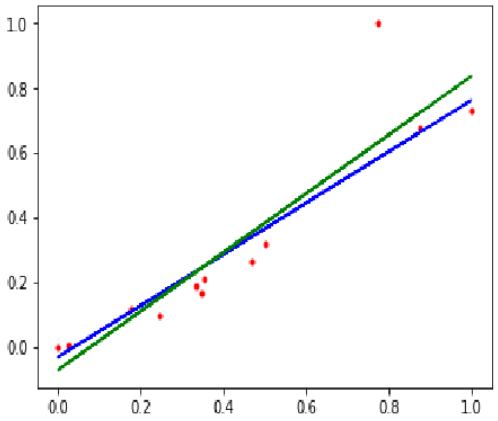

Рисунок 1.6 – Крива визначення найкращого підходу для двигуна Hp проти MSRP [39]

Запропонована у роботі [42] модель виконує систематичне порівняння для оцінки ціни вживаного автомобіля, використовуючи дерева з градієнтним посиленням, які вчаться з попередніх результатів для підвищення точності прогнозу. Він виводить найбільш корельовані характеристики та використовує їх для отримання кінцевого результату.

(a) (b)

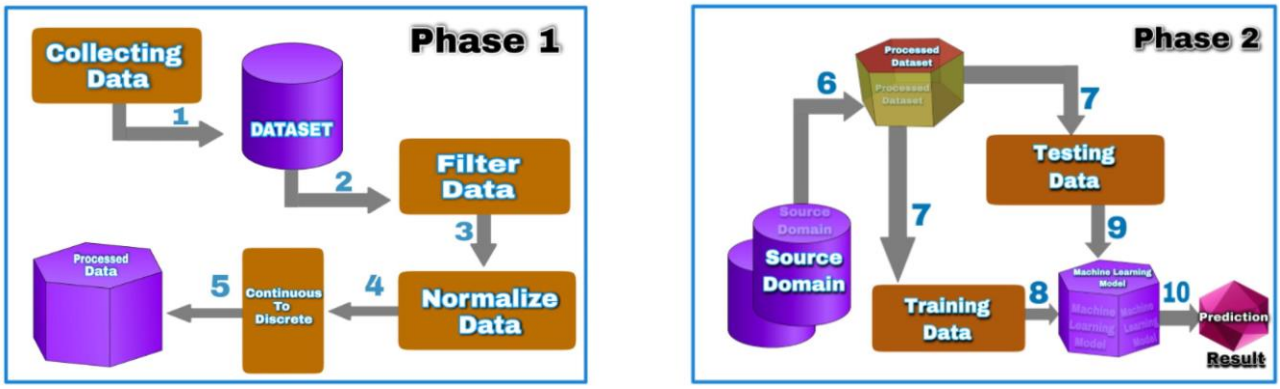

Рисунок 1.7 – Модель оцінки ціни вживаного автомобіля:
(a) Рамкова фаза-1; (b) Рамкова фаза-2. [42]

Фаза 1 запропонованого підходу пов’язана зі збором даних з онлайн-джерела та їх обробкою для перетворення в упорядковану форму.

Необроблений набір даних було зібрано для подальшої обробки шляхом фільтрації даних і видалення надмірності [34] .

Збір набору даних виконується із різних джерел дослідження даних. Для цього аналізу використовується онлайн-набір даних про вживані автомобілі. Він містить численні записи з великою кількістю різних автомобілів.

Попередня обробка даних – тут дані фільтруються, а атрибути, які містять нульові значення, видаляються. Після цього проходить процес нормалізації, щоб не було надлишкового вмісту та не було відсутніх даних. Після цього перетворюємо категоричні значення в числові значення та відокремлюємо цільову змінну.

Етап 2 є найважливішою частиною нашого підходу. Тут модель реалізована за допомогою стандартного набору даних. Автори надали структуровані дані у формі рядків і стовпців до моделі прогнозування [42] . Після того, як дані повністю готові, екземпляри даних розділяються на дві частини – дані навчання та дані тестування для перевірки точності.

Наведені вище два кроки повторюються для інших попередньо існуючих алгоритмів машинного навчання, щоб перевірити точність цих моделей, а потім порівнюються з нашою моделлю, щоб визначити покращення в порівнянні з попередніми моделями.

1.3. Формування наборів даних для прогнозування вартості вживаних автомобілів

У роботі набори даних, використані в цьому дослідженні [43], були отримані з двох різних платформ, а саме Autovit.ro та Mobile.de, які є відомими веб-сайтами, присвяченими продажу вживаних автомобілів у Румунії та Німеччині відповідно. Autovit.ro в першу чергу обслуговує ринок Румунії, з локалізованим фокусом, обмеженим географічними кордонами країни.

На відміну від цього, Mobile.de є ширшою платформою, яка обслуговує німецький ринок вживаних автомобілів, відомий своєю найбільшою часткою ринку в Європейському Союзі. Однак варто зазначити, що Mobile.de також широко використовується окремими особами в сусідніх країнах, включаючи Румунію.

Загалом дослідниками у березні 2023 року з Autovit.ro було зібрано 30264 оголошення про автомобілі, а з Mobile.de – 1308575 записів. Відповідно до умов кожного веб-сайту розміщення дублікатів оголошень про один і той самий транспортний засіб заборонено. Для забезпечення цілісності даних і усунення фальшивої реклами, яка могла б негативно вплинути на процеси навчання та прогнозування, було використано техніку фільтрації викидів. Цю процедуру фільтрації проводили разом із додатковими етапами попередньої обробки для підтримки якості даних.

Набір даних був випадковим чином розділений на 80% набору для навчання та 20% набору для перевірки; однак ми забезпечили збалансований розподіл марок автомобілів у кожній підгрупі. Примітно, що реклама автомобіля може включати кілька зображень, що призводить до кількох записів у наборі даних (по одному для кожного зображення). Однак було вжито заходів для того, щоб набори для навчання та перевірки містили не однакову рекламу, а різні реклами, кожна з відповідними зображеннями.

Марка та модель автомобіля були найбільш відповідними категоріальними ознаками для нашого завдання. На рис. 1.8-1.9 зображено частотний розподіл 10 найпопулярніших марок автомобілів у наборах даних. Встановлено, що результати показують схожість, оскільки вони демонструють значне збігання моделей, які найчастіше зустрічаються на німецькому та румунському ринках. Таке вирівнювання можна пояснити значним припливом імпорту автомобілів до Румунії, зокрема у формі вживаних транспортних засобів, що походять з Німеччини.

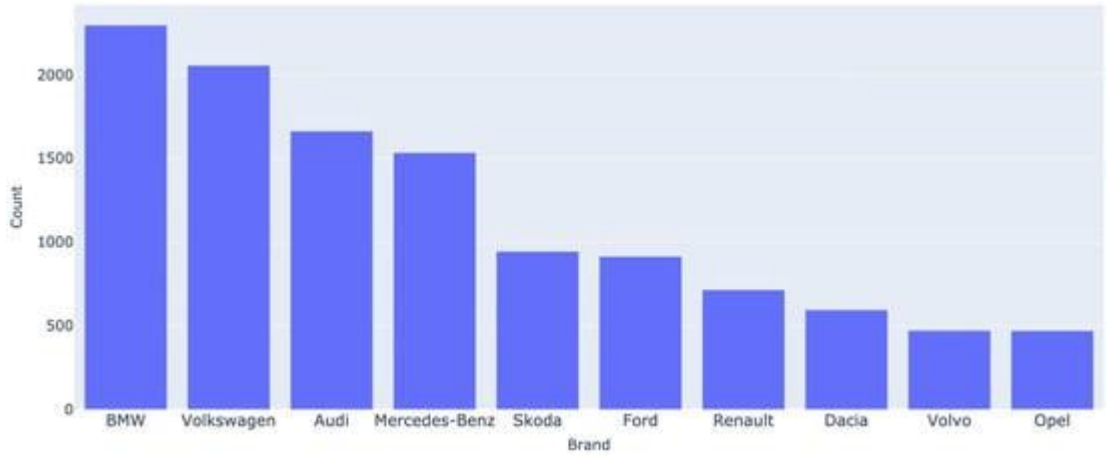

Рисунок 1.8 – Десять найпопулярніших брендів – Autovit.ro [43]

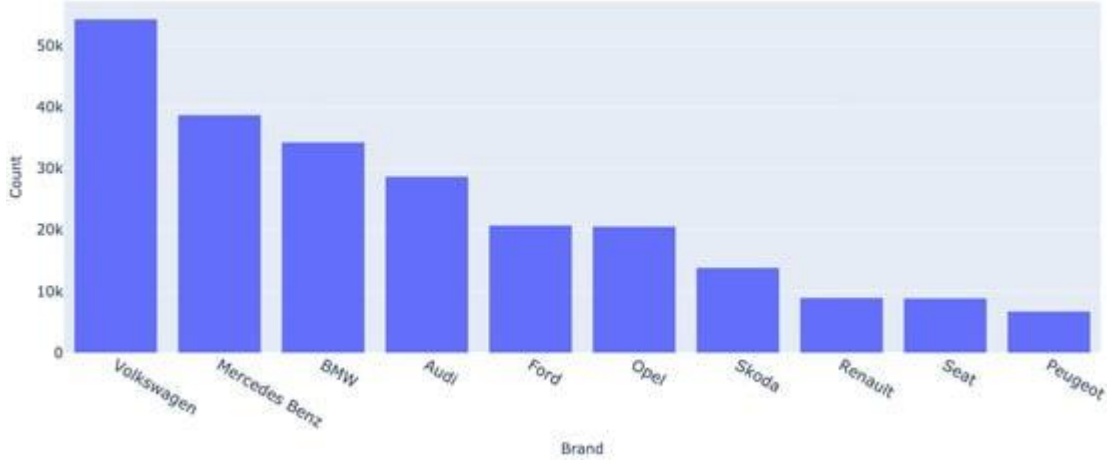

Рисунок 1.8 – Десять найпопулярніших брендів–Mobile.de [43]

Наступні функції мали такий самий розподіл і мали тенденцію слідувати тій самій моделі від одного набору даних до іншого, навіть якщо їхні значення не враховувалися під час створення секцій навчання та перевірки. Дані про рік випуску показують, що більшість транспортних засобів були виготовлені між 2015 і 2020 роками, приблизний вік яких становив близько 5 років на момент розміщення оголошення.

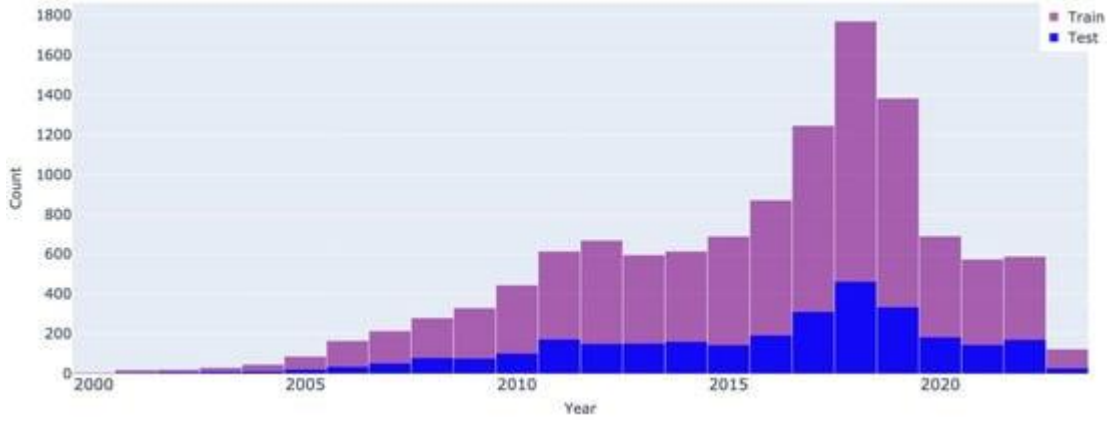

Рисунок 1.9 – Рік виробництва – Autovit.ro [43]

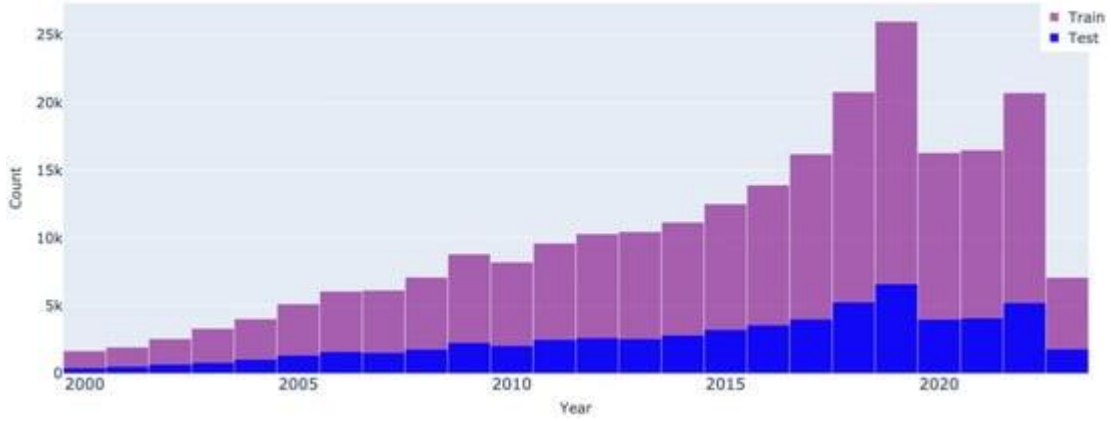

Рисунок 1.10 – Рік виробництва – Mobile.de [43]

Отже, між двома наборами даних спостерігалася різниця в категорії коробки передач. Пропозиція на Autovit.ro підкреслило значну перевагу автомобілів з автоматичною коробкою передач, приблизно на 50% більше автомобілів з автоматичною коробкою передач, ніж з ручною коробкою передач. Тип коробки передач, як правило, мав істотний вплив на ціну продажу. Однак різниця була невеликою на німецькому ринку, і цифри були збалансовані між двома класами.

1.4. Аналіз стану використання автоматичного створення моделей машинного навчання

Машинне навчання має багато переваг, успішне використання машинного навчання вимагає великих зусиль від експертів, враховуючи, що жоден алгоритм не може досягти високої продуктивності з усіма можливими проблемами [37]. Незважаючи на те, що дослідники знайомі з даними, їм все ще часто бракує досвіду машинного навчання, необхідного для застосування цих методів до великих джерел даних.

Дослідники можуть і працюють з експертами в галузі обробки даних [41], але інтерактивний процес зазвичай потребує багато часу та зусиль з обох сторін. Не тільки це, але дані та людський досвід, як правило, недоступні, особливо в окремих установах. Таким чином, важко розробити та розгорнути рішення машинного навчання, оскільки вся робота починається з тривалого процесу надання даних, продовжується пошуком потрібних співавторів і передбачає безперервну взаємодію між експертами з машинного навчання та експертами домену.

Автоматизація деяких компонентів, які вимагають досвіду людини, дозволить швидше створювати, перевіряти та розгортати рішення машинного навчання, а отже, легше отримувати переваги від покращення якості у предметній галузі. Керуючись цією метою в усіх галузях, AutoML став новою галуззю досліджень з метою автоматичної оптимізації частин конвеєра машинного навчання, як показано на рис. 1.11.

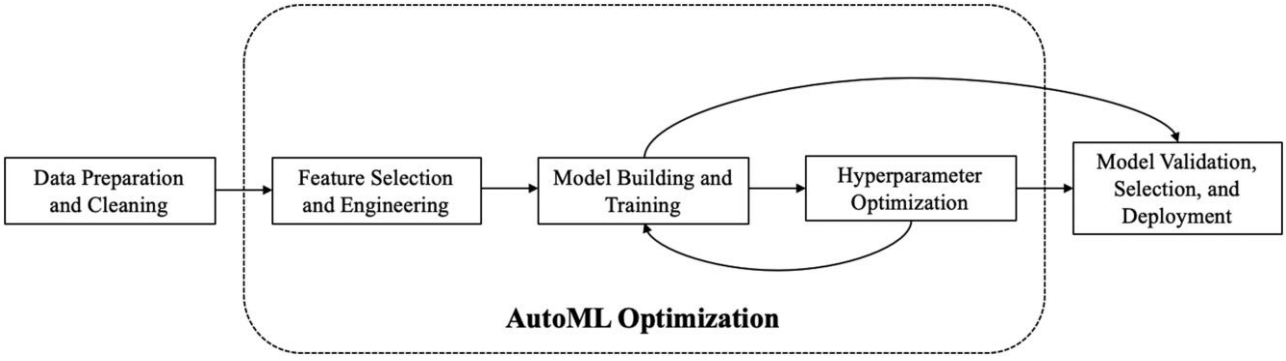

Рисунок 1.11 – Типові компоненти конвеєра задач машинного навчання

Перший крок полягає в підготовці даних. Це передбачає завантаження та очищення даних для використання в системі, а також застосування будь-яких перетворень, нормалізацій або кодувань.

Наступний крок включає вибір функцій, які будуть використані при створенні моделі. Це також може включати розробку функцій, яка є процесом використання знань предметної області для створення нових функцій, які допоможуть покращити модель машинного навчання. Наступні етапи передбачають ітераційний процес, під час якого створюється, тренується, оптимізується, перевіряється та вибирається даний алгоритм машинного навчання для використання для певної проблеми.

В останні роки з’явилися різні рішення AutoML для оптимізації одного або кількох із цих компонентів, деякі з яких є продуктом змагань AutoML Challenge у 2015–2018 роках. ChaLearn AutoML Challenges 1 зосереджено на вирішенні проблем машинного навчання під керівництвом без втручання людини, враховуючи деякі обчислювальні обмеження (рис. 1.12).

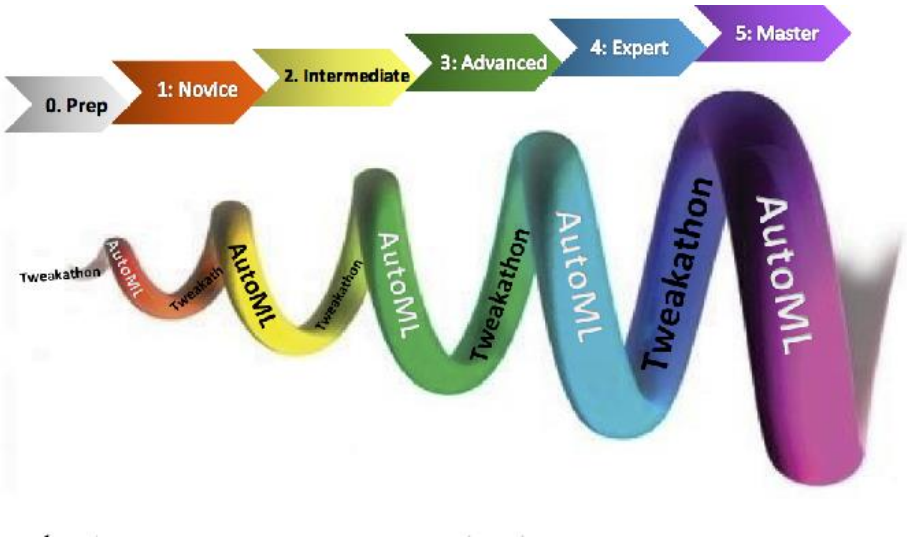

Рисунок 1.12 – Циклічність автоматичного машинного навчання

Хоча офіційне визначення та огляд AutoML існують, вони спрямовані на аудиторію широкого профілю. AutoML особливо корисний для різних сфер. Окремі автори намагаються автоматизувати – автоматизовану розробку

функцій, оптимізацію гіперпараметрів, оптимізатори конвеєрів (звертається на більше ніж один компонент) і пошук нейронної архітектури.

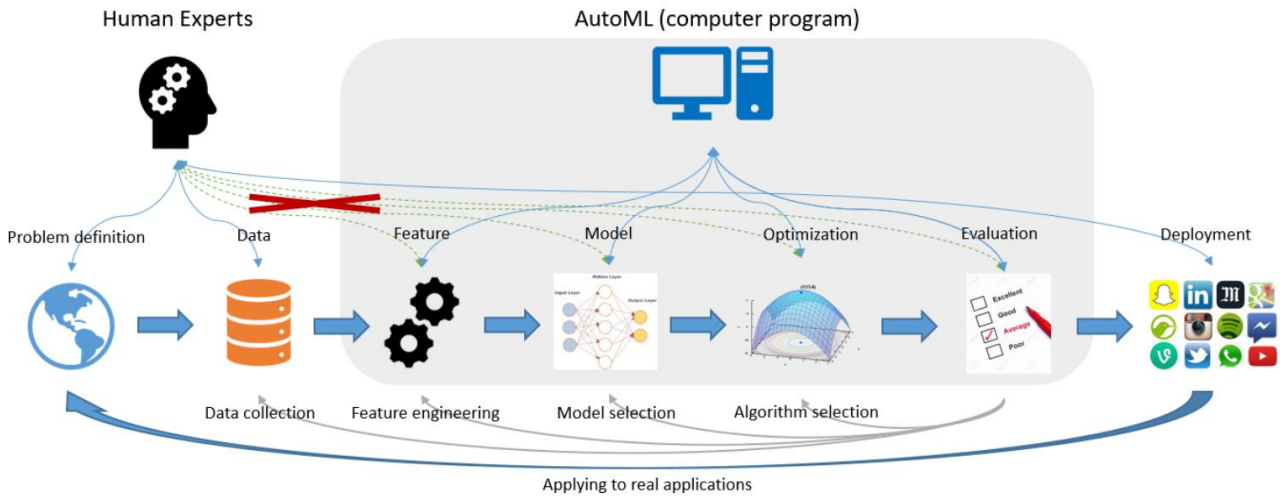

Рисунок 1.12 – Типовий конвеєр програми машинного навчання із використанням AutoML

Таким чином, автоматизоване машинне навчання (AutoML) не просто залишається академічною мрією, але й привертає більше уваги практиків. Якщо можна вилучити людину з машинного навчання під час створення додатків, то можна забезпечити швидке розгортання рішень машинного навчання в організаціях, ефективно перевірити та порівняти продуктивність розгорнутих рішень і змусити експертів більше зосереджуватися на проблемах із більшою додаткою та бізнес-цінністю. Це зробить машинне навчання набагато доступнішим для використання в реальному світі, що призведе до нових рівнів компетенції та налаштування.

1.5. Завдання кваліфікаційної роботи

На основі виконаного аналізу можна сказати, що тема кваліфікаційної роботи «Прогнозування вартості вживаних автомобілів із вибором програмної платформи автоматичного створення моделі машинного навчання» є

актуальною і має практичне значення. Зокрема, вона пов'язана з практичними аспектами нашого повсюдного життя, оскільки багато людей цікавляться покупкою та продажем вживаних автомобілів. Прогнозування вартості може допомогти як покупцям, так і продавцям зробити більш обґрунтовані рішення.

Усе свідчить про обрану тему, яка є достатньо цікавою, важливою та актуальною для дослідження. У роботі слід виконати наступні завдання:

1) виконати збір та аналіз даних про вартість вживаних автомобілів;

2) визначити ключові параметри, які впливають на вартість вживаних автомобілів та запропонувати їх автоматичну обробку;

3) вибрати програмну платформу для автоматичного створення моделі машинного навчання;

4) провести експерименти та створити модель для прогнозування вартості вживаних автомобілів;

5) порівняти обґрунтовану модель із аналогами та визначити показники точності прогнозування вартості вживаних автомобілів;

6) розробити заходи із охорони праці та визначити економічну ефективність.

Використання програмної платформи для автоматичного створення моделі машинного навчання відкриває можливості для розвитку ефективних та точних прогнозних моделей. Це важливо, інші ринки вживаних автомобілів можуть бути під впливом багатства факторів, і складні моделі можуть краще адаптуватися до змінених умов. Вибір правильної програмної платформи для автоматичного створення моделі машинного навчання для досягнення успіху проекту. Обґрунтований вибір платформи може сприяти зручності розробки та підтримки системи.

РОЗДІЛ 2.
ВИБІР ПРОГРАМНОЇ ПЛАТФОРМИ ТА МЕТОДІВ АВТОМАТИЧНОГО СТВОРЕННЯ МОДЕЛІ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ВАРТОСТІ ВЖИВАНИХ АВТОМОБІЛІВ

2.1. Доцільність та переваги AutoML

AutoML (автоматизоване машинне навчання) відноситься до автоматизованого наскрізного процесу застосування машинного навчання в реальних і практичних сценаріях. Протягом останніх багатьох років у галузі використовувалися переваги машинного навчання. Інструменти машинного навчання також розвивалися з часом із розширенням впровадження. Сьогодні люди можуть легко працювати з машинним навчанням завдяки його простим у використанні та зручним інструментам. Оскільки збір даних і перетворення їх у практичну інформацію достатньо автоматизовано, люди з деякими знаннями технологій і мотивацією можуть працювати з машинним навчанням.

Ці інструменти здатні виконувати повсякденну роботу зі збору даних, додавання структури та узгодженості, де це можливо, а потім починати обчислення. Сучасні інструменти можуть спростити процес збору даних і зберігання цієї інформації в рядках і стовпцях.

AutoML має кілька важливих переваг:

1. Підвищує ефективність, автоматично запускаючи повторювані завдання. Це дозволяє дослідникам даних більше зосереджуватися на проблемах, а не на моделях.

2. Передбачає використання автоматизованих конвеєрів ML, що допомагають уникнути можливих помилок, спричинених ручною роботою.

3. AutoML забезпечує демократизацію машинного навчання та дозволяє кожному використовувати функції ML.

4. Створює модель, виконує стратифіковану перехресну перевірку та оцінку показники моделі.

5. Автоматично налаштовує гіперпараметри моделі.
6. Аналізує продуктивність моделі за допомогою різних графіків.
7. Знайде найкращу модель наприкінці експерименту.
8. Робить прогнози на основі нових/невідомих даних.
9. Забезпечує збереження / завантаження моделі для майбутнього використання.

2.2. Вибір програмної платформи для прогнозування вартості вживаних автомобілів

AutoML має багаторічну історію. На даний час з'явилося багато чудових фреймворків AutoML. Коротко проаналізуємо кілька поширених фреймворків. Нижче наведено сім інструментів, які спрощують використання алгоритмів машинного навчання.

2.2.1. Бібліотека машинного навчання PyCaret

PyCaret – це бібліотека машинного навчання з відкритим вихідним кодом і низьким кодом на Python, мета якої скоротити час циклу від гіпотези до ідеї (рис. 2.1). Вона добре підходить для досвідчених спеціалістів із обробки даних, які хочуть підвищити продуктивність своїх експериментів з машинного навчання, використовуючи PyCaret у своїх робочих процесах.

Рисунок 2.1 – Бібліотека машинного навчання PyCaret [44]

2.2.2. Програмний пакет для автоматизованого машинного навчання SKLearn

Auto-SKLearn – це програмний пакет для автоматизованого машинного навчання, створений на основі scikit-learn. Auto-SKLearn звільняє користувача машинного навчання від вибору алгоритму та налаштування гіперпараметрів (рис. 2.2). Він включає такі методи розробки функцій, як One-Hot, стандартизація цифрових функцій і PCA . Модель використовує оцінювачі SKLearn для обробки проблем класифікації та регресії .

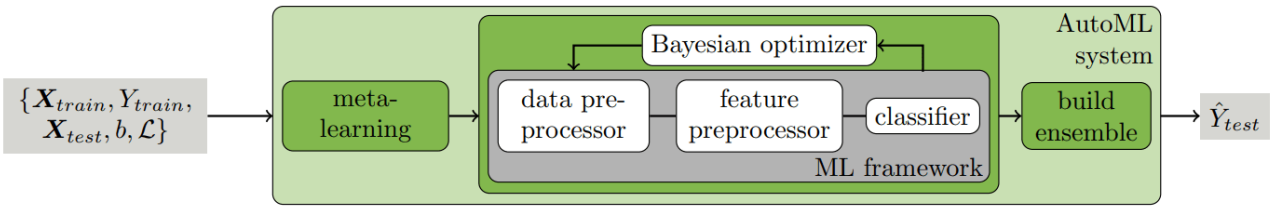

Рисунок 2.2 – Бібліотека машинного навчання Auto-SKLearn [45]

Auto-SKLearn створює конвеєр і використовує пошук Байєса для оптимізації цього каналу. У рамках ML додано два компоненти для налаштування гіперпараметрів за допомогою байєсівських міркувань: метанавчання використовується для ініціалізації оптимізаторів за допомогою Байєса та оцінки конструкції автоматичного збору конфігурації під час процесу оптимізації.

Auto-SKLearn добре працює в середніх і малих наборах даних, але він не може створювати сучасні системи глибокого навчання з найвищою продуктивністю у великих наборах даних.

2.2.3. Бібліотека машинного навчання MLBox

MLBox – це потужна бібліотека Python для автоматизованого машинного навчання. Згідно з офіційним документом, він забезпечує такі функції, як швидке читання та розподілена повторна обробка/очищення/форматування

даних , високонадійний вибір функцій і виявлення витоків, а також точну оптимізацію гіперпараметрів, найсучасніші прогнозні моделі для класифікації та регресії (Deep Learning, Stacking, LightGBM, ...), прогнозування з інтерпретацією моделі.

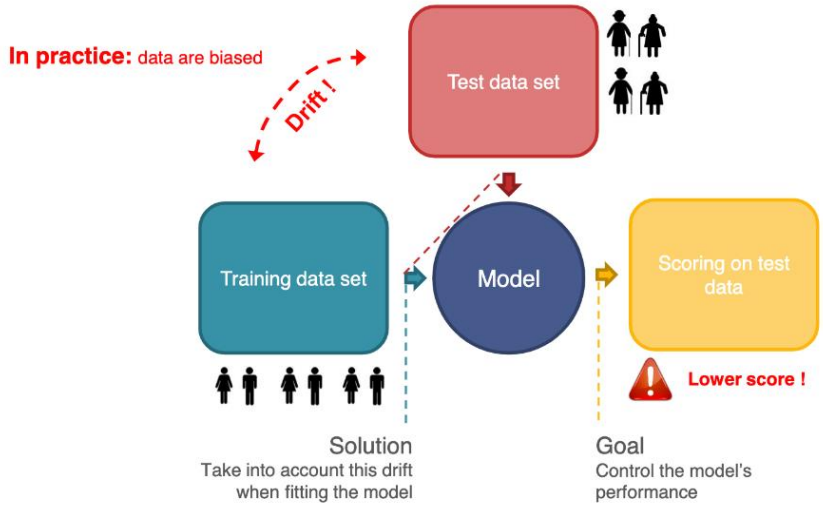

Рисунок 2.3 – Бібліотека машинного навчання MLBox [46]

Основний пакет MLBox містить 3 підпакети:

- ✓ Попередня обробка: читання та попередня обробка даних;
- ✓ Оптимізація: тестування або оптимізація широкого кола учнів;
- ✓ Прогноз: передбачення цілі на тестовому наборі даних.

2.2.4. Інструмент оптимізації конвеєрів TPOT

TPOT – це інструмент оптимізації конвеєрів на основі дерева, який використовує генетичні алгоритми для оптимізації конвеєрів машинного навчання. TPOT створено на основі scikit-learn і використовує власні методи регресора та класифікатора. TPOT досліджує тисячі можливих конвеєрів і знаходить той, який найкраще відповідає даним.

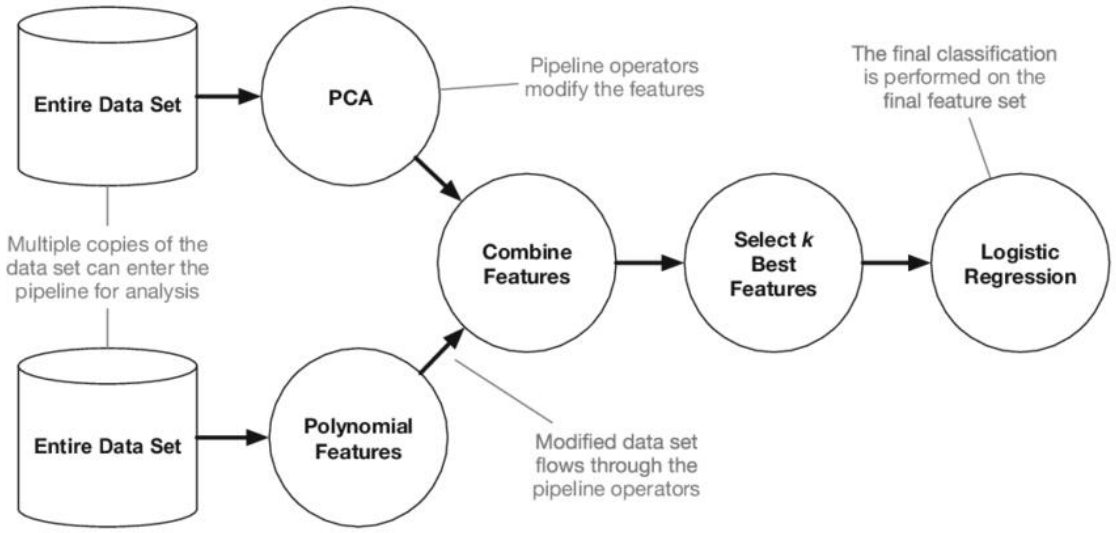

Рисунок 2.4 – Інструмент оптимізації конвеєрів на основі дерева TPOT [47]

2.2.5. Платформа H2O

H2O – це платформа з відкритим вихідним кодом розподіленого машинного навчання в пам'яті, розроблена H2O.ai. H2O підтримує як R, так і Python. Він підтримує найпоширеніші статистичні алгоритми та алгоритми машинного навчання, включаючи машини з градієнтним посиленням, узагальнені лінійні моделі, глибоке навчання тощо.

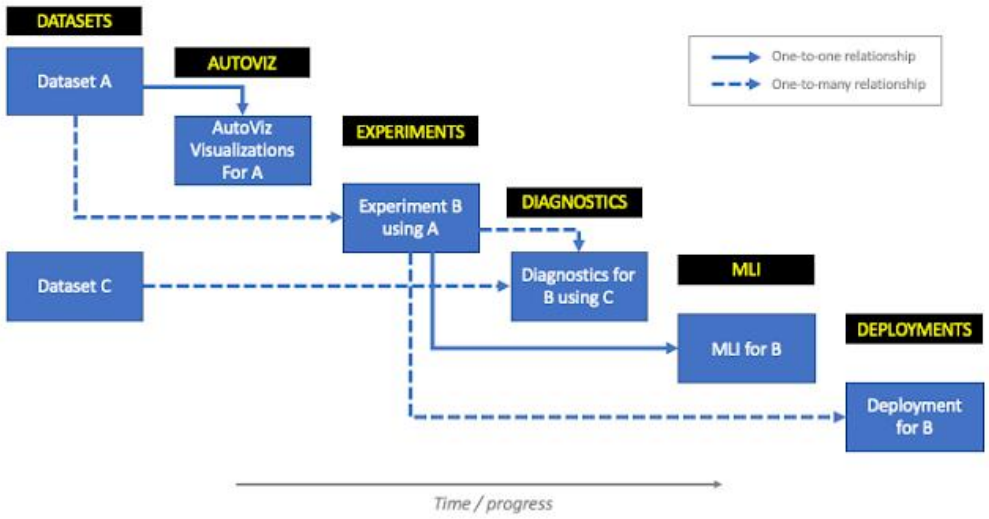

Рисунок 2.5 – Платформа H2O з відкритим вихідним кодом розподіленого машинного навчання [48]

H2O містить модуль автоматизованого машинного навчання та використовує власні алгоритми для створення конвеєрів. Він використовує вичерпний пошук методів розробки функцій і гіперпараметрів моделі для оптимізації конвеєрів.

H2O автоматизує деякі складні завдання з обробки даних і машинного навчання, такі як розробка функцій, перевірка моделі, коригування моделі, вибір моделі та розгортання моделі. Крім того, вона забезпечує автоматичну візуалізацію та інтерпретацію машинного навчання (MLI).

2.2.6. Бібліотека Auto-Keras

Auto-Keras – це бібліотека програмного забезпечення з відкритим кодом для автоматизованого машинного навчання (AutoML), розроблена DATA Lab. Auto-Keras, створена на основі інфраструктури глибокого навчання Keras, надає функції для автоматичного пошуку архітектури та гіперпараметрів моделей глибокого навчання.

Auto-Keras відповідає класичному дизайну Scikit-Learn API, тому його легко використовувати. Поточна версія надає функцію автоматичного пошуку гіперпараметрів під час глибокого навчання.

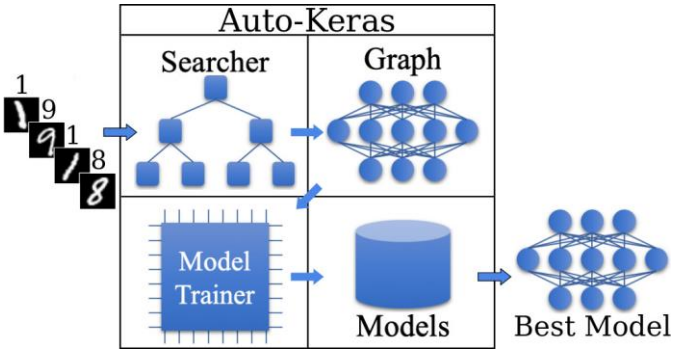

Рисунок 2.6 – Бібліотека Auto-Keras програмного забезпечення з відкритим кодом [49]

В Auto-Keras тенденція полягає в тому, щоб спростити машинне навчання за допомогою алгоритмів автоматичного пошуку нейронної архітектури (NAS). NAS в основному використовує набір алгоритмів, які автоматично налаштовують моделі, щоб замінити інженерів/практиків глибокого навчання.

2.2.7. Платформа DataRobot

DataRobot включає різноманітні методи регресії, починаючи від найпростіших (лінійна регресія) до складних статистичних класичних регресійних моделей і закінчуючи більш складними методами, включаючи посилення градієнта та нейронні мережі.

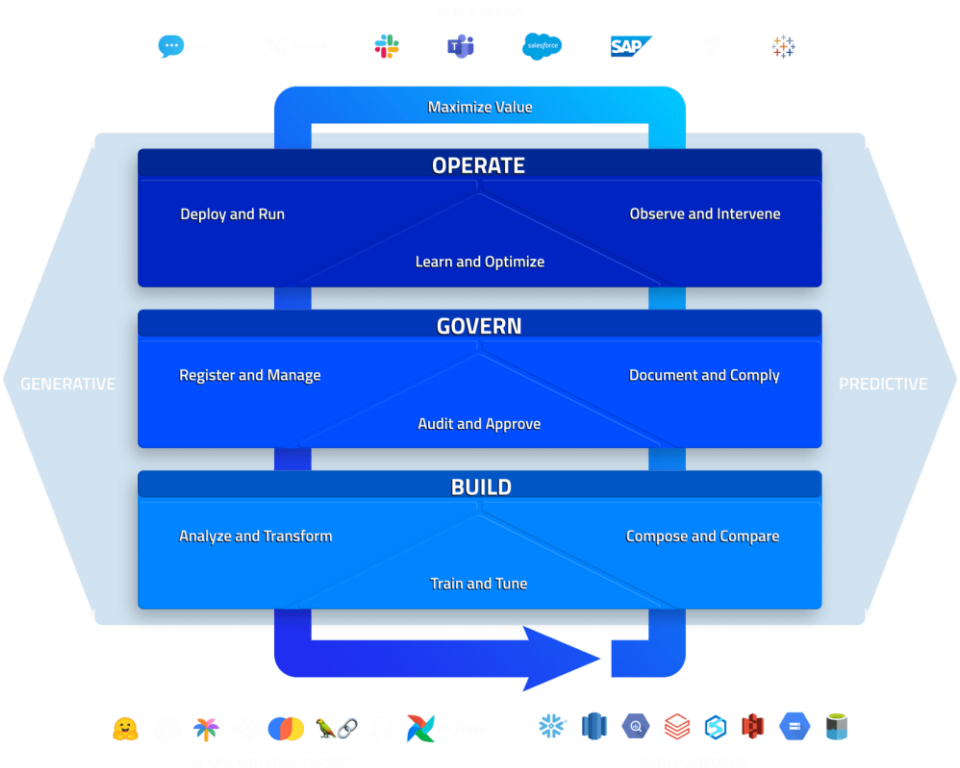

Рисунок 2.7 – Платформа DataRobot [50]

Платформа також може вирішувати прості задачі двійкової класифікації, а також дуже складні задачі багатокласової класифікації з до 100 різними категоріями.

За допомогою DataRobot легко автоматизувати створення таких моделей машинного навчання – з безпрецедентною прозорістю, щоб зрозуміти прогнози, які вони роблять.

Науковці із напрямку обробки даних можуть прискорити розробку моделей машинного навчання, використовуючи AutoML та реалізувати ефективне машинне навчання. Суть AutoML полягає в автоматизації повторюваних завдань, таких як створення конвеєра та налаштування гіперпараметрів, щоб дослідники даних могли витрачати більше часу на поточні бізнес-проблеми в практичних сценаріях. AutoML також дозволяє кожному замість невеликої групи людей використовувати ефективні технології машинного навчання.

У нашій роботі для прогнозування вартості вживаних автомобілів пропонується використовувати інструмент оптимізації конвеєрів на основі дерева TPOT. Це інструмент автоматичного машинного навчання, який можна використовувати для регресії. Він може використовуватися для пошуку найкращої архітектури моделі та параметрів для заданої задачі регресії. Щоб використовувати TPOT для регресії, необхідно створити новий проект TPOT і вибрати задачу регресії. Потім можна надати TPOT дані для навчання та тестування. TPOT буде використовувати ці дані для пошуку найкращої архітектури моделі та параметрів.

2.3. Методи створення моделі машинного навчання для прогнозування вартості вживаних автомобілів

TPOTRegressor виконує інтелектуальний пошук у конвеєрах машинного навчання, які можуть містити контрольовані регресійні моделі, препроцесори, методи вибору функцій і будь-які інші оцінювачі чи трансформатори, які належать API scikit-learn . TPORegressor також шукатиме гіперпараметри всіх об’єктів у конвеєрі.

За замовчуванням TPORegressor здійснюватиме пошук у широкому діапазоні керованих регресійних моделей, трансформаторів та їхніх гіперпараметрів. Однак моделі, трансформатори та параметри, які шукає TPORegressor, можна повністю налаштувати за допомогою config_dict параметра.

TPOTRegressor використовує наступні методи для пошуку найкращої архітектури моделі та параметрів. Генетичний алгоритм – це алгоритм пошуку, який генерує нові рішення шляхом комбінування та модифікації існуючих рішень. TPOTRegressor використовує генетичний алгоритм для генерування різних комбінацій архітектур моделей та параметрів.

Формула для TPOTRegressor виглядає наступним чином:

$$f(x) = score_cv(x), \tag{2.1}$$

де x – рішення, яке генерується TPOTRegressor; $score_cv(x)$ – функція оцінки, яка визначає, наскільки добре x вирішує задачу регресії за допомогою перехресної перевірки.

Функція $score_cv(x)$ може бути визначена різними способами. Наприклад, вона може бути заснована на метриці продуктивності моделі, такій як RMSE або R-квадрат.

Перехресна перевірка – це метод оцінки моделей, який використовує кілька наборів даних для навчання та оцінки моделі. TPOTRegressor використовує перехресну перевірку для оцінки ефективності різних комбінацій архітектур моделей та параметрів.

$$score = \frac{\sum_{i=1}^{k} score_{cv}(X_i, y_i)}{k}, \tag{2.2}$$

де k – кількість фолдів (піднаборів) у перехресній перевірці; X_i, y_i – показники ефективності, показники для *i*-го фолда.

Фільтрація – це метод пошуку, який відкидає рішення, які не відповідають певним критеріям. TPOTRegressor використовує фільтрацію для

відкидання рішень, які не досягають певного рівня продуктивності або мають занадто багато параметрів.

Формула для TPOTRegressor виглядає наступним чином:

$$x_i \in S \Leftrightarrow fcv(x_i) \geq t, \quad (2.3)$$

де x_i – рішення, яке генерується TPOTRegressor; S – множина допустимих рішень; $fcv(x_i)$ – оцінка ефективності моделі на i-му наборі даних за допомогою перехресної перевірки; t – порогове значення.

Функція $fcv(x_i)$ може бути визначена різними способами. Наприклад, вона може бути заснована на метриці продуктивності моделі, такій як RMSE або R-квадрат.

TPOTRegressor також використовує ряд інших методів для покращення ефективності пошуку, включаючи наступні. Збереження найкращих рішень – TPOTRegressor зберігає найкращі рішення, знайдені на кожному кроці, щоб уникнути повторного пошуку вже знайдених рішень. TPOTRegressor використовує різні розміри популяції, щоб досліджувати різні простори пошуку. TPOTRegressor використовує різні швидкості мутації, щоб генерувати різні варіанти рішень.

Ці методи допомагають TPOTRegressor знаходити ефективні архітектури моделей та параметри для різних задач машинного навчання.

РОЗДІЛ 3.
РЕЗУЛЬТАТИ ПІДГОТОВКИ ДАНИХ ТА АВТОМАТИЧНОГО СТВОРЕННЯ МОДЕЛІ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ВАРТОСТІ ВЖИВАНИХ АВТОМОБІЛІВ

3.1. Підготовка даних для прогнозування вартості вживаних автомобілів

Для прогнозування вартості вживаних автомобілів нами обрано доступний набір даних із платформи Kaggle [51]. Дані, що описують ціну доступних вживаних автомобілів на ринку із їх характеристиками, що є основою для виявлення взаємозв'язків між чинниками, що впливають на ціну автомобілів. Метою наших досліджень було підготувати дані та на підставі них виконати автоматичного створення моделі машинного навчання із використанням TPOTRegressor. Це забезпечує інтелектуальний пошук у конвеєрах машинного навчання, ефективної регресійної моделі.

У подальшому запропоновану модель можна використовувати для бізнес-планування, щоб зрозуміти, як саме змінюються ціни в залежності від незалежних змінних.

Із вибраного набору даних нами представлено їх атрибути, які характеризують окремі чинники впливу на вартість вживаних автомобілів:

1. car_ID – ідентифікатор автомобіля;
2. symboling – позначення;
3. CarName – марка автомобіля;
4. fueltype – тип пального;
5. aspiration – вид вприску;
6. doornumber – кількість дверей;
7. carbody – кузов;
8. drivewheel – привід;
9. enginelocation – розташування двигуна;

10. wheelbase – колісна база;
11. carlength – довжина автомобіля;
12. carwidth – ширина автомобіля;
13. carheight – висота автомобіля;
14. curbweight – споряджена маса автомобіля;
15. enginetype – тип двигуна;
16. cylindernumber – кількість циліндрів;
17. enginesize – об'єм двигуна автомобіля;
18. fuelsystem – паливна система;
19. boreratio – діаметр поршня;
20. stroke – хід поршня;
21. compressionratio – ступінь стиснення;
22. horsepower – потужність двигуна автомобіля;
23. peakrpm – частота обертання двигуна автомобіля;
24. citympg – міська витрата пального;
25. highwaympg – витрата пального по трасі;
26. price – вартість автомобіля;

Набір даних містить 26 змінних, які характеризують вживані автомобілі, та 1 змінну, яка вказує на вартість автомобіля. Набір даних використовується для вирішення задачі прогнозування вартості автомобіля. Для цього визначається залежність між змінними, що характеризують вартість автомобіля. Отримана залежність використовується для прогнозування вартості вживаних автомобіля на тестових даних.

Для вирішення задачі підготовки та аналізу даних про вартість вживаних автомобіля було використано Jupyter Notebook. Jupyter Notebook – це відкритий проєкт, який забезпечує розробку програмного забезпечення, що дозволяє проводити інтерактивні обчислення на різних мовах програмування.

Набір даних, який містить інформацію про вартість вживаних автомобілів, було завантажено у блокнот Jupyter Notebook. Для цього було імпортовано необхідні бібліотеки, як показано на рис. 3.1.

```
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        from sklearn.model_selection import train_test_split
```

Рисунок 3.1 – Завантаження бібліотек у блокнот Jupyter Notebook

Далі ми завантажили та вивели масив початкових даних (рис. 3.2).

```
In [2]: df=pd.read_csv('F:\Dipl_2023\Dipl_mag_126\Досяк/CarPrice_Assignment.csv')
        df.head()
```

Out[2]:

	car_ID	symboling	CarName	fueltype	aspiration	doornumber	carbody	drivewheel	enginelocation	wheelbase	...	enginesize	fuelsystem	boreratio
0	1	3	alfa-romero giulia	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47
1	2	3	alfa-romero stelvio	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47
2	3	1	alfa-romero Quadrifoglio	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	2.68
3	4	2	audi 100 ls	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	3.19
4	5	2	audi 100ls	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	3.19

5 rows × 26 columns

Рисунок 3.2 – Результати представлення фрагменту масиву із початковими даними

Після цього ми провели детальний аналіз масиву із початковими даними. Встановлено загальні характеристики масиву даних про вартість вживаних автомобілів (рис. 3.3). Зокрема, маємо:

- ✓ Кількість екземплярів – 205 од;
- ✓ Кількість стовпців – 26 од;

Типи даних:

- ✓ Числові – float64 (8 стовпців), int64 (8 стовпців);
- ✓ Текстові – object (10 стовпців);
- ✓ Відсутні значення – Немає;
- ✓ Розмір пам'яті – 41.8+ KB.

```
In [5]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 26 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   car_ID            205 non-null    int64
 1   symboling         205 non-null    int64
 2   CarName           205 non-null    object
 3   fueltype          205 non-null    object
 4   aspiration        205 non-null    object
 5   doornumber        205 non-null    object
 6   carbody           205 non-null    object
 7   drivewheel        205 non-null    object
 8   enginelocation    205 non-null    object
 9   wheelbase         205 non-null    float64
 10  carlength         205 non-null    float64
 11  carwidth          205 non-null    float64
 12  carheight         205 non-null    float64
 13  curbweight        205 non-null    int64
 14  enginetype        205 non-null    object
 15  cylindernumber    205 non-null    object
 16  enginesize        205 non-null    int64
 17  fuelsystem        205 non-null    object
 18  boreratio         205 non-null    float64
 19  stroke            205 non-null    float64
 20  compressionratio  205 non-null    float64
 21  horsepower        205 non-null    int64
 22  peakrpm           205 non-null    int64
 23  citympg           205 non-null    int64
 24  highwaympg        205 non-null    int64
 25  price             205 non-null    float64
dtypes: float64(8), int64(8), object(10)
memory usage: 41.8+ KB
```

Рисунок 3.3 – Результати детального аналізу масиву із початковими даними

Наступний крок передбачає очищення та підготовку даних (рис 3.4) – це процес, який полягає у виправленні помилок, заповненні пропущених значень та перетворенні даних у формат, придатний для подальшого аналізу. Цей процес є важливим етапом будь-якого дослідження, оскільки він впливає на якість отриманих результатів.

Помилки у даних можуть виникнути з різних причин, наприклад, через людський фактор, технічні збої або неякісний збір даних. Визначення наявності помилок є першим кроком у процесі очищення даних.

Помилки у даних можна виправити вручну або за допомогою автоматизованих методів. Вибір методу залежить від типу помилок та їхньої кількості.

Пропущені значення можуть виникнути з різних причин, наприклад, через неповний збір даних або помилки введення даних. Пропущені значення можна заповнити за допомогою різних методів, наприклад, шляхом усереднення, використання найближчих сусідів або застосування моделей машинного навчання.

```
In [6]: #Splitting company name from CarName column
        CompanyName = df['CarName'].apply(lambda x : x.split(' ')[0])
        df.insert(3,"CompanyName",CompanyName)
        df.drop(['CarName'],axis=1,inplace=True)
        df.head()
```

Out[6]:

	car_ID	symboling	CompanyName	fueltype	aspiration	doornumber	carbody	drivewheel	enginelocation	wheelbase	...	enginesize	fuelsystem	boreratio
0	1	3	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47
1	2	3	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47
2	3	1	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	2.68
3	4	2	audi	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	3.19
4	5	2	audi	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	3.19

5 rows × 26 columns

```
In [7]: df.CompanyName.unique()

Out[7]: array(['alfa-romero', 'audi', 'bmw', 'chevrolet', 'dodge', 'honda',
               'isuzu', 'jaguar', 'maxda', 'mazda', 'buick', 'mercury',
               'mitsubishi', 'Nissan', 'nissan', 'peugeot', 'plymouth', 'porsche',
               'porcshce', 'renault', 'saab', 'subaru', 'toyota', 'toyouta',
               'vokswagen', 'volkswagen', 'vw', 'volvo'], dtype=object)
```

Рисунок 3.4 – Результати аналізу наявності помилок у даних

Встановлено, що у стовпці «CompanyName» допущено помилки у їх описі. Зокрема, зустрічаються різні представлення даних:

maxda = мазда

Nissan = nissan

porsche = porcshce

toyota = toyouta

vokswagen = volkswagen = vw

Виправлення невірних значень виконуємо із використанням коду (рис. 3.5), що забезпечує очищення та стандартизацію марки автомобіля у стовпці «CompanyName» датафрейму df.

```
In [8]: df.CompanyName = df.CompanyName.str.lower()

        def replace_name(a,b):
            df.CompanyName.replace(a,b,inplace=True)

        replace_name('maxda','mazda')
        replace_name('porcshce','porsche')
        replace_name('toyouta','toyota')
        replace_name('vokswagen','volkswagen')
        replace_name('vw','volkswagen')

        df.CompanyName.unique()

Out[8]: array(['alfa-romero', 'audi', 'bmw', 'chevrolet', 'dodge', 'honda',
               'isuzu', 'jaguar', 'mazda', 'buick', 'mercury', 'mitsubishi',
               'nissan', 'peugeot', 'plymouth', 'porsche', 'renault', 'saab',
               'subaru', 'toyota', 'volkswagen', 'volvo'], dtype=object)
```

Рисунок 3.5 – Результати виправлення невірних значень у масиві із початковими даними

Отже, очищення та підготовка даних є важливим етапом будь-якого дослідження, який дозволяє отримати якісні результати та уникнути помилок.

3.2. Аналіз даних для прогнозування вартості вживаних автомобілів

Нами побудовано два графіки, які характеризують розподіл цін на вживані автомобілі – гістограми та діаграми розмаху (рис. 3.6).

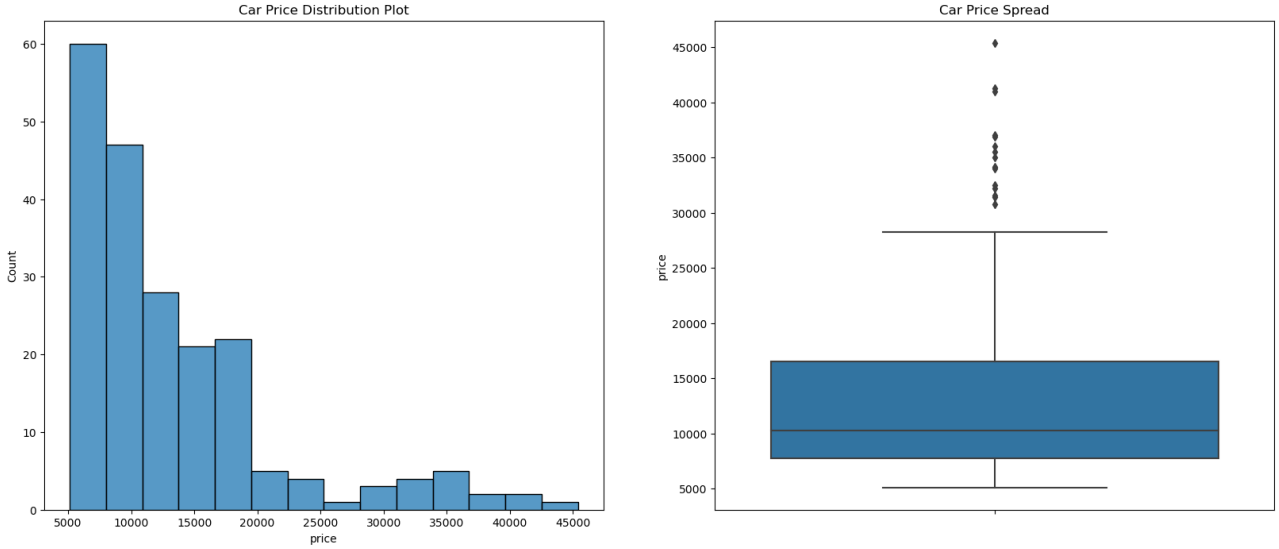

Рисунок 3.6 – Результати кількісного оцінення вартості вживаних автомобілів

Встановлено, що більшість автомобілів у наборі даних мають ціну нижчою 15000$. Існує значна різниця між середнім значенням та медіаною розподілу вартості вживаних автомобілі. Точки даних далеко розкидані від середнього значення, що свідчить про велику дисперсію вартості вживаних автомобілів (85% цін нижче 18500$, тоді як решта 15% знаходяться в діапазоні від 18500$ до 45400$).

На наступному етапі проводимо аналіз категорійних даних, які впливають на вартість вживаних автомобілів (рис. 3.7).

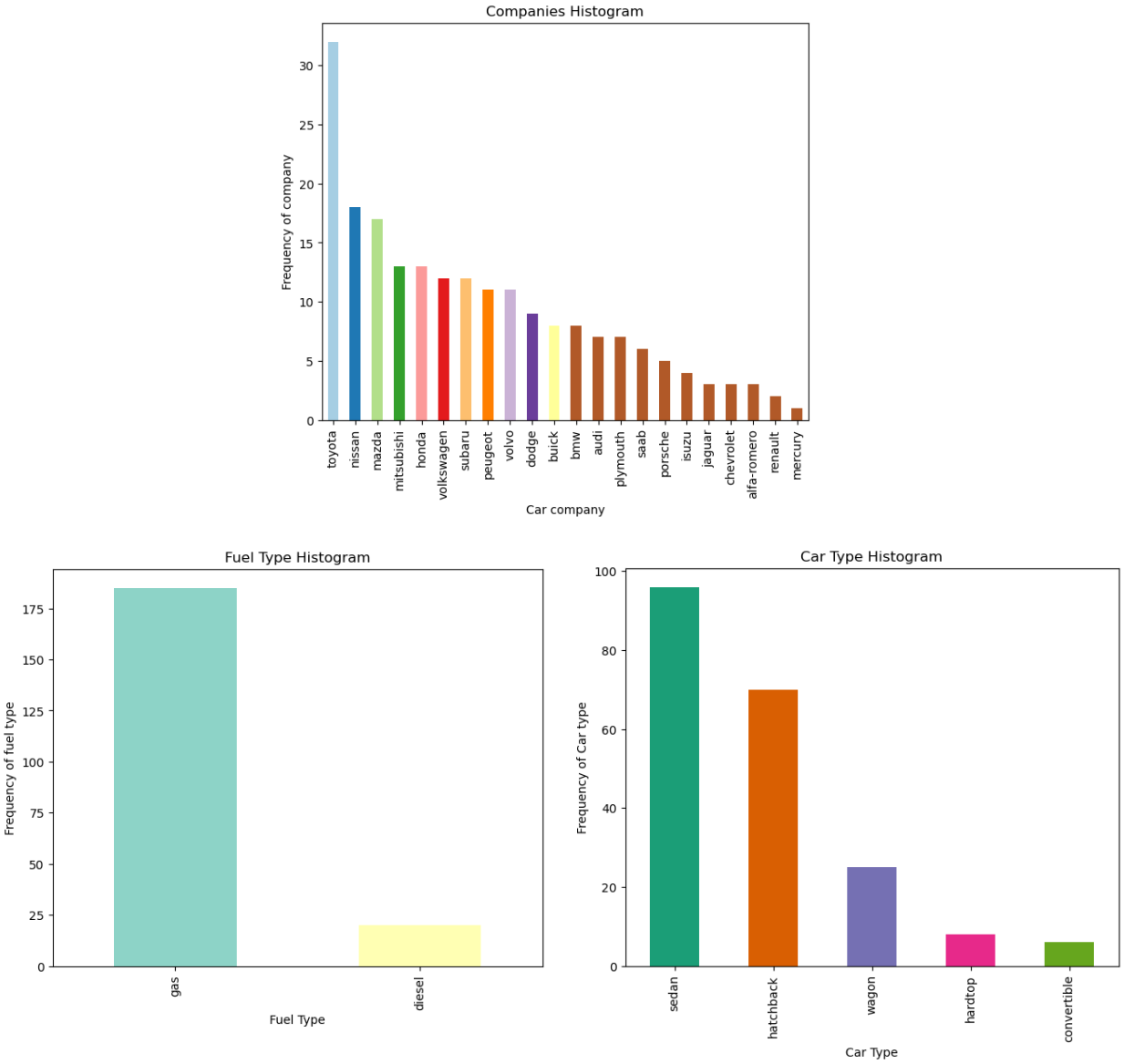

Рисунок 3.7 – Тенденції зміни категорійних даних, які впливають на вартість вживаних автомобілів

Встановлено, що найбільше даних є про автомобілі «Тойота». Кількість автомобілів на бензиновому паливі більша, ніж на дизельному. Седан є найпопулярнішим типом автомобіля.

Також нами проаналізовано вплив інших чинників на вартість вживаних автомобілів. Результати представлено на рис. 3.8.

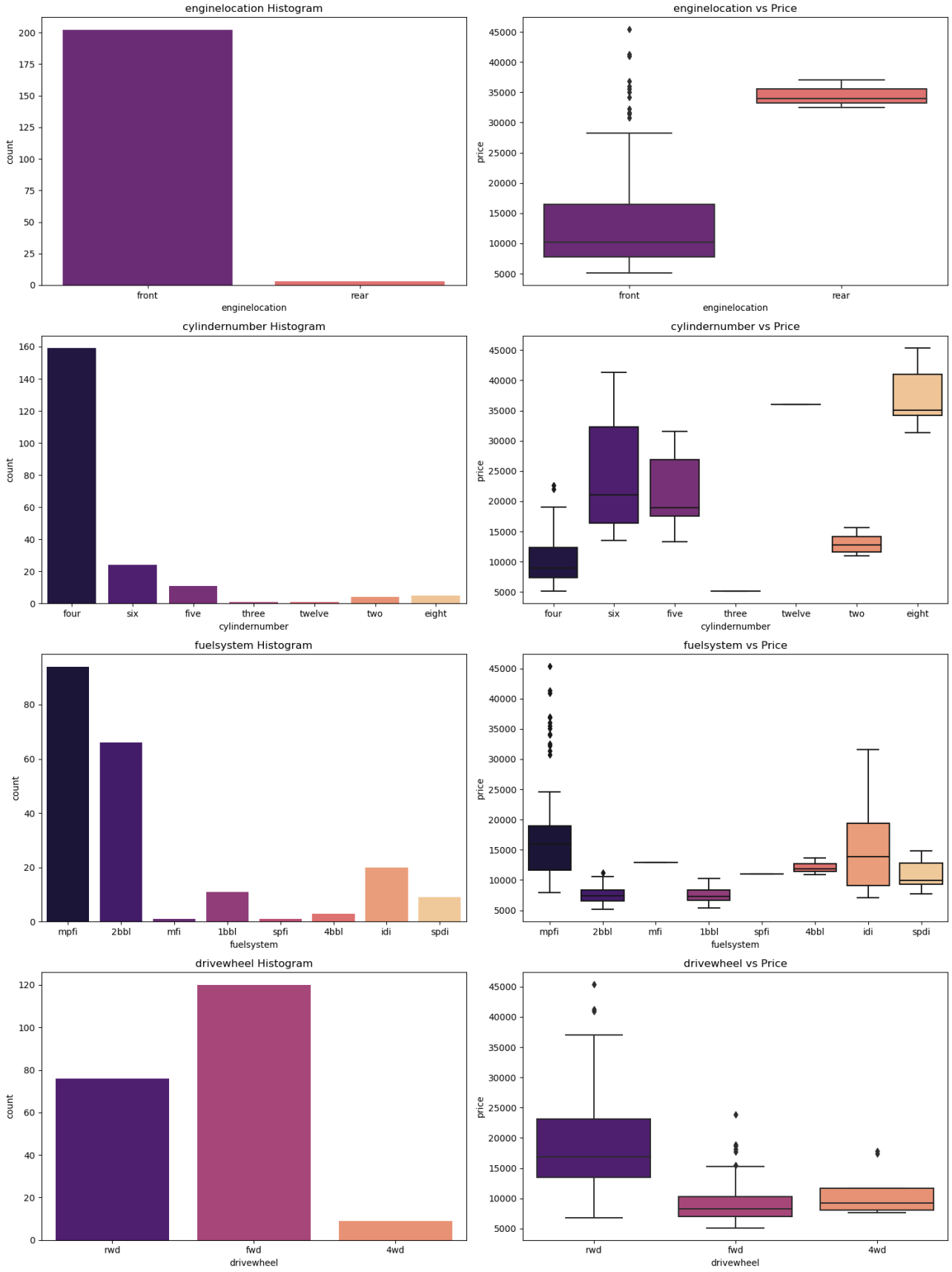

Рисунок 3.8 – Тенденції зміни чинників, які впливають на вартість вживаних автомобілів

Встановлено, що дуже мало даних для категорій розташування двигунів, щоб зробити висновок. Найпоширеніша кількість циліндрів – чотири, шість та п'ять. Хоча вісім циліндрів мають найвищий ціновий діапазон. Найпоширенішими типами паливних систем є mpfi та 2bbl. mpfi та idi мають найвищий ціновий діапазон. Але для інших категорій даних недостатньо, щоб зробити якісь значущі висновки. Дуже суттєва різниця в категорії ведучих коліс. Більшість автомобілів високого цінового діапазону, мають ручне керування.

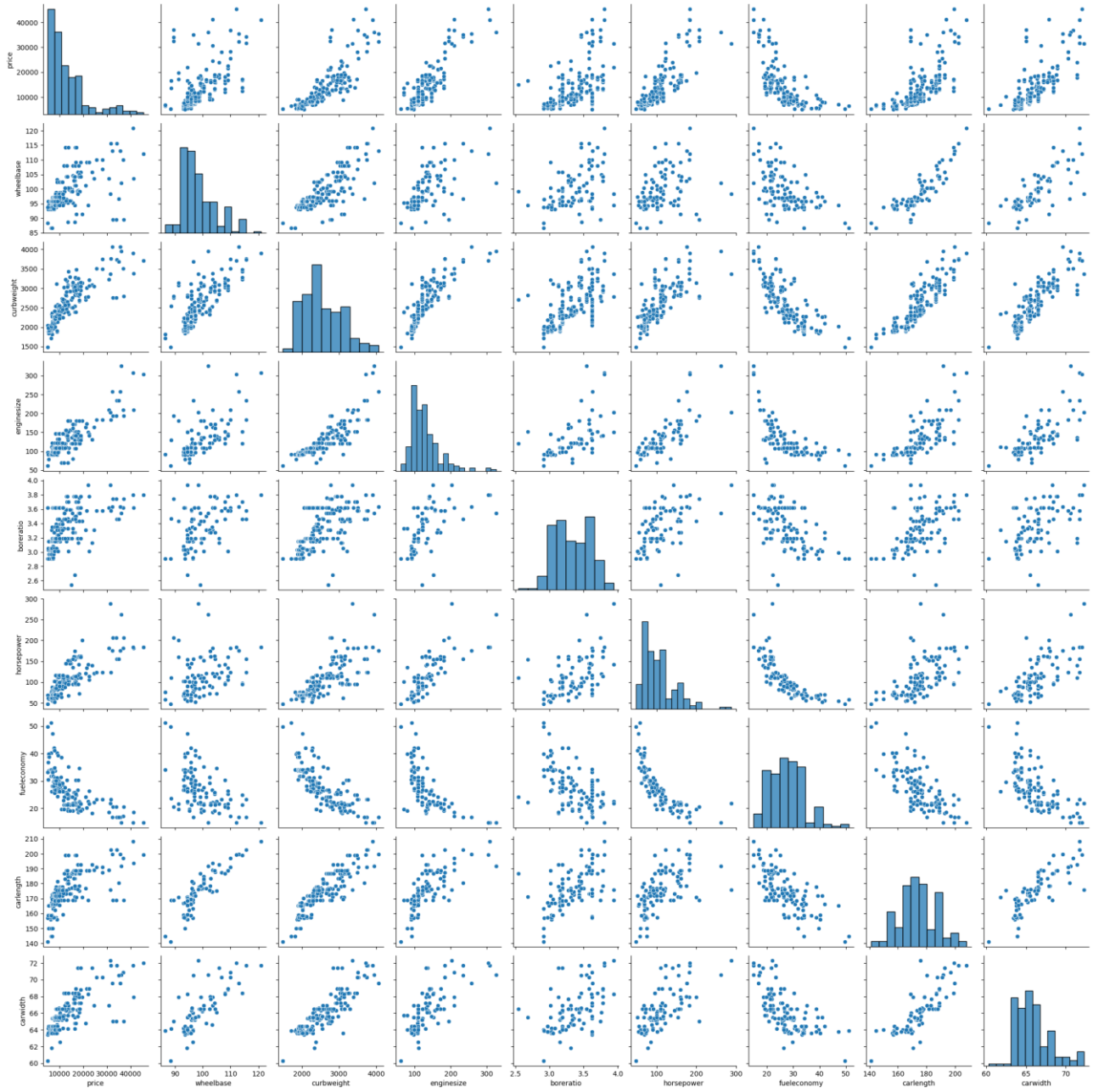

Рисунок 3.9 – Тенденції зміни чинників, які впливають на вартість вживаних автомобілів

Нами створено матрицю парних графіків для візуалізації взаємозв'язків між усіма числовими змінними у датафрейму, що характеризує вартість вживаних автомобілів (рис. 3.9). Діагональ матриці містить гістограми розподілу кожної змінної. Верхній трикутник (крім діагоналі) містить діаграми розсіювання, що показують зв'язок між парами змінних. Нижній трикутник (крім діагоналі) містить криві оцінки щільності Kernel Density Estimation (KDE), які відображають розподіл спільної ймовірності пар змінних.

Матриця парних графіків дозволяє швидко оцінити наявність сильних або слабких кореляцій між змінними, а також виявити потенційні нелінійні зв'язки. Ця інформація може бути корисною для вибору відповідних методів машинного навчання та інтерпретації результатів.

3.3. Створення моделі машинного навчання для прогнозування вартості вживаних автомобілів

Для створення моделі машинного навчання з метою прогнозування вартості вживаних автомобілів пропонується використати TPOTRegressor. Це потужний інструмент, який використовують для автоматизації процесу створення моделей машинного навчання. Він може бути корисним для будь-якого, хто хоче створювати ефективні моделі машинного навчання.

TPOTRegressor працює наступним чином. Він спочатку генерує випадковий набір моделей машинного навчання. Потім він оцінює ці моделі на даних навчання та зберігає найкращу модель. Після цього використовує генетичне програмування для мутації та рекомбінації найкращої моделі, щоб створити нові моделі. Повторюється цей процес до досягнення заданої кількості поколінь або до того, як буде знайдена модель, яка не може бути покращена.

```
In [69]: from tpot import TPOTRegressor
         from sklearn.model_selection import train_test_split
         from sklearn.metrics import mean_squared_error

         # Split the data into training and testing sets
         X_train_new, X_test_new, y_train, y_test = train_test_split(X_train_new, y_train, test_size=0.2, random_state=42)

         # Create and fit the TPOT regressor
         tpot = TPOTRegressor(generations=5, population_size=20, random_state=42, verbosity=2)
         tpot.fit(X_train_new, y_train)

         # Access the best pipeline and its characteristics
         best_pipeline = tpot.fitted_pipeline_
         print("Selected Model Name:", best_pipeline.steps[-1][1].__class__.__name__)

         # If available, print hyperparameters of the selected model
         if 'decision_function' in best_pipeline.named_steps:
             print("Selected Model Hyperparameters:", best_pipeline.named_steps['decision_function'].get_params())
         elif 'predict' in best_pipeline.named_steps:
             print("Selected Model Hyperparameters:", best_pipeline.named_steps['predict'].get_params())

         # Evaluate the performance on the test set
         y_pred = tpot.predict(X_test_new)
         mse = mean_squared_error(y_test, y_pred)
         print(f'Mean Squared Error: {mse}')
```

Рисунок 3.10 – Код із використанням бібліотеки TPOT для автоматизованої процесу побудови моделі прогнозування вартості вживаних автомобілів

Цей код (рис. 3.10) використовує бібліотеку TPOT для автоматизованої оптимізації процесу побудови моделі регресії для прогнозування вартості вживаних автомобілів. Насамперед виконується розділення даних на поточний та тестовий набори – X_train_new, X_test_new, y_train, y_test = train_test_split(X_train_new, y_train, test_size=0.2, random_state=42).

При цьому передбачено, що train_test_split забезпечує розділення даних на навчальний та тестовий набори із співвідношенням 80/20.

На наступному кроці виконується побудова та оптимізація моделі з використанням TPOT – tpot = TPOTRegressor(generations=5, population_size=20, random_state=42, verbosity=2) tpot.fit(X_train_new, y_train).

У цій кроці використовується TPORegressor для побудови та оптимізації моделей. Генерація та розмір населення забезпечують кількість поколінь та розмір кожного покоління в оптимізаційному процесі. Параметр verbosity встановлює рівень виведення інформації під час оптимізації.

Передбачається отримання кращого пайплайну та його характеристик – best_pipeline = tpot.fitted_pipeline_ print("Selected Model Name:", best_pipeline.steps[-1][1].__class__.__name__). Тут вибирається найкращий пайплайн після завершення оптимізації, і виводиться назва вибраної моделі.

Пайплайн машинного навчання – це послідовність етапів, які застосовуються до даних для створення моделі машинного навчання. Пайплайни можуть бути корисними для автоматизації процесу створення моделей машинного навчання, а також для забезпечення того, щоб модель була створена на чистих і якісних даних.

На наступному кроці виконується виведення гіперпараметрів вибраної моделі – if 'decision_function' in best_pipeline.named_steps: print("Selected Model Hyperparameters:", best_pipeline.named_steps['decision_function'].get_params()) elif 'predict' in best_pipeline.named_steps: print("Selected Model Hyperparameters:", best_pipeline.named_steps['predict'].get_params()).

Якщо вибрана модель має метод decision_function, то виводяться її гіперпараметри, інакше, якщо є метод predict, виводяться гіперпараметри цього методу.

Після цього нами передбачено оцінку продуктивності моделі на тестовому наборі даних – y_pred = tpot.predict(X_test_new) mse = mean_squared_error(y_test, y_pred) print(f'Mean Squared Error: {mse}').

Проведена оцінка ефективності моделі на тестовому наборі даних за допомогою середньоквадратичної помилки.

Запропонований код використовує TPOT для автоматичного визначення та оптимізації структури моделі з використанням генетичного програмування. Він автоматично вибирає найкращий алгоритм та його гіперпараметри для задачі регресії на основі вказаних параметрів оптимізації.

3.4. Результати обґрунтування моделі машинного навчання для прогнозування вартості вживаних автомобілів

На підставі машинного навчання із використанням бібліотеки TPOT для автоматизованої процесу побудови моделі прогнозування вартості вживаних автомобілів (рис. 3.10) нами отримано характеристики цієї моделі (табл. 3.1).

Таблиця 3.1 – Характеристики розробленої моделі прогнозування вартості вживаних автомобілів

Показники	Значення
Модель	ElasticNetCV
Середньоквадратична помилка (Mean Squared Error)	0.0053
Найкращий внутрішній CV скор у поколіннях 1-4	-0.0049
Найкращий внутрішній CV скор у поколінні 5	-0.0046
Найкращий пайплайн	ElasticNetCV(ZeroCount(GradientBoostingRegressor (input_matrix, alpha=0.9, learning_rate=0.01, loss=quantile, max_depth=10, max_features=0.25, min_samples_leaf=9, min_samples_split=12, n_estimators=100, subsample=0.9000000000000001)), l1_ratio=0.5, tol=1e-05)

Встановлено, що найкращою виявилася для наших даних модель ElasticNetCV, яка використовує алгоритм генетичного програмування для автоматичного пошуку найкращої моделі машинного навчання для конкретної задачі. Вона входить до складу бібліотеки TPOT, яка є інструментом для автоматизованого машинного навчання. ElasticNetCV – це модель регресії, яка поєднує в собі L1 та L2 регуляризатори. L1 регуляризатор сприяє скороченню кількості ненульових коефіцієнтів у моделі, а L2 регуляризатор сприяє скороченню значень коефіцієнтів.

Для наших даних ElasticNetCV було використано для прогнозування ціни вживаного автомобіля на основі його характеристик. Найкращий пайплайн, який знайшов ElasticNetCV, включає в себе GradientBoostingRegressor, який є моделлю ансамблю, що складається з декількох простих моделей. GradientBoostingRegressor – це ефективна модель для задач регресії з множинними кореляційними змінними.

Встановлено, що середньоквадратична помилка (Mean Squared Error) для найкращого пайплайна становить 0,0053. Це означає, що середня різниця між прогнозами моделі та фактичними значеннями становить 0,0053. Це досить низька помилка, що свідчить про те, що модель може ефективно прогнозувати вартість вживаного автомобіля.

У подальшому виконано алгоритм ансамблювання, тобто використано не одну модель XGBoost. Це забезпечило можливість підібрати найкращі базові алгоритми стекінгу. Стекінг (Stacked Generalization або Stacking) полягає у використанні базових класифікаторів, що забезпечують прогнозування (мета-ознак) та використання їх як ознак для деякого «узагальнюючого» алгоритму (мета-алгоритму). Тобто, основною ідеєю стекінгу є перетворення вихідної ознаки відповідно до завдання у нову ознаку, точками якої є прогнозування базових алгоритмів.

Для виконання стекінгу насамперед виконується вибір окремих пар довільних підмножин даних із навчальної вибірки. Після цього кожну із них слід навчити за базовими алгоритми та прогнозувати із їх використанням цільову змінну. Отримані значення стають об'єктами виконання нового прогнозування.

У подальшому нами створено код, що використовує бібліотеку matplotlib та numpy для візуалізації кривої навчання (learning curve) та аналізу точності прогнозів для моделей, що навчалася (рис. 3.11).

Насамперед нами здійснено визначення функції plot_learning_curve для побудови кривої навчання. Це функція додаткового оцінювача (estimator), назва графіка (title), дані для навчання та цільову зміну (Xi y), можливість задати обмеження для осі (ylim), параметри для крос-валідації (cv, n_jobs), і розміри поточного набору (train_sizes). Вона виводить графік кривої навчання, який показує, як змінюється точність моделі зі збільшенням об’єму тренувальних даних.

```
In [70]: import matplotlib.pyplot as plt
import numpy as np
from sklearn.model_selection import learning_curve

# Visualize the learning curve
def plot_learning_curve(estimator, title, X, y, ylim=None, cv=None, n_jobs=None, train_sizes=np.linspace(0.1, 1.0, 5)):
    plt.figure()
    plt.title(title)
    if ylim is not None:
        plt.ylim(*ylim)
    plt.xlabel("Training examples")
    plt.ylabel("Score")
    train_sizes, train_scores, test_scores = learning_curve(
        estimator, X, y, cv=cv, n_jobs=n_jobs, train_sizes=train_sizes)
    train_scores_mean = np.mean(train_scores, axis=1)
    train_scores_std = np.std(train_scores, axis=1)
    test_scores_mean = np.mean(test_scores, axis=1)
    test_scores_std = np.std(test_scores, axis=1)
    plt.grid()

    plt.fill_between(train_sizes, train_scores_mean - train_scores_std,
                     train_scores_mean + train_scores_std, alpha=0.1,
                     color="r")
    plt.fill_between(train_sizes, test_scores_mean - test_scores_std,
                     test_scores_mean + test_scores_std, alpha=0.1, color="g")
    plt.plot(train_sizes, train_scores_mean, 'o-', color="r",
             label="Training score")
    plt.plot(train_sizes, test_scores_mean, 'o-', color="g",
             label="Cross-validation score")

    plt.legend(loc="best")
    return plt

# Plot learning curve
title = "Learning Curve (LinearSVR)"
plot_learning_curve(tpot, title, X_train_new, y_train, cv=5)
plt.show()

# Plot prediction accuracy
plt.scatter(y_test, y_pred, color='blue')
plt.plot([min(y_test), max(y_test)], [min(y_test), max(y_test)], linestyle='--', color='red', linewidth=2)
plt.xlabel('Actual Prices')
plt.ylabel('Predicted Prices')
plt.title('Actual Prices vs. Predicted Prices')
plt.show()
```

Рисунок 3.11 – Код для візуалізації кривої навчання (learning curve) та аналізу точності прогнозів для моделей

Функція plot_learning_curve стосується моделі, яку попередньо навчили. Отриманий графік відображає вплив розміру поточного набору на точність моделі. На графіку, відображено фактичну вартість вживаних автомобілів (y_test) та прогнозовану вартість вживаних автомобілів (y_pred) для тестового набору даних. Також показана лінія ідентичності для порівняння того, наскільки добре навчена модель для прогнозування вартості вживаних автомобілів.

Нами здійснено побудову графіка поколінь та визначення їх кращих показників CV у порівняння з реальними значеннями (рис. 3.12).

У цьому прикладі використовуються реальні значення для поколінь від 1 до 15.

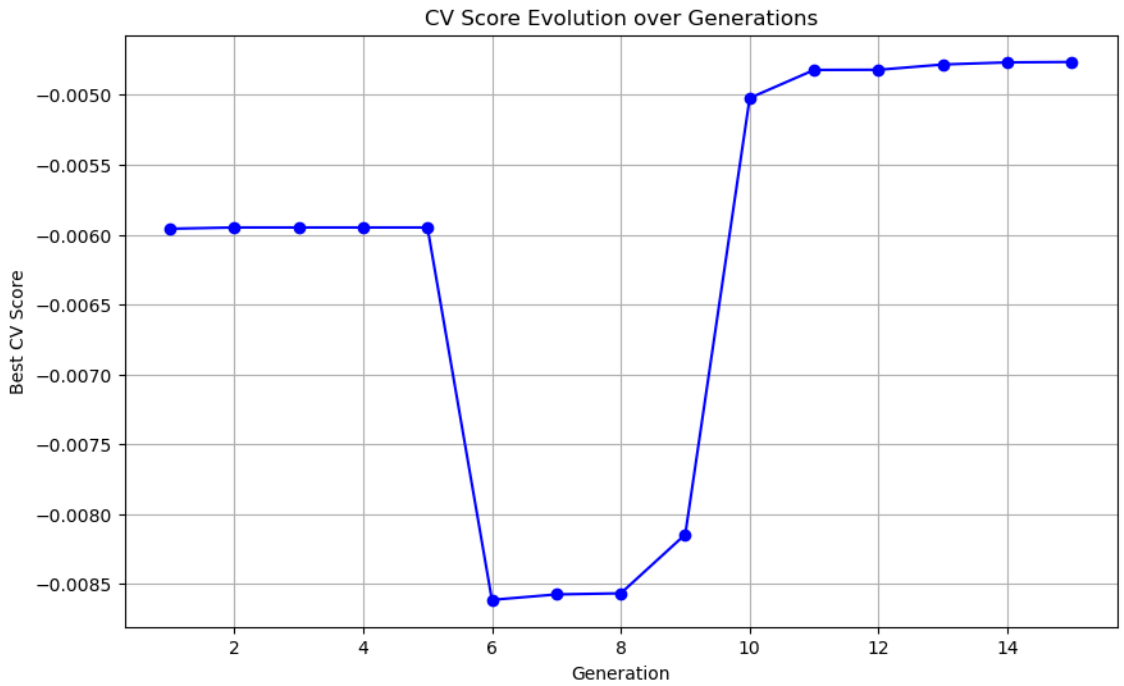

Рисунок 3.12 – Графік точності прогнозів із використанням моделі під час її навчання впродовж 15 поколінь

Із представленого графіка видно, що від 5 до 6 покоління знижується точність моделі. Водночас, після 8 покоління навчання точність починає зростати.

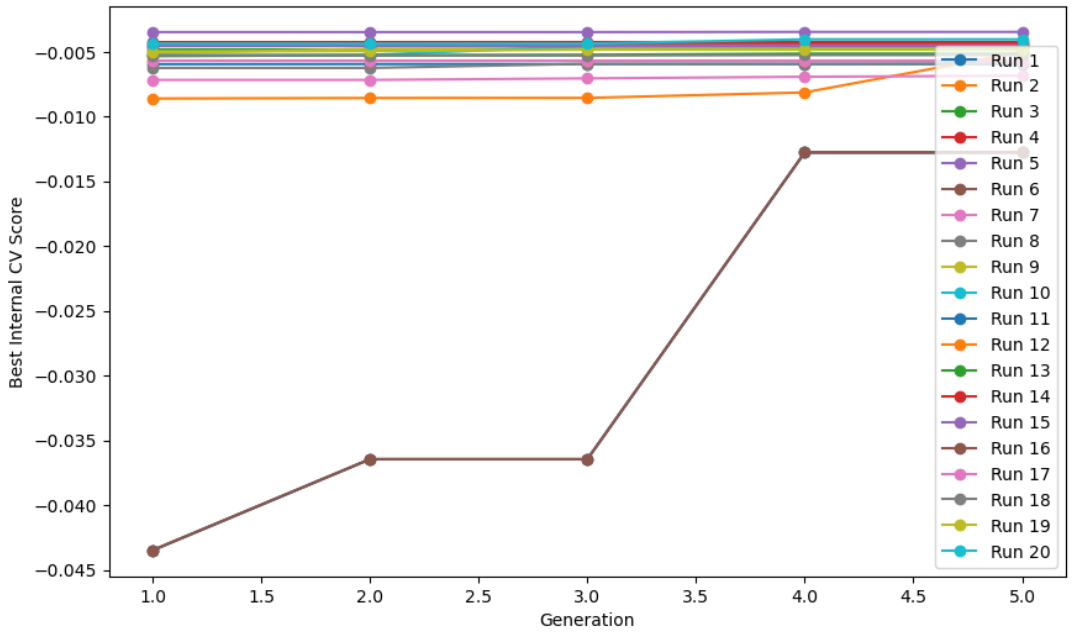

Рисунок 3.13 – Результати еволюції алгоритму оптимізації для 20 різних запусків

На графіку представлені результати еволюції алгоритму оптимізації для 20 різних запусків (Run 1, Run 2, ..., Run 20) протягом п'яти поколінь (Generation 1, Generation 2, ..., Generation 5). Окрема лінія на графіку представляє один із запусків, що відображена індивідуалізованим кольором та стилем лінії.

За результатами видно, як кращі внутрішні крос-валідаційні оцінки (Best Internal CV Score) змінюються з плином часу (поколінь). Цей графік забезпечує можливість визначення точності алгоритму оптимізації впродовж кількох поколінь та представляє динаміку отриманих результатів точності.

На підставі отриманих досліджень нами побудовано графік відображення процесу навчання (Learning Curve) (рис. 3.14).

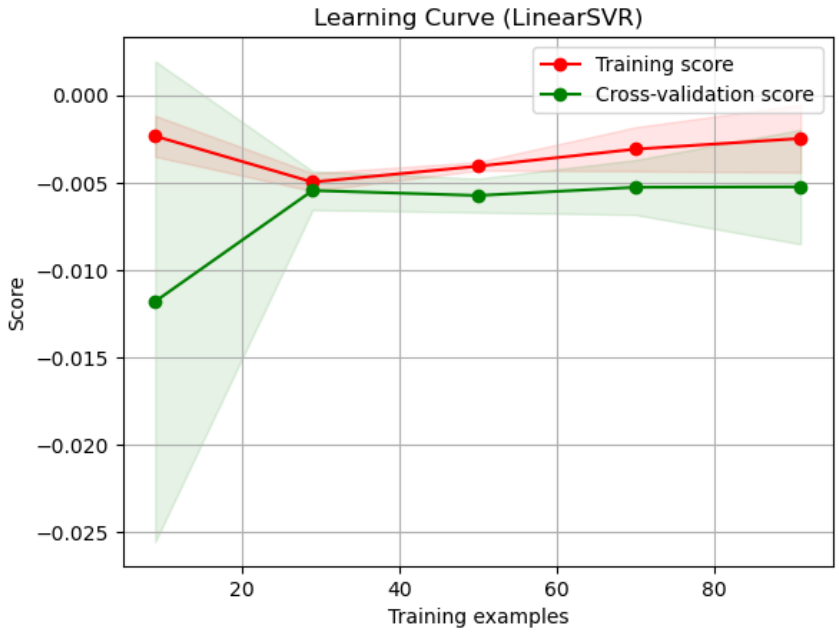

Рисунок 3.14 – Графік відображення процесу навчання (Learning Curve)

Отриманий графік процесу навчання моделі вказує на те, що запропонована модель забезпечує найкращі значення за 28 епох (прикладів) навчання. Зі зростанням прикладів навчання її точність зменшується вона поступово перенавчається. Отже, за 28 епох (прикладів) криві навчання та валідації збігаються, що свідчить про хорошу узгодженість моделі. При цьому до 22 прикладів спостерігається недонавчання, а після 40 прикладів її перенавчання.

Розглянемо точність виконання прогнозів із використанням запропонованої моделі (фактична вартість вживаних автомобілів порівняно із прогнозованою) на підставі графіка, представленого на рис. 3.15.

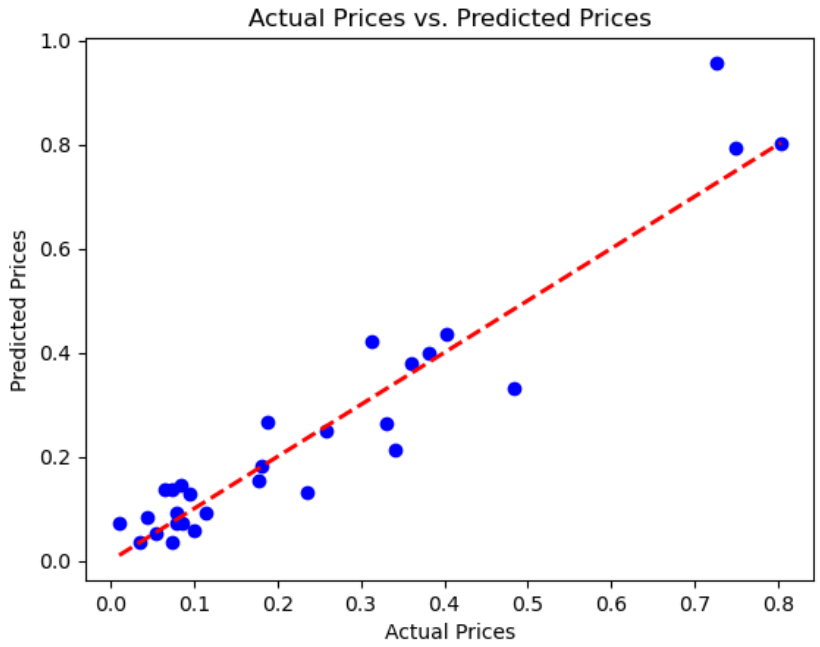

Рисунок 3.15 – Графік точності виконання прогнозів із використанням запропонованої моделі (фактична вартість вживаних автомобілів порівняно із прогнозованою)

Представлений графік точності прогнозів (фактична вартість вживаних автомобілів порівняно із прогнозованою) свідчить про те, що отримуються достатньо точні прогнози із використанням моделі, як отримані прогнозні значення незначно відрізняються від фактичних значень вартості вживаних автомобілів.

3.5. Результати порівняння запропонованої моделі із існуючими аналогами

Для порівняння запропонованої моделі прогнозування вартості вживаних автомобілів розглянемо OLS Regression (Ordinary Least Squares Regression), або

лінійна регресія методом найменших квадратів, яка навчання на цих же даних. OLS Regression це широко використовуваний метод статистичного аналізу, що застосовується для моделювання лінійних взаємозв'язків між однією залежною змінною (y) і однією або кількома незалежними змінними (x).

Результати навчання моделі прогнозування вартості вживаних автомобілів на основі алгоритму OLS Regression представлено на рис. 3.16.

OLS Regression Results

Dep. Variable:	price	R-squared:	0.899
Model:	OLS	Adj. R-squared:	0.896
Method:	Least Squares	F-statistic:	308.0
Date:	Sat, 23 Dec 2023	Prob (F-statistic):	1.04e-67
Time:	22:03:32	Log-Likelihood:	181.06
No. Observations:	143	AIC:	-352.1
Df Residuals:	138	BIC:	-337.3
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0824	0.018	-4.480	0.000	-0.119	-0.046
horsepower	0.4402	0.052	8.390	0.000	0.336	0.544
carwidth	0.3957	0.046	8.677	0.000	0.306	0.486
hatchback	-0.0414	0.013	-3.219	0.002	-0.067	-0.016
Highend	0.2794	0.022	12.591	0.000	0.236	0.323

Omnibus:	29.385	Durbin-Watson:	1.955
Prob(Omnibus):	0.000	Jarque-Bera (JB):	98.010
Skew:	0.692	Prob(JB):	5.22e-22
Kurtosis:	6.812	Cond. No.	12.9

Рисунок 3.16 – Результати навчання моделі прогнозування вартості вживаних автомобілів на основі алгоритму OLS Regression

Загалом, результати навчання моделі прогнозування вартості вживаних автомобілів на основі алгоритму OLS Regression є хорошими. Рівень детермінації (R-squared) становить 0,899, що означає, що модель пояснює 89,9% варіації ціни автомобіля. Скоригований рівень детермінації (Adjusted R-squared) становить 0,896, що означає, що модель пояснює 89,6% варіації ціни автомобіля, враховуючи кількість незалежних змінних.

Значення F-статистики (308,0) є статистично значущим (p-значення менше 0,05), що означає, що модель є статистично значущою. Це свідчить про те, що модель може генерувати прогнози, які є більш точними, ніж випадкові прогнози. Коефіцієнти регресії також є статистично значущими (p-значення менше 0,05). Це означає, що кожна незалежна змінна має значний вплив на залежну змінну.

У таблиці нижче наведено порівняння двох моделей прогнозування вартості вживаних автомобілів за різними показниками якості (табл. 3.2).

Таблиця 3.2 – Характеристика моделей прогнозування вартості вживаних автомобілів

Показники	Модель 1	Модель 2
Алгоритм	OLS Regression	Ансамблювання на основі XGBoost
R-squared	0.899	0.947
Adj. R-squared	0.896	0.945
F-statistic	308.0	494.2
Prob (F-statistic)	$1.04\ 10^{-67}$	$2.672\ 10^{-69}$
Log-Likelihood	181.06	142.19
AIC	-352.1	-276.3
BIC	-337.3	-265.4

На основі цих показників (табл. 3.2) можна вважати, що Модель 2 (Ансамблювання на основі XGBoost) виявляється кращою моделлю у порівнянні з Моделлю 1 (OLS Regression). Це підтверджується вищими значеннями коефіцієнтів детермінації, статистичною значущістю та іншими показниками якості моделі.

Для порівняння показників точності моделі на основі алгоритму OLS Regression (Модель 1) та запропонованої нами моделі на основі алгоритму ансамблювання (Модель 2) побудовано гістограми, які представлено на рис. 3.17.

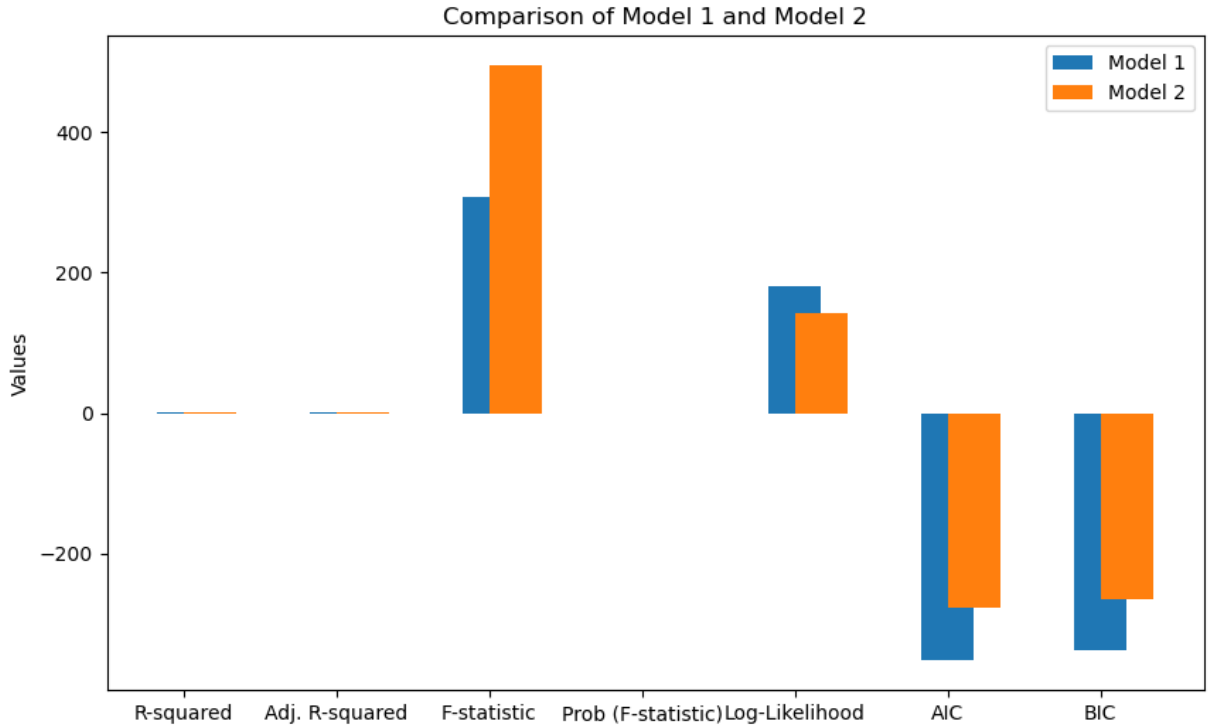

Рисунок 3.17 – Гістограми порівняння показників точності моделі на основі алгоритму OLS Regression (Модель 1) та запропонованої нами моделі на основі алгоритму ансамблювання (Модель 2) для прогнозування вартості вживаних автомобілів

За всіма показниками запропонована нами у цій роботі модель (модель 2) має кращі результати, ніж модель 1 (на основі алгоритму OLS Regression). Зокрема, модель 2 має вищий R-squared, Adj. R-squared, F-statistic, Prob (F-statistic), Log-Likelihood та нижчий AIC та BIC. Це означає, що модель 2 краще пояснює залежність між змінними, що характеризують вживані автомобілі, та вплив на їх вартість. Модель 2 також має більшу ймовірність правильного прогнозування вартості вживаних автомобілів.

РОЗДІЛ 4.
ОХОРОНА ПРАЦІ ТА БЕЗПЕКА У НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1. Аналіз шкідливих чинників під час роботи за комп'ютером

Робота за комп'ютером є одним із найпоширеніших видів праці в сучасному світі. Вона має ряд переваг, таких як комфортність, можливість дистанційної роботи тощо. Однак, робота за комп'ютером також пов'язана з певними шкідливими факторами, які можуть негативно впливати на здоров'я працівників. Характеристика небезпечних та шкідливих чинників під час роботи за компютером наведена у таблиці 4.1.

Таблиця 4.1. Характеристика небезпечних та шкідливих чинників

Чинники	Складова	Характеристика
Фізичні	Механічне навантаження	Тривале статичне положення тіла, обмеження рухливості, вібрація.
	Оптичне навантаження	Тривала робота з екраном комп'ютера, недостатня освітленість, різке світло.
	Акустичні навантаження	Підвищений шум, тривалий контакт із шумом.
	Хімічні навантаження	Шкідливі речовини, що виділяються з матеріалів меблів, оргтехніки тощо.
Психофізіологічні	Ментальне навантаження	Тривала концентрація уваги, монотонна робота, необхідність швидко реагувати на дії.
	Емоційне навантаження	Стрес, перенапруження.

Здоров’ю працівників, які працюють за комп’ютером, можуть загрожувати такі захворювання:

- ✓ опорно-рухового апарату – остеохондроз, радикуліт, тунельний синдром тощо;
- ✓ очей – короткозорість, далекозорість, астигматизм, кон’юнктивіт, синдром сухого ока тощо;
- ✓ серцево-судинної системи – гіпертонія, аритмія, інфаркт міокарда тощо;
- ✓ нервової системи – неврози, депресія, безсоння тощо.

4.2. Розробка заходів щодо покращення умов праці під час роботи за комп’ютером

Для покращення умов праці під час роботи за комп’ютером необхідно вжити таких заходів. Створювати умови для раціонального розміщення робочого місця: висота столу та стільця повинні відповідати росту працівника, відстань від очей до екрана комп’ютера повинна становити 50-70 см, рівень освітлення повинен бути достатнім, але не занадто яскравим, шум повинен бути знижений до допустимих норм.

Забезпечувати регулярно відпочинок для очей та опорно-рухового апарату: кожні 20-30 хвилин робити перерву на 5-10 хвилин, під час якої встати з-за столу, поворухнутися, подивитися вдаль.

Організовувати роботу так, щоб працівники мали можливість чергувати розумову та фізичну діяльність, уникати монотонної роботи. Створювати умови для психологічного комфорту працівників: підтримувати нормальний мікроклімат у приміщенні, забезпечити достатню вентиляцію та освітлення, уникати стресових ситуацій.

Крім того, працівникам, які працюють за комп’ютером, необхідно дотримуватися таких правил:

- ✓ регулярно проходити медичні огляди;
- ✓ виконувати вправи для очей та опорно-рухового апарату;
- ✓ зменшувати кількість часу, який проводиться за комп'ютером.

Дотримання цих заходів допоможе попередити негативний вплив шкідливих факторів праці під час роботи за комп'ютером і зберегти здоров'я працівників.

4.3. Розробка логічно-імітаційної моделі процесу виникнення травм під час роботи за комп'ютером

Для побудови логіко-імітаційної моделі процесу, формування і виникнення аварії та травми під час роботи за комп'ютером складемо список базових подій. Вони лежатимуть у основі даної моделі. Кожному пункту списку присвоюємо певне значення ймовірності виникнення. Нижче подано сам список:

1. Стан контролю з охорони праці ..... **$P_1 = 0,2$;**
2. Несерйозне відношення до проходження обслуговування комп'ютерної техніки ..... **$P_2 = 0,1$;**
3. Відсутність захисних засобів ..... **$P_3 = 0,2$;**
4. Невисока наційність ..... **$P_4 = 0,03$;**
5. Використання застарілого обладнання ..... **$P_6 = 0,02$;**
6. Пошкодження ізоляції ..... **$P_7 = 0,4$;**
7. Досвід роботи виконавця ..... **$P_{12} = 0,35$.**
8. Професійний рівень виконавця ..... **$P_{13} = 0,5$;**
9. Психофізіологічний стан виконавця ..... **$P_{14} = 0,083$;**

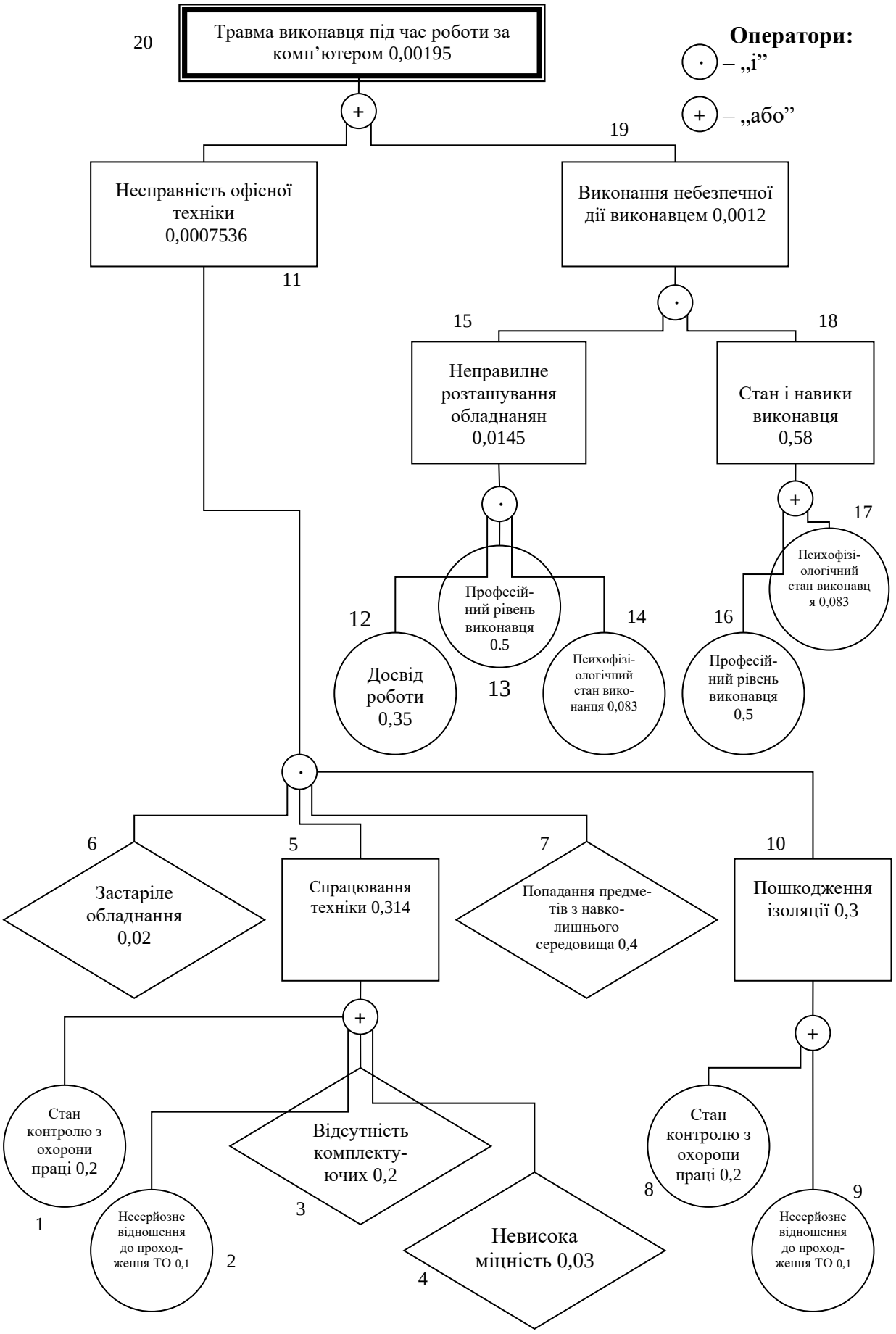

Рис. 4.1. Логіко-імітаційна модель процесу формування та виникнення аварії та травми під час роботи за комп'ютером

На основі даного списку будуємо матрицю логічних взаємозв'язків між окремими пунктами, графічне представлення якої зображено на рис. 4.1.

Розрахуємо ймовірності виникнення подій, що входять у дану логіко-імітаційну модель процесу роботи за комп'ютером (на прикладі ймовірності отримання травми виконавця).

Ймовірність виникнення події P_5 визначаємо наступним чином:

$$P_5 = 02 + 0{,}1 + 0{,}2 + 0{,}003 - 0{,}2 \cdot 0{,}1 - 0{,}2 \cdot 0{,}03 - 0{,}2 \cdot 0{,}03 - 0{,}1 \cdot 0{,}2 - 0{,}1 \cdot 0{,}03 - $$
$$- 0{,}2 \cdot 0{,}03 + 0{,}2 \cdot 0{,}1 \cdot 0{,}2 + 0{,}1 \cdot 0{,}2 \cdot 0{,}03 + 0{,}2 \cdot 0{,}1 \cdot 0{,}2 + 0{,}2 \cdot 0{,}1 \cdot 0{,}03 - $$
$$- 0{,}2 \cdot 0{,}1 \cdot 0{,}2 \cdot 0{,}03 = 0{,}314$$

Ймовірність виникнення події P_{10} визначаємо так:

$$P_{10} = 0{,}2 + 0{,}1 = 0{,}3.$$

Ймовірність виникнення події P_{11} визначаємо:

$$P_{11} = 0{,}02 \cdot 0{,}314 \cdot 0{,}4 \cdot 0{,}3 = 0{,}00075.$$

Ймовірність виникнення події P_{15} визначаємо наступним чином:

$$P_{15} = 0{,}35 \cdot 0{,}5 \cdot 0{,}083 = 0{,}0145.$$

Ймовірність події P_{18}:

$$P_{18} = 0{,}5 + 0{,}083 = 0{,}58.$$

Ймовірність події P_{19}:

$$P_{19} = 0{,}0145 \cdot 0{,}083 = 0{,}0012.$$

Ймовірність події P_{20}:

$$P_{20} = 0{,}00075 + 0{,}0012 = 0{,}00195.$$

Ймовірність травми рівна ймовірності виникнення аварії, бо остання можлива лише за умови роботи виконавця за комп'ютером.

Логічно-імплікаційна модель є ефективним інструментом для вивчення процесу виникнення травм під час роботи за комп'ютером. Вона дозволяє детально описати процес виникнення травм та оцінити ризик їх виникнення. Це може бути використано для розробки ефективних заходів щодо попередження травм.

4.4. Розробка заходів щодо безпеки у надзвичайних ситуаціях

Оцінка ризику виникнення надзвичайних ситуацій передбачає визначення імовірності та наслідків виникнення надзвичайних ситуацій. Для цього необхідно провести аналіз потенційних джерел надзвичайних ситуацій, їхніх характеристик та можливих наслідків.

Плани дій на випадок надзвичайних ситуацій є документами, які регламентують порядок дій працівників та населення у разі виникнення надзвичайних ситуацій. Плани дій на випадок надзвичайних ситуацій повинні бути розроблені для всіх об’єктів, які можуть бути задіяні у надзвичайних ситуаціях.

Матеріально-технічна база для реагування на надзвичайні ситуації включає в себе засоби індивідуального захисту, засоби пожежогасіння, засоби медичної допомоги, а також інші необхідні засоби та обладнання.

Забезпечення матеріально-технічної бази для реагування на надзвичайні ситуації здійснюється органами влади та місцевого самоврядування, а також підприємствами та організаціями.

Навчання та тренування з питань безпеки у надзвичайних ситуаціях спрямовані на підвищення рівня знань та навичок працівників та населення щодо дій у разі виникнення надзвичайних ситуацій. Навчання та тренування з питань безпеки у надзвичайних ситуаціях проводяться органами влади та місцевого самоврядування, а також підприємствами та організаціями.

Розробка заходів щодо безпеки у надзвичайних ситуаціях є важливим завданням, яке забезпечує захист життя та здоров’я людей, майна та навколишнього середовища від можливих негативних наслідків надзвичайних ситуацій.

РОЗДІЛ 5.
ВИЗНАЧЕННЯ ЕКОНОМІЧНОЇ ЕФЕКТИВНОСТІ ВІД ВИКОРИСТАННЯ РОЗРОБЛЕНОЇ МОДЕЛІ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ВАРТОСТІ ВЖИВАНИХ АВТОМОБІЛІВ

Мета даного розділу полягає у визначенні економічної ефективності від використання розробленої моделі машинного навчання для прогнозування вартості вживаних автомобілів. Для визначення економічної ефективності моделі необхідно провести порівняння доходу, який може бути отриманий від використання моделі, з витратами на її розробку та використання. Для цього необхідно:

- ✓визначити джерела доходу від використання моделі;
- ✓визначити витрати на розробку та використання моделі;
- ✓розрахувати економічну ефективність моделі.

Джерело доходу від використання запропонованої моделі може бути продаж прогнозів вартості автомобілів та зниження витрат на закупівлю автомобілів.

У нашому розрахунку розглядається перший варіант джерела доходу – продаж прогнозів вартості автомобілів.

Витрати на розробку та використання моделі включають в себе такі витрати:

$$C_{рв} = C_{зд} + C_р + C_п, \qquad (5.1)$$

де $C_{зд}$ – витрати на збирання та обробку даних, грн;

$C_р$ – витрати на розробку моделі, грн;

$C_п$ – витрати на підтримку моделі, грн.

У даному випадку розглядаються витрати на збирання та обробку даних, розробку моделі та її підтримку протягом одного року.

Економічна ефективність від використання розробленої моделі машинного навчання для прогнозування вартості вживаних автомобілів визначається за формулою:

$$E = Д - C_{рв}, \quad (5.2)$$

Де E – економічна ефективність від використання розробленої моделі машинного навчання для прогнозування вартості вживаних автомобілів, грн/рік;

$Д$ – дохід від використання моделі для прогнозування вартості вживаних автомобілів, грн/рік;

$C_{рв}$ – витрати на розробку та використання моделі для прогнозування вартості вживаних автомобілів, грн/рік.

Вважаємо, що модель буде забезпечувати дохід від прогнозу вартості автомобілів за ціною 750 гривень за один прогноз.

Якщо модель буде використовуватися для прогнозування вартості 50 автомобілів на місяць, то дохід від неї складе:

$$Д_м = 750 \cdot 50 = 37500 \text{ грн/місяць.}$$

У рік модель буде приносити дохід у розмірі:

$$Д_р = 37500 \cdot 12 = 450000 \text{ грн/рік.}$$

Приймаємо, що витрати на збирання та обробку даних складуть $C_{зд} = 50000$ грн., витрати на розробку моделі складуть $C_р = 100000$ грн., а витрати на підтримку моделі складуть $C_п$=25000 грн. на рік.

Підставивши відповідні значення у формулу (5.1) отримаємо:

$$C_{рв} = 50000 + 100000 + 25000 = 175000 \text{ грн.}$$

Виконаємо розрахунок економічної ефективності від використання моделі машинного навчання для прогнозування вартості вживаних автомобілів:

$$E = 450000 - 175000 = 275000 \text{ грн/рік.}$$

Термін окупності капіталовкладень у розроблену модель машинного навчання для прогнозування вартості вживаних автомобілів визначається за формулою:

Таблиця 5.1 – Результати визначення економічної ефективності від використання розробленої моделі машинного навчання для прогнозування вартості вживаних автомобілів

№ п/п	Назва показників	Одиниця виміру	Значення
1	Витрати на збирання та обробку даних	грн.	50000
2	Витрати на розробку моделі	грн.	100000
3	Витрати на підтримку моделі	грн.	25000
4	Витрати на розробку та використання моделі	грн.	175000
5	Річний дохід	грн.	450000
6	Економічна ефективність	грн./рік	275000
7	Термін окупності капіталовкладень	років	0,64

$$T_{ок} = \frac{C_{рв}}{E}. \tag{5.3}$$

Підставивши значення у (5.3) маємо:

$$T_{ок} = \frac{175000}{275000} = 0,64 \text{ року}.$$

У даному розділі було проведено визначення економічної ефективності від використання розробленої моделі машинного навчання для прогнозування вартості вживаних автомобілів. Для цього були визначені джерела доходу від використання моделі, витрати на розробку та використання моделі, а також розраховано економічну ефективність моделі. Результати розрахунку показали, що економічна ефективність моделі становить 275000 грн/рік, а термін окупності становитиме 0,67 року. Це означає, що використання моделі може принести значний економічний ефект.

ВИСНОВКИ І ПРОПОЗИЦІЇ

Виконана кваліфікаційна робота стосується вивчення та аналізу різних програмних платформ для автоматичного створення моделей машинного навчання та їх застосування для ефективного прогнозування вартості вживаних автомобілів. Результати цього дослідження можуть сприяти вдосконаленню методів оцінки автомобільного ринку та забезпечення якісних інформаційних рішень у сфері купівлі та продажу автотранспорту.

Нами проаналізовано особливості прогнозування вартості вживаних автомобілів. Встановлено, що визначення вартості вживаних автомобілів є складним завданням через багатофакторність впливу на ціни. Основними факторами, які впливають на ціну вживаного автомобіля, є рік випуску, пробіг, тип палива та трансмісія. Відомі підходи дають достатньо точні результати у прогнозуванні вартості вживаних автомобілів, яка становить 93,73%.

Нами виконано аналіз використовуваних технологій машинного навчання для прогнозування вартості вживаних автомобілів. Встановлено, що у існуючих підходах процес формування набору даних виявився критичним етапом. Існує потреба систематизувати та стандартизувати отримані дані для ефективного навчання моделей, що є ключовим для досягнення високої точності прогнозу.

Відомі аналогічні дослідження, де набори даних були отримані з двох різних платформ, а саме Autovit.ro та Mobile.de, які є відомими веб-сайтами, присвяченими продажу вживаних автомобілів у Румунії та Німеччині відповідно. Пропозиція на Autovit.ro показала, що існує значно більше автомобілів з автоматичною коробкою передач, приблизно на 50% ніж з ручною коробкою передач. Тип коробки передач, як правило, мав істотний вплив на ціну продажу. Однак аналогічна різниця була невеликою на німецькому ринку, і кількість пропозиції була майже однаковою між цими двома класами.

Нами проведено аналіз стану використання автоматичного створення моделей машинного навчання. Встановлено, що у усіх предметних галузях,

AutoML став новою галуззю досліджень з метою автоматичної оптимізації частин конвеєра машинного навчання, як показано на рис. 1.11. В останні роки з’явилися різні рішення AutoML для оптимізації одного або кількох із цих компонент (рис. 1.12).

Типовий конвеєр програми машинного навчання із використанням AutoML передбачає автоматизовану розробку функцій, оптимізацію гіперпараметрів, використання оптимізаторів конвеєрів (звертається на більше ніж один компонент) і пошук ефективної архітектури моделі. Сьогодні автоматизоване машинне навчання (AutoML) не просто залишається ефективним теоретичним інструментом, але й привертає більше уваги практиків.

Відповідно до цього нами сформульовано тему, мету та завдання кваліфікаційної роботи, яка стосується прогнозування вартості вживаних автомобілів із вибором програмної платформи автоматичного створення моделі машинного навчання.

Нами проаналізовано процес автоматизованого машинного навчання (AutoML) та встановлено його переваги. На даний час з’явилося багато чудових фреймворків AutoML. Нами наведено сім інструментів, які спрощують використання алгоритмів машинного навчання. Кожен із них має свої переваги та недоліки.

У нашій роботі для прогнозування вартості вживаних автомобілів пропонується використовувати інструмент оптимізації конвеєрів на основі дерева TPOT. Щоб використовувати TPOT для регресії, необхідно створити новий проект TPOT і вибрати задачу регресії. Потім можна надати TPOT дані для навчання та тестування. TPOT буде використовувати ці дані для пошуку найкращої архітектури моделі та параметрів.

TPOTRegressor виконує інтелектуальний пошук у конвеєрах машинного навчання, які можуть містити контрольовані регресійні моделі, препроцесори, методи вибору функцій і будь-які інші оцінювачі чи трансформатори, які належать API scikit-learn . TPORegressor також шукатиме гіперпараметри всіх

об’єктів у конвеєрі. TPOTRegressor використовує проаналізовані нами методи для пошуку найкращої архітектури моделі та параметрів. Ці методи допомагають TPOTRegressor знаходити ефективні архітектури моделей та параметри для різних задач машинного навчання.

Для прогнозування вартості вживаних автомобілів нами обрано доступний набір даних із платформи Kaggle. Дані, що описують ціну доступних вживаних автомобілів на ринку із їх характеристиками, що є основою для виявлення взаємозв’язків між чинниками, що впливають на ціну автомобілів. Метою наших досліджень було підготувати дані та на підставі них виконати автоматичного створення моделі машинного навчання із використанням TPOTRegressor. Це забезпечує інтелектуальний пошук у конвеєрах машинного навчання, ефективної регресійної моделі.

Набір даних містить 26 змінних, які характеризують вживані автомобілі, та 1 змінну, яка вказує на вартість автомобіля. Для вирішення задачі підготовки та аналізу даних про вартість вживаних автомобіля було використано Jupyter Notebook. Jupyter Notebook – це відкритий проєкт, який забезпечує розробку програмного забезпечення, що дозволяє проводити інтерактивні обчислення на різних мовах програмування.

Нами проведено підготовку даних до машинного навчання. Встановлено, що у стовпці «CompanyName» допущено помилки у їх описі. Зокрема, зустрічаються різні представлення даних. Виправлення невірних значень виконуємо із використанням коду (рис. 3.5), що забезпечує очищення та стандартизацію марки автомобіля у стовпці «CompanyName» датафрейму df.

Нами виконано аналіз даних для прогнозування вартості вживаних автомобілів. Побудовано два графіки, які характеризують розподіл цін на вживані автомобілі – гістограми та діаграми розмаху (рис. 3.6). Встановлено, що більшість автомобілів у наборі даних мають ціну нижчою 15000$. Існує значна різниця між середнім значенням та медіаною розподілу вартості вживаних автомобілі.

На наступному етапі проводимо аналіз категорійних даних, які впливають на вартість вживаних автомобілів (рис. 3.7). Встановлено, що найбільше даних є про автомобілі «Тойота». Кількість автомобілів на бензиновому паливі більша, ніж на дизельному. Седан є найпопулярнішим типом автомобіля.

Також нами проаналізовано вплив інших чинників на вартість вживаних автомобілів. Результати представлено на рис. 3.8. Встановлено, що дуже мало даних для категорій розташування двигунів, щоб зробити висновок. Найпоширеніша кількість циліндрів – чотири, шість та п'ять. Хоча вісім циліндрів мають найвищий ціновий діапазон. Найпоширенішими типами паливних систем є mpfi та 2bbl. mpfi та idi мають найвищий ціновий діапазон.

Нами створено матрицю парних графіків для візуалізації взаємозв'язків між усіма числовими змінними у датафрейму, що характеризує вартість вживаних автомобілів (рис. 3.9). Діагональ матриці містить гістограми розподілу кожної змінної. Верхній трикутник (крім діагоналі) містить діаграми розсіювання, що показують зв'язок між парами змінних. Матриця дозволяє швидко оцінити наявність сильних або слабких кореляцій між змінними, а також виявити потенційні нелінійні зв'язки.

Написаний код (рис. 3.10) використовує бібліотеку TPOT для автоматизованої оптимізації процесу побудови моделі регресії для прогнозування вартості вживаних автомобілів. Розділення даних на навчальний та тестовий набори виконано із співвідношенням 80/20. Використовується TPORegressor для побудови та оптимізації моделей. Генерація та розмір населення забезпечують кількість поколінь та розмір кожного покоління в оптимізаційному процесі. Параметр verbosity встановлює рівень виведення інформації під час оптимізації.

На підставі машинного навчання із використанням бібліотеки TPOT для автоматизованої процесу побудови моделі прогнозування вартості вживаних автомобілів (рис. 3.10) нами отримано характеристики цієї моделі (табл. 3.1). Встановлено, що найкращою виявилася для наших даних модель ElasticNetCV, яка використовує алгоритм генетичного програмування для автоматичного

пошуку найкращої моделі машинного навчання для нашої задачі. Найкращий пайплайн, який знайшов ElasticNetCV, включає в себе GradientBoostingRegressor, який є моделлю ансамблю, що складається з декількох простих моделей. У подальшому використано алгоритм ансамблювання, тобто не одну модель XGBoost. Це забезпечило можливість підібрати найкращі базові алгоритми стекінгу.

У результаті виконаних досліджень нами здійснено побудову графіка поколінь та визначення їх кращих показників CV у порівняння з реальними значеннями (рис. 3.12). У цьому прикладі використовуються реальні значення для поколінь від 1 до 15. Із представленого графіка видно, що від 5 до 6 покоління знижується точність моделі. Водночас, після 8 покоління навчання точність починає зростати. Нами подано результати еволюції алгоритму оптимізації для 20 різних запусків протягом п'яти поколінь. За результатами видно, як кращі внутрішні крос-валідаційні оцінки (Best Internal CV Score) змінюються з плином часу (поколінь).

Отриманий графік процесу навчання моделі вказує на те, що запропонована модель забезпечує найкращі значення за 28 епох (прикладів) навчання. Зі зростанням прикладів навчання її точність зменшується вона поступово перенавчається. Отже, за 28 епох (прикладів) криві навчання та валідації збігаються, що свідчить про хорошу узгодженість моделі. При цьому до 22 прикладів спостерігається недонавчання, а після 40 прикладів її перенавчання. Представлений графік точності прогнозів свідчить про те, що запропонована модель дає достатньо точні прогнози щодо вартості вживаних автомобілів.

На основі проведених досліджень (табл. 3.2) можна вважати, що запропонована нами модель (Модель 2 – Ансамблювання на основі XGBoost) виявилася кращою моделлю у порівнянні з Моделлю 1 (OLS Regression). Це підтверджується вищими значеннями коефіцієнтів детермінації, статистичною значущістю та іншими показниками якості моделі.

Розроблені заходи щодо охорони праці та безпеки у надзвичайних ситуаціях під час роботи за комп’ютером лежать в основі створення безпечних умов праці.

Результати розрахунку показали, що економічна ефективність від використання запропонованої моделі становить 275000 грн/рік, а термін окупності капіталовкладень становитиме 0,67 року. Це означає, що використання моделі може принести значний економічний ефект.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Введення в машинне навчання за допомогою Python и Scikit-Learn. URL: https://habr.com/ua/company/mlclass/blog/247751/ (дата звернення: 26.08.2023).

2. Жидецький В.Ц., Джигирей В.С., Мельников О.В. Основи охорони праці. Підручник. Вид. 5-е, доповнене. Львів: Афіша, 2012. 350с.

3. Класифікація в Python з Scikit-Learn та Pandas. URL: https://stackabuse.com/classification-in-pythonwith-scikit-learn-and-pandas/ (дата звернення: 21.07.2023).

4. Tryhuba A., Tryhuba I., Bashynsky O., Kondysiuk I., Koval N., Bondarchuk L., Conceptual Model of Management of Technologically Integrated Industry Development Projects. IEEE 15th International Conference on Computer Sciences and Information Technologies, CSIT-2020, IEEE, Lviv, 2020, pp. 155–158. doi: 10.1109/CSIT49958.2020.9321903.

5. Koval N., Tryhuba A., Kondysiuk I., Tryhuba I., Boiarchuk O., Rudynets M., Grabovets V., Onyshchuk V. Forecasting the fund of time for performance of works in hybrid projects using machine training technologies. 3rd International Workshop on Modern Machine Learning Technologies and Data Science Workshop, MoMLeT and DS 2021, CEUR Workshop Proceedings 2917, Lviv-Shatsk, 2021, pp. 196–206.

6. Kondysiuk I., Tryhuba A., Bashynsky O., Grabovets V., Dembitskyi, V., Myskovets, I. Formation and risk assessment of stakeholders value of motor transport enterprises development projects. IEEE 16th International Conference on Computer Sciences and Information Technologies, CSIT-2021, IEEE, Lviv, 2021, pp. 307–310. doi: 10.1109/CSIT52700.2021.9648739.

7. Tryhuba A., Kondysiuk I., Tryhuba I., Koval N., Boiarchuk O., Tatomyr A. Intellectual information system for formation of portfolio projects of motor transport enterprises, in: I Workshop Information Technologies in Energy and Agro-

industrial Complex, ITEA-WS 2021, CEUR Workshop Proceedings 3109, Dubliany, Lviv region, 2021, pp. 44–52.

8. Tryhuba, A., Kondysiuk I., Tryhuba I., Lub P. Approach and Software for Risk Assessment of Stakeholders of Hybrid Projects of Transport Enterprise. CEUR Workshop Proceedingsthis link is disabled, 2022, 3295, pp. 86–96.

9. Тригуба А., Тригуба І., Фтома О., Кондисюк І., Коваль Н. Системний підхід до оцінення ризиків несвоєчасного виконання робіт в інтегрованих проєктах. Вісник Львівського національного аграрного університету: агроінженерні дослідження. №23. Львів: Львів НАУ, 2019. С. 123-130. (0,84 д. а.).

10. Тригуба А., Тригуба І., Чубик Р., Кондисюк І., Коваль Н., Панюра Я. Прогнозування обсягів заготівлі сировини на території громад із використанням штучних нейронних мереж. Вісник Львівського національного аграрного університету: агроінженерні дослідження. №24. Львів: Львів НАУ, 2020. С.143-151.

11. Тригуба А., Кондисюк І., Коваль Н. Формування портфелів гібридних проєктів автотранспортних підприємств. Вісник Національного технічного університету "ХПІ". Сер. : Стратегічне управління, управління портфелями, програмами та проєктами : зб. наук. праць – Харків : НТУ "ХПІ", 2021. – № 2 (4). – С. 67-72.

12. Тригуба А., Кондисюк І., Коваль Н., Тригуба І., Боярчук Ок., Боярчук Ол. Планування часу виконання робіт у гібридних проєктах. Вісник Національного технічного університету "ХПІ". Сер. : Стратегічне управління, управління портфелями, програмами та проєктами : зб. наук. праць Харків : НТУ "ХПІ", 2022. № 2 (6). С. 64-71.

13. Tryhuba A., Sholudko Y., Kondysiuk I. Justification of the configuration of the logistic delivery system of perishable agricultural products. 2nd International Conference on Agriculture, Technology, Engineering and Sciences (ICATES 2019), 2019. Lviv. P. 144.

14. Тригуба А.М., Кондисюк І., Коваль Н. Алгоритм прийняття управлінських рішень в умовах невизначеності із використанням машинного навчання. Вчені Львівського національного аграрного університету виробництву: каталог інноваційних розробок за заг. ред. В. В. Снітинського, І. Б. Яціва. Вип. 20. Львів: Львів. нац. аграр. ун-т, 2020. С. 39.

15. Тригуба А.М., Ратушний Р.Т., Кондисюк І., Коваль Н. Рівні та особливості моделювання гібридних проєктів розвитку територіальних систем. Управління проєктами: стан та перспективи: матеріали XVI Міжнар. конф. – Миколаїв: НУК, 2020. С. 74-75.

16. Тригуба А.М., Кондисюк І.В. Метод формування портфелів гібридних проєктів автотранспортних підприємств. Вчені Львівського національного аграрного університету виробництву: каталог інноваційних розробок за заг. ред. В. В. Снітинського, І. Б. Яціва. Вип. 21. Львів: Львів. нац. аграр. ун-т, 2021. С. 52.

17. Тригуба А.М., Тригуба І.Л., Кондисюк І.В., Коваль Н.Я. Планування змісту та часу виконання робіт у гібридних проєктах із використанням штучних нейронних мереж. Тези доп. XVII-ї Міжн. конф. Управління проєктами у розвитку суспільства: Управління проєктами в умовах пандемії COVID-19». Київ: КНУБА, 2021. С.279-284.

18. Коваль Н.Я., Кондисюк І.В., Тригуба А.М. Алгоритм навчання нейронної мережі для планування часу виконання робіт у гібридних проєктах. Молодь у світі сучасних технологій за тематикою: Сучасні інформаційні технології: стан та перспективи розвитку : матеріали міжнар. наук.-практ. конф. (4 червня 2021 р., м. Херсон) / за заг. ред. Г.О. Райко. – Херсон: Видавництво ФОП Вишемирський В. С., 2021. – С. 153-156.

19. Тригуба А., Пташник В., Татомир А., Коваль Н.Я., Кондисюк І.В. Використання штучних нейронних мереж для прогнозування складових гібридних проєктів. Теорія і практика розвитку агропромислового комплексу та сільських територій: матеріали XXII Міжнародного науково-парктичного форуму, 5-7 жовтня 2021р.: у 2 т. Львів: ННВК «АТБ», 2021. Т.2. С. 96-100.

20. Тригуба А.М., Кондисюк І.В., Татомир А.В., Шолудько Я.В., Боярчук О.В. Інтелектуальна інформаційна система формування портфелів проєктів автотранспортних підприємств. Інформаційні технології в енергетиці та агропромисловому комплексі: матеріали X-ї міжнародної наукової конференції, присвяченої 165-річчю університету. Львів-Дубляни, 2021, С. 113–115.

21. Тригуба А.М., Кондисюк І.В. Алгоритм оптимізації портфелів гібридних проєктів автотранспортних підприємств. Вчені Львівського національного аграрного університету виробництву: каталог інноваційних розробок за заг. ред. В. В. Снітинського, І. Б. Яціва. Вип. 22. Львів: Львів. нац. аграр. ун-т, 2022. С. 26.

22. Тригуба А.М., Тригуба І.Л., Коваль Н.Я., Кондисюк І.В. Використання моделі SARIMA для прогнозування проектного середовища гібридних проектів заготівлі молока на території громад. Тези доп. XIX-й Міжнародній науково-практичній конференції «Управління проектами у розвитку суспільства». Київ: КНУБА, 2022. С.279-284.

23. Лехман С.Д., Рублев В.І., Рябцев Б.І. Запобігання аварійності і травматизму у сільському господарстві. К.: Урожай, 1993. 267 с.

24. Новоселецький О.М., Якубець О.В. Моделювання кредитоспроможності юридичних осіб на основі дискримінантного аналізу та нейронних мереж. Нейронечіткі технології моделювання в економіці. Науково-аналітични журнал. Київ, 2014. №3. С.120-151.

25. Огляд методів класифікації у машинному навчанні за допомогою Scikit-Learn. URL: https://tproger. ru/translations/scikit-learn-in-python/s://stackabuse.com/classification-in-python-with-scikit-learn-and-pandas/ (дата звернення: 15.09.2023).

26. Productionizing Distributed XGBoost to Train Deep Tree Models with Large Data Sets at. Uber URL: https://www.uber.com/en-UA/blog/productionizing-distributed-xgboost/ (дата звернення: 09.11.2023).

27. Lessmann S. , Voß S. . Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy. Int. J. Forecast., 33 (4) (2017), pp. 864-877.

28. Noor K. , Jan S. Vehicle Price Prediction System using Machine Learning Techniques. Int. J. Comput. Appl., 167 (9) (2017)

29. Chen C. , Hao L. , Xu C. Comparative analysis of used car price evaluation models. AIP Conf. Proc., 1839 (2017), Article 020165.

30. Monburinon N., Chertchom P., Kaewkiriya T., Rungpheung S., Buya S. and Boonpou P. Prediction of prices for used car by using regression models, 5th International Conference on Business and Industrial Research (ICBIR), Bangkok, 2018, pp. 115-119.

31. Pal N. , Arora P. , Palakurthy S.S. , Sundraraman D. , Kohli P. How much is my car worth? A methodology for predicting used cars prices using Random Forest. Future of Information and Communications Conference (FICC) (2018)

32. Gegic E. , Isakovic B. , Keco D. , Masetic Z. , Kevric J. Car Price Prediction using Machine Learning Techniques. TEM J., 8 (1) (2019), pp. 113-118

33. Chandak A. , Ganorkar P. , Sharma S. , Bagmar A. , Tiwari S. Car Price Prediction Using Machine Learning. Int. J. Comput. Sci. Eng., 7 (5) (2019)

34. Celika O. , Omer Osmanoglu U. Prediction of the Prices of Second-Hand Cars. Eur. J. Sci. Technol., 16 (2019), pp. 77-83.

35. Jayadera T.P., Jayamanne D.J. Fair Price Prediction System for Used Cars in Sri Lanka Using Machine Learning and Robotic Process Automation, International Conference On Business Innovation (ICOBI), Colombo, 2019.

36. Venkatasubbu P. , Ganesh M. Used Cars Price Prediction using Supervised Learning Techniques. Int. J. Eng. Adv. Technol. (IJEAT), 9 (1S3) (2019)

37. Laveena D'Costa A. , Wilson D'Souza, Abhijith K. , Maria Varghese D. Predicting True Value of Used Car using Multiple Linear Regression Model. Int. J. Recent Technol. Eng. (IJRTE), 8 (5S) (2020)

38. Viswapriya E. , Sharma D.S.S. , Sathya Kiran G. Vehicle Price Prediction using SVM Techniques. Int. J. Innov. Technol. Explor. Eng., 9 (8) (2020)

39. Kiran S. Prediction of Resale Value of the Car Using Linear Regression Algorithm. Int. J. Innov. Sci. Res. Technol., 5 (7) (2020)

40. Samruddhi K. , Ashok Kumar R. Used Car Price Prediction using K-Nearest Neighbor Based Model. Int. J. Innov. Res. Appl. Sci. Eng., 4 (3) (2020)

41. Nair A. , Ramkumar P. , Mahadevan S. , Prakash C. , Dixit S. , Murali G. , Vatin N.I. , Epifantsev K. , Kumar K. Machine Learning for Prediction of Heat Pipe Effectiveness. Energies, 15 (9) (2022), p. 3276.

42. Arora P., Gupta H. , Singh A. Forecasting resale value of the car: Evaluating the proficiency under the impact of machine learning model. Volume 69, Part 2, 2022, Pages 441-445.

43. Dutulescu A. Car Price Quotes Driven by Data-Comprehensive Predictions Grounded in Deep Learning Techniques. *Electronics* 2023, *12*(14), 3083.

44. Pycaret. URL: https://pycaret.org/ (дата звернення: 15.11.2023).

45. Auto-SKLearn. URL: https://www.automl.org/automl/auto-sklearn/ (дата звернення: 16.11.2023).

46. MLBox. URL: https://pypi.org/project/mlbox/ (дата звернення: 20.11.2023).

47. TPOT. URL: https://wiki.pathmind.com/automl-automated-machine-learning-ai (дата звернення: 21.11.2023).

48. H2O. URL: https://www.h2o.ai/products/h2o-automl/ (дата звернення: 23.11.2023).

49. Auto-Keras. URL: https://autokeras.com/ (дата звернення: 25.11.2023).

50. DataRobot. URL: https://www.datarobot.com/ (дата звернення: 26.11.2023).

51. Набір даних із платформи Kaggle про вартість вживаних автомобілів. URL: https://www.kaggle.com/datasets/goyalshalini93/car-data?resource=download