

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»**

**Інститут прикладного системного аналізу  
Кафедра математичних методів системного аналізу**

До захисту допущено  
Завідувач кафедри  
\_\_\_\_\_ Оксана ТИМОЩУК  
«\_\_» \_\_\_\_\_ 20\_\_ р.

**Дипломна робота**

**на здобуття ступеня бакалавра  
за освітньо-професійною програмою «Системний аналіз і управління»  
спеціальності 124 «Системний аналіз»  
на тему: «Система для аналізу та прогнозування вартості оренди  
торговельних приміщень»**

Виконав:

Студент ІV курсу, групи КА-71  
Возняк Володимир Зіновійович \_\_\_\_\_

Керівник:

професор, д.т.н., Бідюк П.І. \_\_\_\_\_

Консультант з економічного розділу:

доцент, к.е.н., Рощина Н.В. \_\_\_\_\_

Консультант з нормоконтролю:

доцент, к.т.н., Коваленко А.Є. \_\_\_\_\_

Рецензент: професор, д.т.н., Архипов О.Є. \_\_\_\_\_

Засвідчую, що у цій дипломній роботі  
немає запозичень з праць інших авторів  
без відповідних посилань.

Студент \_\_\_\_\_

**Національний технічний університет України**  
**«Київський політехнічний інститут імені Ігоря Сікорського»**  
**Інститут прикладного системного аналізу**  
**Кафедра математичних методів системного аналізу**

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 124 «Системний аналіз»

Освітньо-професійна програма «Системний аналіз і управління»

ЗАТВЕРДЖУЮ

Завідувач кафедри

\_\_\_\_\_ Оксана ТИМОЩУК

«\_\_» \_\_\_\_\_ 20\_\_ р.

### ЗАВДАННЯ

на дипломну роботу студенту

**Возняка Володимира Зіновійовича**

1. Тема роботи «Система для аналізу та прогнозування вартості оренди торговельних приміщень», керівник роботи Бідюк Петро Іванович, професор, докт. технічних наук, затверджені наказом по університету від «26» травня 2021р. №1344-с

2. Термін подання студентом роботи 7 червня 2021 р.

3. Вихідні дані до роботи: модель дискретного вибору, модель ARIMA

4. Зміст роботи: дослідження та особливості предметної області торгівельних приміщень, процес аналізу та прогнозування цін на оренду

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо): презентація

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Економічний	Рощина Н.В., доцент		

7. Дата видачі завдання: 15.03.2021 \_\_\_\_\_

## Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Затвердження теми ДР	15.03.2021-21.03.2021	виконано
2	Ознайомлення зі структурою БДР згідно з Положенням про державну атестацію студентів НТУУ «КПІ ім. І. Сікорського»	15.03.2021-28.03.2021	виконано
3	Ознайомлення з ДСТУ 3008-95 та стандарти ЄСПД	22.03.2021-28.03.2021	виконано
4	Проведення дослідження за темою БДР під керівництвом керівника	29.03.2021-18.04.2021	виконано
5	Завершення роботи над першим варіантом частини БДР	19.04.2021-02.05.2021	виконано
6	Проведення роботи над експериментальною частиною БДР	03.05.2021-16.05.2021	виконано
7	Проведення роботи над програмним продуктом	17.05.2021-26.05.2021	виконано
8	Оформлення БДР та аналіз отриманих результатів	17.05.2021-26.05.2021	виконано

Студент

В.З. Возняк

Керівник

П.І. Бідюк

## РЕФЕРАТ

Дипломна робота: 114 с., 8 табл., 41 рис., 2 додатки, 34 джерела.

МОДЕЛЬ ДИСКРЕТНОГО ВИБОРУ, ЛОГІСТИЧНА РЕГРЕСІЯ, ПРОГНОЗУВАННЯ, МОДЕЛЬ ARIMA, ЧАСОВІ РЯДИ.

Об'єкт дослідження: ціни на оренду торговельних приміщень.

Мета дослідження: зробити аналіз цін на оренду реальних торговельних приміщень, розробити власні моделі для прогнозування попиту на приміщення та ціни на оренду у зручному для користувача вигляді.

Використані моделі: у програмній реалізації було використано модель дискретного вибору (логістичну регресію) та модель ARIMA.

Отриманні результати: побудовано модель для прогнозування попиту на приміщення із точністю 93,5%; побудовано модель для прогнозування ціни на оренду із MAPE 3,34%.

У рамках подальшого дослідження пропонується підвищувати точність моделей, протестувати поведінку моделей на даних із інших ТРЦ, зробити моделі універсальними для будь-яких вхідних даних.

## ABSTRACT

Thesis: 114 p., 8 tabl., 41 fig., 2 appendices, 34 sources.

DISCRETE CHOICE MODELS, LOGISTIC REGRESSION, PREDICTION, MODEL ARIMA, TIME SERIES.

Object of study: prices of retail space rent value.

Purpose: to make the analysis of prices for real retail space rent value, to develop own models for forecasting demand for retail spaces and prices for rent in a user-friendly form.

Used models: in the software implementation discrete choice model (Logistic Regression) and model ARIMA were used.

Results: Forecasting the demand for retail spaces model was built with an accuracy of 93,5%; forecasting the rental price model was built with a MAPE of 3,34%.

As part of further research, it is proposed to increase the accuracy of models, test the behavior of models on data from other malls, to make models universal for any input data.

## ЗМІСТ

ВСТУП .....	8
РОЗДІЛ 1 ТЕОРЕТИЧНІ АСПЕКТИ ВСТАНОВЛЕННЯ ОРЕНДНОЇ ПЛАТИ НА РИНКУ КОМЕРЦІЙНОЇ НЕРУХОМОСТІ .....	10
1.1 Дослідження категорії нерухомості та її сучасна класифікація.....	10
1.2 Аналіз економічних відносин на ринку комерційної нерухомості.....	15
1.3 Дослідження питання розрахунку орендної плати в Україні.....	21
Висновки .....	24
РОЗДІЛ 2 ПОБУДОВА МОДЕЛЕЙ ДЛЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ВАРТОСТІ ОРЕНДИ ТОРГОВЕЛЬНИХ ПРИМІЩЕНЬ.....	26
2.1 Побудова моделі дискретного вибору .....	26
2.2 Побудова моделі авторегресії.....	39
Висновки .....	44
РОЗДІЛ 3 ЗАСТОСУВАННЯ МОДЕЛЕЙ НА РЕАЛЬНИХ ДАНИХ ТРЦ “ВЕЛЕС” ДЛЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ЦІН НА ОРЕНДУ.....	46
3.1 Опис використовуваних даних ТРЦ “Велес” .....	46
3.2 Застосування моделі дискретного вибору для аналізу цін на оренду й прогнозування попиту приміщень .....	46
3.3 Застосування моделі ARIMA для прогнозування цін на оренду .....	61
3.4 Опис програмного продукту.....	69
Висновки .....	70
РОЗДІЛ 4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ .....	71
4.1 Постановка завдання.....	71
4.2 Обґрунтування функцій програмного продукту.....	72
4.3 Обґрунтування системи параметрів ПП .....	75
4.4 Аналіз експертного оцінювання параметрів .....	77
4.5 Аналіз рівня якості варіантів реалізації функцій.....	81
4.6 Економічний аналіз варіантів розробки ПП.....	82

4.7 Вибір кращого варіанту ПП техніко-економічного рівня .....	87
Висновки .....	88
ВИСНОВКИ.....	89
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	90
ДОДАТОК А ЛІСТИНГ ПРОГРАМНОГО ПРОДУКТУ.....	93
ДОДАТОК Б ПРЕЗЕНТАЦІЯ .....	105

## ВСТУП

Актуальність обраної теми. Ринок комерційної нерухомості України переживає не найкращі часи, проте за статистикою 2020 року попит на оренду торговельних та офісних приміщень за останні роки істотно активувався. Для отримання стабільного прибутку від здачі приміщень в оренду орендодавцю необхідно правильно оцінювати власні об'єкти, фактори, що впливають на формування орендної плати на них та ситуацію на ринку комерційної нерухомості. Саме тому проводиться аналіз ринку в цілому і окремих об'єктів оренди. На жаль, часто оцінка вартості оренди виконується не на найвищому рівні, не враховуються впливові ціноутворюючі фактори, тощо. Тому в даній роботі ми плануємо побудувати такі моделі оцінки вартості оренди, в яких по максимуму будуть враховані фактори, що є найбільш вагомими. Такі моделі дозволять орендодавцю коректно оцінити потенційний дохід від здачі приміщень в оренду, та виокремити переваги від покращення деяких характеристик торговельних та офісних приміщень.

Метою дослідження є аналіз методик формування орендної плати комерційної нерухомості в Україні та моделювання оцінки вартості оренди торговельної та офісної нерухомості на прикладі ТРЦ “Велес”, що знаходиться в Івано-Франківську, створення програмної реалізації обчислення вартості оренди комерційних приміщень торгових центрів, враховуючи особливості побудованої моделі.

Для виконання сформульованої мети дослідження, нами були поставлені такі завдання:

1. Розглянути поняття нерухомості, ринку комерційної нерухомості та орендної плати.
2. Проаналізувати методики розрахунку орендної плати в Україні.
3. Розглянути теоретичні відомості про модель дискретного вибору та модель ARIMA.

4. Побудувати модель дискретного вибору для прогнозування попиту на приміщення та модель ARIMA для прогнозування ціни на реальних даних ТРЦ «Велес».

5. Створити програмний продукт в середовищі Jupyter Notebook мовою програмування Python задля реалізації моделей.

Об'єктом дослідження є діяльність ТРЦ «Велес» як орендодавця на ринку комерційної нерухомості Івано-Франківської області.

Предметом дослідження є моделі оцінювання вартості оренди торговельної нерухомості та їх програмна реалізація.

Щодо джерельної бази досліджень, дана робота ґрунтується на аналізі законодавства України, науково-методичної літератури, методичних посібників, наукових статей, періодичних видань та напрацювань вітчизняних та іноземних вчених і дослідників, а також практичних розробок за темою комерційної нерухомості та оцінюванні вартості оренди.

Наукова новизна полягає у побудові модифікованої моделі, яка враховуватиме специфіку діяльності підприємства, що здає в оренду приміщення торговельної та офісної нерухомості, а практична цінність полягає у розробці програми розрахунку вартості оренди на основі побудованої моделі для торгових та офісних центрів на прикладі ТРЦ «Велес».

## РОЗДІЛ 1 ТЕОРЕТИЧНІ АСПЕКТИ ВСТАНОВЛЕННЯ ОРЕНДНОЇ ПЛАТИ НА РИНКУ КОМЕРЦІЙНОЇ НЕРУХОМОСТІ

### 1.1 Дослідження категорії нерухомості та її сучасна класифікація

З правової точки зору нерухомість – це об’єкти права власності господарюючих суб’єктів і громадян, відносно яких здійснюються правочини, такі як: купівля-продаж, оренда, передача прав на володіння і користування припиненням цього права.

З економічної точки зору ринок нерухомості є категорією, котра пов’язана зі створенням умов, необхідних для здійснення виробничої, комерційної, соціальної, екологічної та іншої діяльності. Будь-які об’єкти нерухомості призначенні для обслуговування цілком конкретних потреб, так ринок житла – для забезпечення житлового сектору, ринок промислових об’єктів призначений для забезпечення виробничо-технологічних процесів, ринок земельних ділянок – для забезпечення сільськогосподарського виробництва, промислово-цивільного будівництва, рекреаційної діяльності.

Розвиток теоретичних засад дослідження проблем визначення поняття «нерухомість» і об’єктів нерухомості пов’язаний з іменами таких вчених, як Асаул А.М., Брижань І.А., Шаркова І.М., Григор’єв В.В., Коваленко М.А., Кіхая Ю.В., Радванська Л.М., які в рамках більш глибокого дослідження ринку нерухомості, розкривають і сутність цього поняття. Гриценко Є.А. розглядає концептуальні основи виникнення і функціонування ринку нерухомості. Шаркова І.М. розкриває зміст понять нерухомого майна, ринку нерухомості в Україні. Асаул А.М., Брижань І.А. розглядають сутнісні характеристики об’єктів нерухомості, виводять сутність цих об’єктів з поняття основні засоби, характеризують склад об’єктів нерухомого майна. Частково визначення об’єктів нерухомості торкалися Коваленко М.А., Радванська І.М.

За ЦК України [34] до нерухомості (нерухомих речей, нерухомого майна) належать земельні ділянки, а також об’єкти, розташовані на земельній ділянці, переміщення яких є неможливим без їхнього знецінення та зміни їхнього при-

значення. З наведеного визначення випливає матеріальний критерій зарахування об'єкту до нерухомості, яким є ступінь зв'язку речі з землею. За таким критерієм О. Дзера відносить до нерухомості земельні ділянки та все, що розташоване на них і міцно з ними пов'язане, тобто об'єкти, переміщення яких неможливе без нанесення значної шкоди його вартості і призначенню. [20]. Законодавче визначення віднесення об'єктів цивільних прав до нерухомого майна дозволяє виділити такі критерії його здійснення: природні властивості об'єктів (земельні ділянки, ділянки надр, відокремлені водні об'єкти), фізичні властивості об'єктів (тісний зв'язок із землею, неможливість переміщення без знецінення та зміни їхнього призначення – ліси, будівлі, споруди та інші об'єкти), об'єкти, вказані у законі.

Розкривають зміст поняття «нерухомість» також наступні законодавчі акти, зокрема: Закон України «Про іпотеку» – нерухоме майно (нерухомість) – це земельні ділянки, а також об'єкти, розташовані на земельній ділянці і невід'ємно пов'язані з нею, переміщення яких є неможливим без їх знецінення та зміни їх призначення (стаття 1) [25]. У даному означенні підкреслюється невід'ємність нерухомого майна від землі. Закон України «Про збір на обов'язкове державне пенсійне страхування» нерухомістю визнає жилий будинок або його частину, квартиру, садовий будинок, дачу, гараж, іншу постійно розташовану будівлю, а також інший об'єкт, що підпадає під визначення першої групи основних фондів згідно із Законом України «Про оподаткування прибутку підприємств» (стаття 1) [22].

Ураховуючи наведені законодавчі акти, можна виділити такі ознаки нерухомості:

1) ці об'єкти розташовані на земельній ділянці та нерозривно пов'язані з землею, тобто вони не можуть існувати без землі;

2) переміщення зазначених об'єктів є неможливим без їх знецінення та зміни їх призначення. Вказані ознаки є основними і їх розглядати необхідно у сукупності. При цьому, крім ЦК слід ураховувати інші нормативні акти

(спеціальне законодавство), які визначають статус того чи іншого виду нерухомого майна.

Щодо класифікації об'єктів нерухомості, то вона є різноманітною. Розглянемо поділи нерухомості, наведені в законодавстві України.

У ч. 2 ст. 331 ЦК України закріплено невичерпний перелік об'єктів нерухомого майна та проведено його поділ на житлові будинки, будівлі, споруди. Закон України «Про інвестиційну діяльність» від 18.09.1991 р. № 1560-ХІІ (із внесеними змінами та доповненнями) поділяє нерухоме майно на будинки, споруди, устаткування та інші матеріальні цінності (ст. 1) [24], а Закон України «Про оренду державного та комунального майна» від 10.04.1992 р. № 2269-ХІІ (із внесеними змінами та доповненнями) – на будівлі, споруди, приміщення (ст. 4) [26].

У Державному реєстрі прав реєструються права на такі об'єкти нерухомого майна, розташовані на земельній ділянці, переміщення яких неможливе без їх знецінення та зміни призначення:

- 1) житлові будинки;
- 2) квартири;
- 3) будівлі, в яких розташовані приміщення, призначені для перебування людини, розміщення рухомого майна, збереження матеріальних цінностей, здійснення виробництва тощо;
- 4) споруди (інженерні, гідротехнічні тощо) – земельні поліпшення, що не належать до будівель та приміщень, призначені для виконання спеціальних технічних функцій;
- 5) приміщення – частини внутрішнього об'єму житлових будинків, будівель, квартир, обмежені будівельними елементами [23].

У практичній діяльності з управління нерухомістю її об'єкти часто класифікуються за такими ознаками:

- а) спеціалізації
  - спеціалізована нерухомість, яка в силу свого спеціального характеру рідко продається на відкритому ринку для продовження її існуючого

використання (крім випадків, коли вона реалізується як частина бізнесу, який її використовує) (нафтопереробні та хімічні заводи, електростанції). Для об'єктів з такими конструктивними особливостями, розмірами і специфікою ринок не існує взагалі або в даній місцевості;

– неспеціалізована – вся інша нерухомість, на яку існує загальний попит на відкритому ринку для використання в існуючих або аналогічних цілях;

б) ступенем готовності до експлуатації:

- введені в експлуатацію;
- які вимагають реконструкції або капітального ремонту;
- незавершене будівництво;

в) відтворюваності в натуральній формі:

- невідтворювані (земельні ділянки, родовища корисних копалин);
- відтворювані (будівлі, споруди, багаторічні насадження).

г) походження:

– природні об'єкти – земельна ділянка, ліс і багаторічні насадження, відокремлені водні об'єкти і ділянки надр. Ці об'єкти нерухомості називають ще й нерухомістю за природою. Земля як єдине місце проживання всіх людей, основний фактор в будь-якій сфері бізнесу, прямо або побічно бере участь у виробництві всіх інших товарів і благ, є базовим об'єктом нерухомості.

– штучні об'єкти (споруди) – «нерухомість за законом». Вона як категорія нерухомості спирається на «нерухомість за природою».

Найчастіше використовується класифікація об'єктів нерухомого майна по сегментах ринку нерухомості, які сформувалися залежно від специфічних споживчих властивостей і характеристик її різних видів, умов використання, особливостей ринкових угод і дій їх учасників, складу споживачів. Це – земельні ділянки, житлова і нежитлова нерухомість.

Житлова нерухомість представлена різного виду об'єктами, призначеними для постійного проживання в них сімей або окремих індивідів. Вона пов'язана з орендою, будівництвом і експлуатацією житлового фонду. У її складі виділяють малоповерхові (до трьох поверхів), багатоповерхові (від 4 до 9

поверхів), підвищеної поверховості (від 10 до 20 поверхів), висотні (понад 20 поверхів) будинку. Об'єктом житлової нерухомості може також бути кондомініум, секція (під'їзд), поверх в під'їзді, квартира, кімната, дачний будинок.

У складі нежитлової нерухомості виділяють комерційну, промислову, інституційну та рекреаційну нерухомість, готелі та мотелі.

Під комерційною нерухомістю розуміють споруди для роздрібної торгівлі, офіси для бізнесу, стоянки автомобілів, бензоколонки та інші об'єкти сфери обслуговування.

Промислова нерухомість являє собою промислові споруди та інфраструктуру, діючі заводи, технопарки, складські приміщення, спеціалізовані офіси для управлінського персоналу та ін.

Готелі та мотелі представляють особливий клас нерухомості. На відміну від житлової, ці об'єкти призначені для тимчасового, а не постійного користування.

Рекреаційна нерухомість пов'язана зі сферою відпочинку. Це різного роду споруди на території курортних зон, клуби за інтересами (гольф-клуби, яхт-клуби та ін.), Стадіони, плавальні басейни та інші спортивно-оздоровчі комплекси і комплекси розваг.

Інституційна нерухомість – будівлі для урядових і муніципальних органів влади, навчальних закладів, лікарень, санаторіїв та інших функцій спеціального призначення [18].

Часто виділяють нерухомість змішаного типу використання, яка являє собою комбінації наведених вище видів нерухомості.

У даній роботі акцентується увага на торгівельній та офісній нерухомості, що є частиною комерційної нерухомості. Торгівельна нерухомість – це різновид комерційної нерухомості, яка призначена для організації місць роздрібної торгівлі та дрібнооптового продажу товарів і послуг. Основним чинником, який сприяє активному розвитку ринку та залученню національних та іноземних гравців, є значна нестача якісних торгових приміщень (попит значно перевищує

пропозицію). Офісна нерухомість разом із торговою є найбільш динамічно розвиваючим сегментом на ринку комерційної нерухомості України. Зростання ділової активності вітчизняних підприємців і прихід до України іноземних інвесторів викликали потребу в офісних приміщеннях високої якості, нестача в яких призвела до високих орендних ставок, високої прибутковості інвестицій в проєкти офісної нерухомості [33].

## 1.2 Аналіз економічних відносин на ринку комерційної нерухомості

У широкому розумінні поняття «ринок нерухомості» можна трактувати як комплекс відносин, пов'язаних зі створенням нових та експлуатацією вже існуючих об'єктів нерухомості.

Вивченню теоретичних та практичних аспектів функціонування ринку нерухомості приділяли увагу такі вчені: Дж. Фрідман, Н. Ордуей, С. Мочерний, А. Микитчик, М. Коваленко, Л. Радванська, Л. Чубук, І. Крекотень, Р. Ткаченко, С. Шибірина, Д. Бонцевич, С. Максимов, В. Прокопенко та ін. Вказаними вченими визначено основну суть поняття «ринок нерухомості», а також закладено теоретичну основу його державного регулювання.

Найбільш поширене визначення ринку нерухомості, на наш погляд, дали американські вчені Дж. Фрідман та Н. Ордуей. На їхню думку, ринок нерухомості – це певний набір механізмів, за допомогою яких передаються права на власність та пов'язані з нею інтереси. Ринок нерухомості – це також взаємозв'язок суб'єктів ринку на основі принципів, методів та процедур [6]. За означенням С. Мочерного, ринок нерухомості – це особливий вид ринку, на якому об'єктом купівлі-продажу чи застави для отримання кредиту є нерухомість; один із основних видів ринку, який у взаємодії з ринком товарів і послуг та ринком капіталів визначає особливості господарського механізму сучасної регульованої економіки [29].

Тобто основні процеси функціонування ринку нерухомості – це розвиток (створення), управління (експлуатація) та передання прав на нерухомість. На

наш погляд, нерухомість в умовах ринкової економіки слід тлумачити як один із факторів, залучених у процес виробництва товарів і послуг з метою отримання доходу та його забезпечення шляхом обороту в межах товарно-грошових відносин [33].

Таким чином, на ринку комерційної нерухомості виділяють дві його складові: первинний і вторинний. Первинний ринок – економічна ситуація, коли нерухомість як товар вперше з'явився на ринок. Основними продавцями нерухомості у такому разі виступають держава в особі своїх загальнодержавних, регіональних і місцевих органів влади, будівельні компанії – постачальники нежитлових будов. На вторинному ринку нерухомість виступає як товар, раніше вживаний і такий, що належить певному власнику – фізичній або юридичній особі.

На первинному і вторинному ринках нерухомості, полягає в тому, що на них представлені принципово різні за своєю природою продавці.

На первинному ринку продавець – це підприємець, прагне до максимізації свого сукупного доходу, самостійно і оперативно приймати рішення.

На вторинному ринку продавець – це «обиватель», прагне до максимізації доходу за даною угодою, значною мірою орієнтується на поведінку інших продавців та інертний в своїй поведінці.

На первинному ринку продавці-забудовники прагнуть максимізації доходу від загального обсягу продажів, на вторинному ринку продавці-власники – до максимізації доходу від угоди.

Кількість нерухомості, пропонованої на первинному ринку, залежить від нового будівництва. Первинний і вторинний ринки нерухомості взаємопов'язані. Якщо з яких-небудь причин (наприклад, несприятливе екологічне середовище та ін.) збільшиться пропозиція нерухомості на вторинному ринку, це приведе до знецінення нерухомості на первинному ринку [21].

Також можна зустріти наступне означення – ринок нерухомості являє собою сферу вкладення капіталу в об'єкти нерухомості та систему економічних відносин, що виникають при операціях з нерухомістю. Ці відносини виникають між інвесторами при купівлі-продажу нерухомості, іпотеці, здачі об'єктів нерухомості в траст, в оренду тощо. На рис. 1.1 наведено операції, які проводяться на ринку нерухомості Івано-Франківська із об'єктами та їх відсоткове співвідношення.

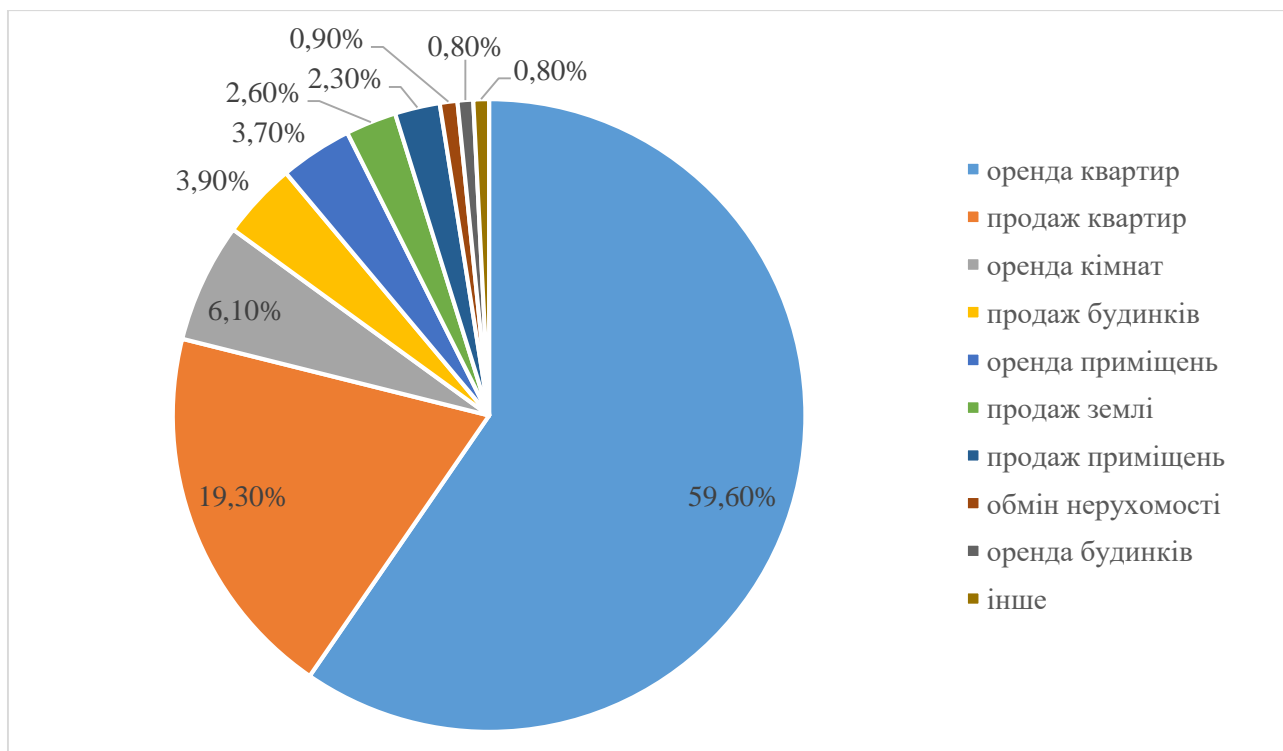


Рисунок 1.1 – Операції, що проводяться на ринку нерухомості Івано-Франківська та частота їх виконання у 2020 році у відсотковому співвідношенні

Бачимо, що найбільше операцій проводиться із процесом оренди квартир, а також їх продажами. Оренда приміщень займає 3,7% всього ринку нерухомості.

Отже, незважаючи на те, що вищевказані визначення не включають в себе однакові формулювання і не спираються на одне джерело, всі вони представляють ринок, як сукупність взаємовідносин між власником і покупцем (орендарем), за коштами яких здійснюється передача прав володіння об'єктом. Тобто ми можемо зробити висновок, що найбільш поширені два товари на

ринку нерухомості – сам об'єкт і послуга (тимчасове користування). Очевидно, що ціни на ці товари відрізняються.

У даній роботі ми будемо розглядати вартість оренди, а не вартість покупки об'єкту, так як для досліджуваного підприємства основним видом діяльності є здача в оренду власних торгівельних та офісних приміщень. Варто зауважити, що величина орендної плати так чи інакше впливає на ціну продажу об'єкта, якщо власник хоче повністю продати права володіння (в нашому випадку це малоймовірно, проте варте для зауваження). Виходить, що орендна плата – основа не тільки при укладенні договору оренди, а й при оцінці повної вартості дохідної нерухомості, що обумовлює значимість вивчення її формування.

Для того, щоб розібратися, як відбувається формування орендної плати, необхідно розібратися в суті орендних відносин.

Економічний сенс орендних відносин полягає в можливості орендаря використовувати майно для вилучення з нього корисних властивостей, а для орендодавця – в отриманні від орендаря плати за використання (використання і володіння) належить орендодавцеві майна (орендної плати). Орендні відносини показують, як один об'єкт нерухомості одночасно забезпечує реалізацію двох економічних інтересів: власника (орендодавця) і користувача (орендаря) [31].

У Законі України «Про оподаткування прибутку підприємств» поняття оренди трактують як «господарську операцію фізичної або юридичної особи, що передбачає відповідно до договору оперативного лізингу (оренди) передання орендарю майна, що підпадає під визначення основного фонду згідно із ст. 8 цього Закону, придбаного або виготовленого орендодавцем на умовах інших, ніж передбачаються фінансовим лізингом (орендою)» [22]. Статтею 793 ГКУ визначено, що оренда – це «договір найму будівлі або іншої капітальної споруди (їх окремої частини) укладається у письмовій формі» [19].

Оренда не змінює власника і означає розподіл функцій власника та господарюючого володаря (користувача). Орендар при цьому є власником виробленої продукції та послуг, доходів від їх реалізації, а також придбаного за

їх рахунок майна. Орендар несе відповідальність за використання орендованої власності.

При проведенні орендарем за рахунок власних коштів та за згодою орендодавця реконструкції майна або будівництва нових будівель та споруд можливо наступне:

- 1) відшкодування орендарю витрат після закінчення договору оренди;
- 2) перехід новоствореного майна у власність орендаря;
- 3) перетворення новоствореного майна в акціонерну (пайову) власність.

Орендар має право на викуп об'єкта оренди, якщо таке право передбачено договором оренди.

Оренда може застосовуватись до будь-якої форми власності: державної, колективної, приватної, тобто орендодавцями можуть бути державні органи, громадяни, товариства.

Основним документом, що регулює відносини між орендарем та орендодавцем, є договір оренди. Умовами договору оренди є:

- мета оренди;
- суб'єкт оренди;
- об'єкт оренди (склад та вартість майна за залишковою відновною вартістю). Перед укладанням договору проводиться оцінка майна та до договору додається баланс;
- термін дії договору;
- розмір орендної плати і строки її внесення;
- порядок використання амортизаційних відрахувань (орендарем або передаються орендодавцю);
- відновлення орендованого майна та умови його повернення або викупу;
- права та обов'язки сторін. Обов'язками орендодавця можуть бути матеріально-технічне забезпечення, обслуговування устаткування, консультаційна допомога; обов'язок орендаря
- підтримувати робочий стан об'єкта;
- умови зміни та розірвання договору.

Наприклад, орендодавець може розірвати договір, якщо орендар використовує майно не за призначенням, погіршує стан майна, не вносить в термін орендну плату. Орендар відшкодовує орендодавцю вартість орендованого майна у разі відчуження цього майна або його знищення чи псування з вини орендаря.

Одностороння відмова від договору оренди не допускається. Договір оренди припиняється у разі:

- закінчення строку, на який його було укладено;
- викупу (приватизації) об'єкта оренди;
- ліквідації суб'єкта господарювання-орендаря;
- загибелі (знищення) об'єкта оренди (наприклад, внаслідок стихійного лиха).

Договір оренди може бути розірваний за згодою сторін.

Щодо орендних відносин на ринку комерційної нерухомості Івано-Франківська, то попит на оренду торгових приміщень в Івано-Франківську за останні роки істотно активувався. Попит на оренду торгових площ переважно формують заклади громадського харчування (кав'ярні, кафе, ресторани), магазини одягу, мобільного зв'язку, побутової хімії, торгівля вживаними товарами (секонд-хенд).

В Івано-Франківську найбільш затребувані комерційні приміщення площею від 20 до 50 кв. м. Основний попит на оренду торговельних площ зосереджений у центрі міста, який умовно обмежується вул. Незалежності, вул. Новгородською, вул. Дністровською та вул. І. Франка.

Фахівці Івано-Франківська виділяють наступні діапазони ціни оренди торгових приміщень площею близько 100 кв. м в місті у залежності від їх місця розташування й людяності:

- 1) «5»: 400-500 грн./кв.м.
- 2) «4»: 250-350 грн./кв.м
- 3) «3»: 150 -200 грн./кв.м.
- 4) «2»: до 100 грн./кв.м.

Місце розташування оцінюється за 5-ти бальною системою: оцінка «5» відповідає топовим місцям у місті з найбільш інтенсивним знаходження потенційних покупців, «2» – інтенсивність потенційних покупців низька.

Аналізуючи все вищесказане, можна зробити висновок про те, що власник комерційної нерухомості орієнтований на отримання доходу і окупність початкових вкладень, шляхом здачі приміщень в оренду. Тому для будь-якої компанії-орендодавця є важливим формування орендної плати. Даний процес вимагає детального вивчення, при чому, як реальної процедури, так і пропонованих в теорії методів.

### 1.3 Дослідження питання розрахунку орендної плати в Україні

Згідно із законодавством України, орендна плата – це фіксований платіж, який орендар сплачує орендодавцеві за користування його майном, причому незалежно від результатів своєї господарської діяльності (ч. 1 ст. 286 ГК) [19]. Обидві сторони можуть керуватися нормами Закону № 2269 і Методикою № 786 (п. 4 ст. 1 Закону № 2269) [26]. Відповідно до даних законодавчих документів, порядок розрахунку орендної плати є наступним:

1) При оренді нежитлового приміщення, яке є частиною будівлі (споруди) орендна плата розраховується за формулою:

$$V_n = \frac{V_b}{P_b} \cdot P_n, \quad (1.1)$$

де  $V_n$  – вартість орендованого приміщення, яке є частиною будівлі (споруди), грн.;

$V_b$  – вартість будівлі (споруди) у цілому (без вартості підвальних приміщень, якщо вони не експлуатуються орендарем), визначена експертним шляхом, грн.;

$P_b$  – площа орендованого приміщення, кв. м;

$P_n$  – площа будівлі (споруди) у цілому (без площі підвальних приміщень, якщо вони не експлуатуються орендарем), кв. м.

2) Визначити розмір плати за перший місяць оренди після укладення договору оренди (або перегляду розміру орендної плати) можна за формулою:

$$O_{\text{пл.міс.}} = O_{\text{пл.}} \cdot 12 \cdot I_{\text{п.р.}} \cdot I_{\text{м.}}, \quad (1.2)$$

де  $O_{\text{пл.}}$  – річна орендна плата, грн.;

$I_{\text{п.р.}}$  – індекс інфляції в період з дати проведення експертної оцінки до дати укладення договору оренди або перегляду розміру орендної плати, грн.;

$I_{\text{м.}}$  – індекс інфляції за перший місяць оренди.

3) Для того, щоб обчислити розмір орендної плати за кожний наступний місяць, потрібно розмір орендної плати за попередній місяць скоригувати на індекс інфляції за поточний місяць.

Якщо ж сторони не бажають застосовувати вказану в цьому законі методику, то розмір і склад орендної плати вони повинні визначити за взаємною угодою (ч. 1 ст. 762 ЦК) [34]. Тобто вартість оренди може встановлюватися орендодавцем будь-яким методом, важливо, щоб розмір орендної плати був погоджений орендарем і записаний в договорі.

Найпростішим (узагальненим) способом для орендодавця встановити вартість оренди згідно з класичною формулою розрахунку ціни, а саме як різницю доходу і витрат.

Як правило, при розрахунку орендної плати орендодавець враховує ряд показників:

- суму амортизаційних відрахувань, розраховану за період, у якому нерухоме майно передається в оренду;
- витрати на обслуговування об'єкта оренди (наприклад, зарплата працівникам, які охороняють, прибирають передане в оренду приміщення, і т. д.);
- відсоток прибутку, який орендодавець планує отримати від здачі нерухомості в оренду.

Окрім витрат, величина орендної плати залежить від попиту і пропозиції. За угодою сторін розмір орендної плати може коригуватися на індекс інфляції.

Щодо комунальних послуг, варто зазначити важливий момент, що ніяких обмежень за складом орендної плати за користування недержавним майном у законодавчих актах не існує. Тому вартість спожитих орендарем комунальних послуг з однаковим успіхом може:

- або включатися до складу орендної плати,
- або компенсуватися орендарем на підставі рахунків орендодавця (комунальних служб), виставлених окремо від рахунків за оренду (так само, як і при оренді нерухомості державної форми власності). При цьому в договорі робиться потрібне застереження.

Важливо, що вартість оренди комерційної нерухомості прямо пропорційно залежить від вартості цієї нерухомості (її ціни). А ціна нерухомості – це не просто ціна певного товару, а інтегральний показник, тобто результуюча величина комплексу складових чинників впливу макросередовища (демографічна ситуація, соціальне та майнове розшарування в суспільстві, міграційна політика, тощо). Також процес формування ціни враховує, з одного боку, переваги споживача (які надають конкретному товару певну цінність), а з іншого – вартість товару (ті витрати, які неминуче несе товаровиробник, перетворюючи ресурси на благо). Таким чином, можна виділити багато факторів, що певною мірою впливають на формування вартості оренди, а перелік і вагомість цих факторів може бути індивідуальним для кожного орендодавця, тому можливі різноманітні специфікації моделей оцінки вартості комерційної нерухомості. Особливості діяльності досліджуваного підприємства (що виступає орендодавцем на ринку комерційної нерухомості) та факторів, що впливають на встановлення вартості оренди, розглянемо в наступному розділі.

## Висновки

З правової точки зору нерухомість – це об'єкти права власності господарюючих суб'єктів і громадян, відносно яких здійснюються правочини, такі як: купівля-продаж, оренда, передача прав на володіння і користування припиненням цього права. З економічної точки зору ринок нерухомості є категорією, котра пов'язана зі створенням умов, необхідних для здійснення виробничої, комерційної, соціальної, екологічної та іншої діяльності.

Можна виділити такі ознаки нерухомості: об'єкти розташовані на земельній ділянці та нерозривно пов'язані з землею, тобто вони не можуть існувати без землі; переміщення зазначених об'єктів є неможливим без їх знецінення та зміни їх призначення.

У даній роботі акцентується увага на торгівельній та офісній нерухомості, що є частиною комерційної нерухомості. Торгівельна нерухомість – це різновид комерційної нерухомості, яка призначена для організації місць роздрібної торгівлі та дрібнооптового продажу товарів і послуг. Основним чинником, який сприяє активному розвитку ринку та залученню національних та іноземних гравців, є значна нестача якісних торгових приміщень (попит значно перевищує пропозицію). Офісна нерухомість разом із торговою є найбільш динамічно розвиваючим сегментом на ринку комерційної нерухомості України.

У широкому розумінні поняття «ринок нерухомості» можна трактувати як комплекс відносин, пов'язаних зі створенням нових та експлуатацією вже існуючих об'єктів нерухомості. На ринку комерційної нерухомості виділяють дві його складові: первинний і вторинний. Первинний ринок – економічна ситуація, коли нерухомість як товар вперше з'явився на ринку. На вторинному ринку нерухомість виступає як товар, раніше вживаний і такий, що належить певному власнику – фізичній або юридичній особі.

За даними операцій, що проводилися на ринку нерухомості Івано-Франківська у 2020 році, оренда приміщень займає 3,7 % усього ринку нерухомості.

У даній роботі ми будемо розглядати вартість оренди, а не вартість покупки об'єкту, так як для досліджуваного підприємства основним видом діяльності є здача в оренду власних торговельних та офісних приміщень.

Важливо, що вартість оренди комерційної нерухомості прямо пропорційно залежить від вартості цієї нерухомості (її ціни). А ціна нерухомості – це не просто ціна певного товару, а інтегральний показник, тобто результуюча величина комплексу складових чинників впливу макросередовища (демографічна ситуація, соціальне та майнове розшарування в суспільстві, міграційна політика тощо).

## РОЗДІЛ 2 ПОБУДОВА МОДЕЛЕЙ ДЛЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ВАРТОСТІ ОРЕНДИ ТОРГОВЕЛЬНИХ ПРИМІЩЕНЬ

### 2.1 Побудова моделі дискретного вибору

Моделі дискретного вибору використовуються для пояснення або прогнозування вибору з множини двох або більше дискретних, тобто взаємовиключних, альтернатив [10]. Наприклад, модель дискретного вибору може бути використана для аналізу того, чому людина вирішила добиратися на роботу пішки, на метро, на машині, або ж на самокаті, або, скажімо, для аналізу факторів, що змушують студентів ставити пріоритетніше один університет понад іншим.

Моделі дискретного вибору діють у рамках раціонального вибору; тобто передбачається, що, зіткнувшись з дискретною множиною опцій, люди обирають варіант максимальної вигоди або корисності. З цього припущення випливає, що корисність вибору є функцією характеристик можливого вибору та характеристик людини, яка робить вибір [15]. Моделі дискретного вибору характеризують цю функцію в сукупності, тим самим дозволяючи статистично зробити висновок про функціональні параметри. Наприклад, застосувавши модель дискретного вибору до набору даних про вибір людиною способу добирання до робочого місця (пішки, метро, машина, самокат) для 100 осіб протягом 2 тижнів, можна виявити, що вибір добирання пов'язаний як з характеристиками вибору (наприклад, вибір ходити пішки може бути притаманний небагатьом людям у загальному від усієї когорти людей), так і взаємодії між виборами та характеристиками людей (наприклад, ті, хто живе на середній відстані від роботи й поруч із стоянкою для самокатів, можуть вибрати частіше приїжджати на роботу саме на цьому виді транспорту).

Моделі дискретного вибору можна відрізнити від стандартних регресійних моделей явним включенням певного набору варіантів, деякі з яких не були обрані. Існує два типи даних про вибір: заявлений вибір та виявлений вибір [8]. Дані про заявлений вибір отримуються з гіпотетичних сценаріїв або з

набору варіантів, представлених опитуваним (наприклад, запитаннями: "Ви хочете більше ходити пішки, чи їхати на метро до роботи?"). У той час, як виявлений вибір аналізують із записаних виборів суб'єктів у минулому. Наприклад, опитувані фіксують, як вони починають добиратися до роботи щодня протягом 2 тижнів, а потім аналізуються результати. Краща форма вибору залежить від питання дослідження. Наприклад, дослідження, спрямоване на збільшення робочої сили в галузі охорони здоров'я у сільській місцевості, використовували б заявлені вибори замість виявлених, оскільки виявлених просто не можна було б отримати із робочих місць, яких не існує. У деяких випадках дослідники можуть використовувати комбінацію обох типів, потенційно для порівняння обох отриманих виборів.

Моделі дискретного вибору характеризуються двоетапною процедурою: спочатку аналітик генерує множину вибору, тобто набір усіх варіантів вибору. Далі застосовує модель із (прихованою) корисністю ліворуч від рівняння та характеристиками суб'єктів, характеристиками вибору та деякою помилкою праворуч [4].

Зупинимося на генеруванні множини вибору. Існує принаймні три типи множин, тобто дискретних змінних, кожна із яких вимагає свою власну статистичну модель.

Дихотомічні, бінарні, чи фіктивні змінні. Вони приймають лише два значення: 0 та 1. Надалі будемо розглядатимемо випадок, коли така змінна знаходиться в лівій частині рівняння, тобто є ендогенною, або ж залежною. У даному випадку з'являються нові труднощі на відміну від ситуації, коли фіктивна змінна є екзогенною.

Так, продовжуючи приклад із добиранням людини до роботи, можемо ввести таку дихотомічну змінну:

$$y_i = \begin{cases} 1, \text{ якщо людина } i \text{ дібралася на роботу пішки;} \\ 0, \text{ у протилежному випадку.} \end{cases}$$

Як інші аналогічні приклади, можна навести рішення купувати сьогодні морозиво, чи ні, або зустрітися із друзями сьогодні, чи ні.

Фіктивні змінні є одними з тих дискретних змінних, що найбільш часто зустрічаються, тому розглянемо цей тип моделей детальніше.

Наведемо приклад з біометричної літератури, де й зародилися моделі для ендогенних фіктивних змінних. Наприклад, розглянемо задачу оцінювання числа засобів для винищування комах. Нехай чутливість  $y_i^*$  комахи  $i$  до інсектицидів є нормально розподіленою величиною:

$$y_i^* \sim N(\mu, \sigma^2)$$

Якщо чутливість комахи менше дози  $x_i$  інсектицидів, то вона гине.

Проблема в тому, що ми не можемо спостерігати за чутливістю  $y_i^*$  кожної окремої комахи. Замість цього ми спостерігаємо за тим, комаха вижила чи ні. Тобто ми визначаємо  $y_i$  так:

$$y_i = \begin{cases} 1, & \text{якщо комаха } i \text{ гине;} \\ 0, & \text{у протилежному випадку.} \end{cases}$$

Тепер можна повернутися до питання про те, яка імовірність того, що комаха  $i$  загине. Це просто імовірність того, що чутливість комахи менше дози:

$$\text{prob}(y_i = 1) = \text{prob}(y_i^* < x_i). \quad (2.1)$$

У такому формулюванні  $y_i$  генерується за наступним правилом:

$$y_i = \begin{cases} 1, & \text{якщо } y_i^* < x_i; \\ 0, & \text{у протилежному випадку.} \end{cases}$$

У нашому прикладі  $y_i^*$  називається латентною (прихованою), чи індексною змінною, тому що вона не спостерігається, на відміну від  $y_i$ . Формування латентних змінних є більш зручним з аналітичної точки зору.

Поліхотомічні змінні. Такі змінні можуть приймати більше 2 можливих значень. У свою чергу поліхотомічні змінні діляться на невпорядковані та впорядковані змінні.

Невпорядковані змінні – це такі змінні, для яких не визначене ранжування вхідних значень. Наприклад, продовжуючи приклад із вибором одною способом добирання до роботи, можемо ввести наступну невпорядковану поліхотомічну змінну:

$$y_i = \begin{cases} 1, \text{ якщо людина } i \text{ йде пішки;} \\ 2, \text{ якщо людина } i \text{ їде на машині;} \\ 3, \text{ якщо людина } i \text{ їде на метро;} \\ 4, \text{ якщо людина } i \text{ їде на самокаті.} \end{cases}$$

Упорядковані змінні. Для таких змінних значення мають природний порядок. Наприклад, припустимо, що в нас є шкала оцінювання: незадовільно, задовільно, добре та відмінно. Тоді можна ввести наступну упорядковану поліхотомічну змінну:

$$y_i = \begin{cases} 1, \text{ якщо студент } i \text{ отримав незадовільно;} \\ 2, \text{ якщо студент } i \text{ отримав задовільно;} \\ 3, \text{ якщо студент } i \text{ отримав добре;} \\ 4, \text{ якщо студент } i \text{ отримав відмінно.} \end{cases}$$

На відміну від попереднього прикладу, тут існує природний порядок змінних: (відмінно) > (добре) > (задовільно) > (незадовільно). Особливим випадком упорядкованих змінних є послідовні змінні. Це буває у випадку, якщо наступна подія залежить від попередніх. Як приклад можна привести одержання наукових ступенів.

І останній тип дискретних змінних – рахункові моделі. У таких моделях змінні можуть приймати цілі значення. Ці моделі менше зустрічаються в економетричних дослідженнях і, звичайно, обробляються традиційними лінійними методами.

Перейдемо безпосередньо до побудови моделі дискретного вибору, а саме логіт-регресії (логістичної регресії). Для початку наведемо трішки незрозумілий на даний момент рис. 2.1, який показує принцип роботи логістичної регресії.

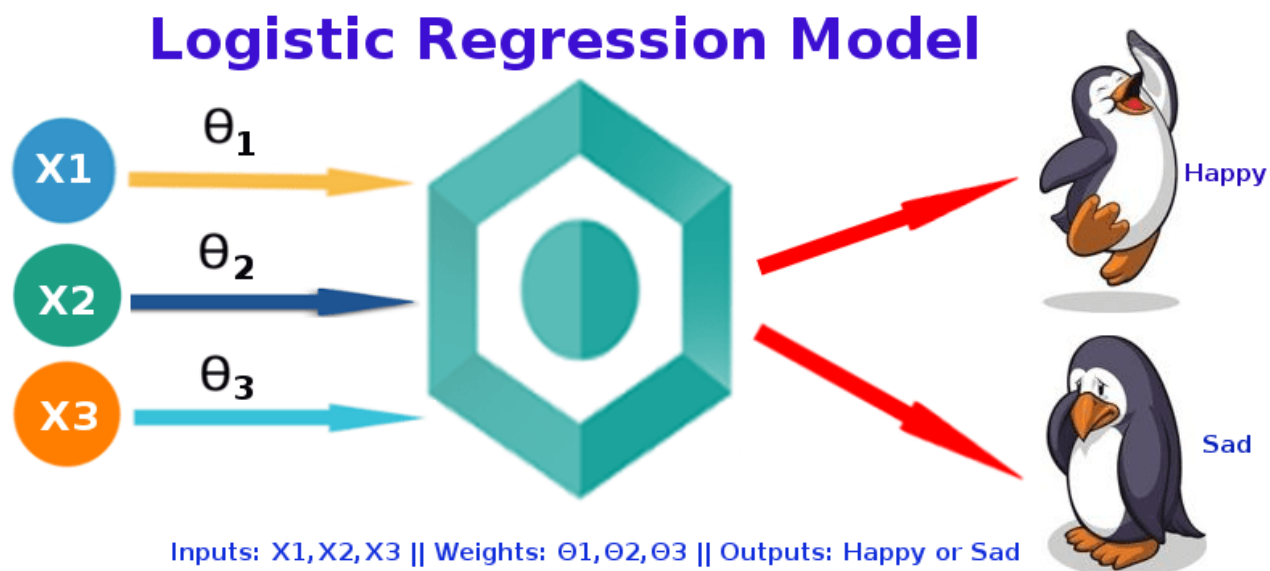


Рисунок 2.1 – Схема роботи логістичної регресії

Вперше логістичну регресію використали в біології на початку ХХ ст. Пізніше вона почала використовуватися для безлічі соціальних потреб. Як ми бачимо із рис. 2.1 логістичну регресію доцільно використовувати, коли залежна змінна (таргет) є категоріальною (дискретною). Наприклад, для прогнозу того, чи користується дане приміщення із якимись параметрами попитом, чи ні. Забігаючи наперед, це впринципі те, що ми й будемо робити в 3 розділі даної дипломної роботи [9].

Розглянемо вище наведений приклад детальніше. У табл. 2.1 наведений приклад даних оренди приміщень із деякими параметрами ( $x_j$ ) та залежною змінною, або таргетом ( $y$ ). Варто відзначити, що ці дані не експериментальні, а теоретичні й наведені в даному розділі лише для того, щоб показати приклад роботи й побудови логістичної регресії.

Таблиця 2.1 – Теоретичні дані оренди приміщень

Площа ( $m^2$ ) $x_1$	Поверх $x_2$	Ряд $x_3$	Тип $x_4$	Чи користується попитом? $y$
2700	1	2	1	1
170	2	1	2	0

Продовження таблиці 2.1

45	1	1	3	1
55	3	1	4	0
...	...	...	...	...

Щоб приступити до побудови логістичної регресії, введемо наступні позначення:  $n$  – кількість змінних;  $x^i$  – признаки (змінні)  $i$ -того приміщення;  $x_j^i$  – значення признаку  $j$  для  $i$ -того приміщення;  $y^i$  – значення таргету для  $i$ -того приміщення. Для даного прикладу:  $n = 4$ ;  $x^1 = (2700; 1; 2; 1)$ ,  $x^2 = (170; 2; 1; 2)$  і т.д.;  $x_1^1 = 2700$ ;  $x_2^1 = 1$ ;  $x_1^2 = 170$ ;  $x_3^1 = 45$  і т.д.;  $y^1 = 1$ ;  $y^2 = 0$  і т.д. Тут також варто відзначити, що змінна  $x_4$  – уже закодована змінна, тобто змінна, яка представлена в числовому форматі для того, щоб алгоритм міг використовувати математичні операції. У початковому форматі ця змінна могла бути представлена в текстовому вигляді: супермаркет (1), одяг (2), кафе (3), аптека (4). Про типи кодувань і переведення текстових змінних у числовий детальніше розповідатиметься в 3 розділі.

Використовуючи признаки (площа, поверх, ряд, тип) із табл. 2.1, ми хочемо спрогнозувати, чи користується попитом дане приміщення. Для цього введемо наступну гіпотезу:

$$h_{\theta}(x^i) = \theta_0 + \theta_1 x_1^i + \theta_2 x_2^i + \theta_3 x_3^i + \theta_4 x_4^i \quad (2.2)$$

Саме  $h_{\theta}(x^i)$  буде прогнозувати значення таргет змінної  $y^i$  для  $i$ -того приміщення за допомогою вектора параметрів  $\theta$ .

У загальному випадку для  $n$  признаков:

$$h_{\theta}(x^i) = \theta_0 + \theta_1 x_1^i + \theta_2 x_2^i + \dots + \theta_n x_n^i \quad (2.3)$$

Покладемо  $x_0 = 1$  для кожного приміщення. Ця змінна називається bias змінною. Тоді загальна формула із (2.3) перетворюється так:

$$h_{\theta}(x^i) = \theta^T x^i, \quad (2.4)$$

$$\text{де } x^i = \begin{pmatrix} x_0^i \\ x_1^i \\ x_2^i \\ \dots \\ x_n^i \end{pmatrix} \in \mathbb{R}^{n+1};$$

$$\theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \dots \\ \theta_n \end{pmatrix} \in \mathbb{R}^{n+1}.$$

Така гіпотеза застосовується для лінійної множинної регресії, проте в даному випадку вона не є дуже зручною, оскільки в даній задачі залежна змінна (таргет) приймає значення лише 0 та 1, тобто:

$$0 \leq h_{\theta}(x^i) \leq 1 \quad (2.5)$$

Для того, щоб ця умова виконувалася, розглянемо нову гіпотезу:

$$h_{\theta}(x^i) = g(\theta^T x^i), \quad (2.6)$$

де

$$g(z) = \frac{1}{1 + e^{-z}} \quad (2.7)$$

І в кінцевому результаті із (2.6) та (2.7) наша гіпотеза перетворюється в такий вигляд:

$$h_{\theta}(x^i) = \frac{1}{1 + e^{-\theta^T x^i}} \quad (2.8)$$

Функцію  $g(z)$  називають сігмоїдою, або ж логістичною функцією. На рис. 2.2 зображено графік цієї функції.

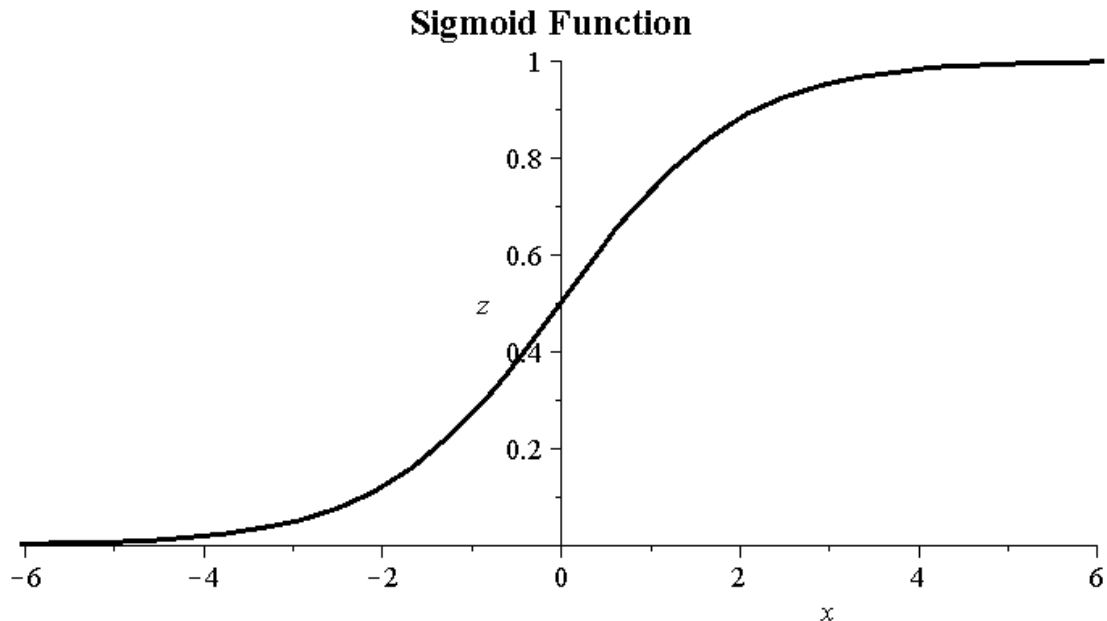


Рисунок 2.2 – Графік сігмоїди

Сігмоїда дає криву у формі букви S, яка може приймати будь-яке дійсне число й відобразити його у значення від 0 до 1. Якщо крива переходить до плюс нескінченності, у спрогнозоване стане 1, а якщо крива переходить до мінус нескінченності, у спрогнозоване стане 0. Якщо вихід сигмоїдної функції більше 0,5, ми можемо класифікувати результат як 1 або ТАК, а якщо він менше 0,5, ми можемо класифікувати його як 0 або НІ. Наприклад, якщо результат дорівнює 0,75, з точки зору ймовірності ми можемо сказати: “Існує 75 відсотків шансів, що дане приміщення користується попитом” [16].

Формульно ми можемо записати це так:

$$h_{\theta}(x^i) = P\{y = 1 | x^i; \theta\} \quad (2.9)$$

Надалі замість  $x^i$  будемо використовувати просто  $x$ , вважаючи, що це просто один екземпляр (у нашому прикладі одне приміщення).

Оскільки на початку ми вводили гіпотезу для лінійної регресії й потім за допомогою функції сігмоїди ввели гіпотезу для логістичної, доцільно буде порівняти ці два типи регресій між собою (рис 2.3).

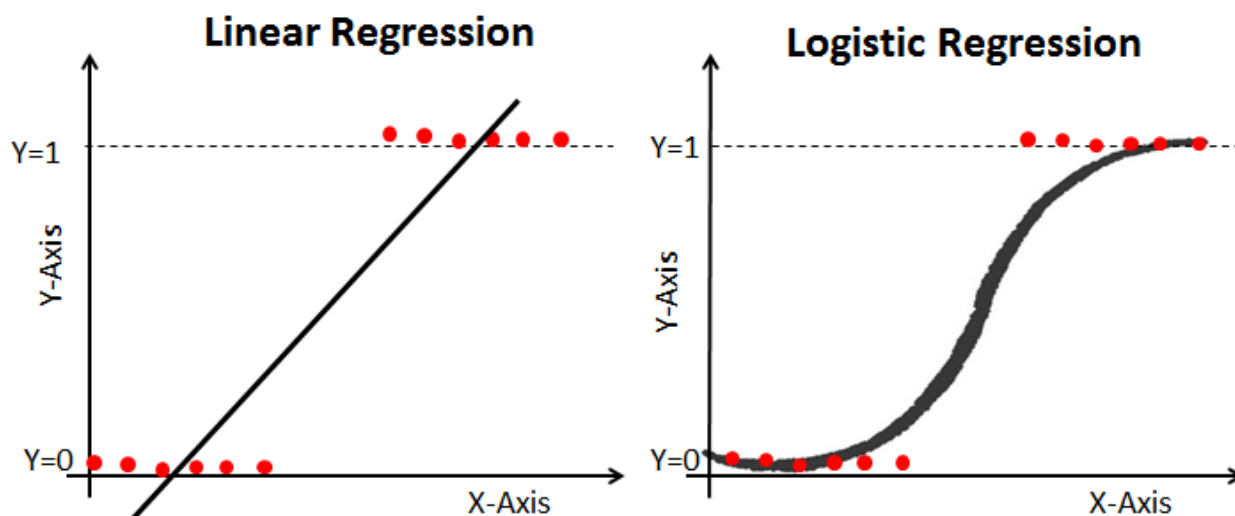


Рисунок 2.3 – Порівняння лінійної та логістичної регресії

Лінійна регресія дає нам неперервний результат, у той час як логістична регресія забезпечує дискретний. Прикладом неперервного виходу є ціна на оренду приміщення. Прикладом ж дискретного виходу – це прогнозування того, чи буде дане приміщення користуватися попитом (те, що ми й розглядали раніше й будемо розглядати більш детально на реальних даних у 3 розділі). Лінійна регресія оцінюється за допомогою методу найменших квадратів (МНК), тоді як логістична регресія оцінюється за допомогою методу максимальної правдоподібності (ММП) [16].

ММП – це метод максимізації "ймовірності", тоді як МНК – метод наближення, що мінімізує відстань. Функція максимізації правдоподібності визначає параметри, які найімовірніше дадуть спостережувані дані. Зі статистичної точки зору ММП встановлює середнє значення та дисперсію як параметри при визначенні конкретних параметричних значень для даної моделі. Цей набір параметрів можна використовувати для прогнозування даних, необхідних для нормального розподілу. Звичайні оцінки найменших квадратів обчислюються шляхом встановлення лінії регресії для заданих точок даних, яка має мінімальну суму квадратичних відхилень (найменша квадратична

помилка). Обидва вони використовуються для оцінки параметрів лінійної регресійної моделі. ММП приймає спільну функцію ймовірності маси, тоді як МНК не вимагає жодних стохастичних припущень для мінімізації відстані.

Введемо функцію втрат для логістичної регресії для одного екземпляру:

$$Cost(h_{\theta}(x), y) = \begin{cases} -\log(h_{\theta}(x)), & y = 1; \\ -\log(1 - h_{\theta}(x)), & y = 0. \end{cases} \quad (2.10)$$

Чому не можна використовувати МНК для логістичної регресії? Тому що тоді ми б отримували не випуклу функцію втрат для параметрів  $\theta$  (рис. 2.4). У такому випадку градієнтний спуск не завжди б сходився в глобальному мінімумі [6].

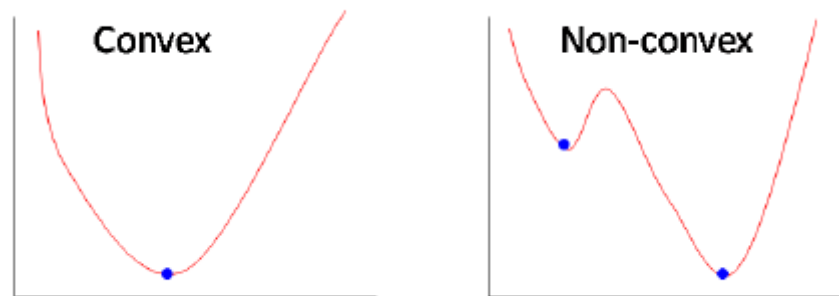


Рисунок 2.4 – Випукла та невивукла функція втрат

Пояснимо формулу 2.10.

Якщо  $y = 1$ , то  $Cost(h_{\theta}(x), y) = -\log(h_{\theta}(x))$  (рис. 2.5).

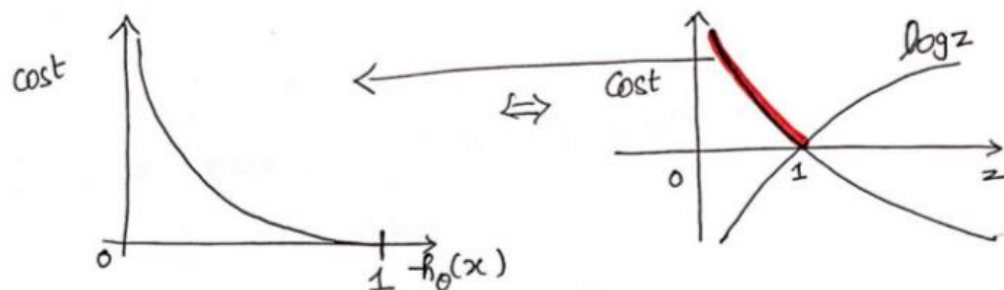


Рисунок 2.5 – Візуалізація (2.10), якщо  $y = 1$

Якщо:

- $Cost = 0 \Rightarrow y = 1 \Rightarrow h_{\theta}(x) = 1$ ;
- $Cost = \text{нескінченність}$  для  $h_{\theta}(x) = 0$ .

Якщо  $h_\theta(x) = 0$ , то це те саме, що  $P\{y = 1|x; \theta\} = 0$  із (2.9).

Якщо  $y = 0$ , то  $Cost(h_\theta(x), y) = -\log(1 - h_\theta(x))$  (рис. 2.6).

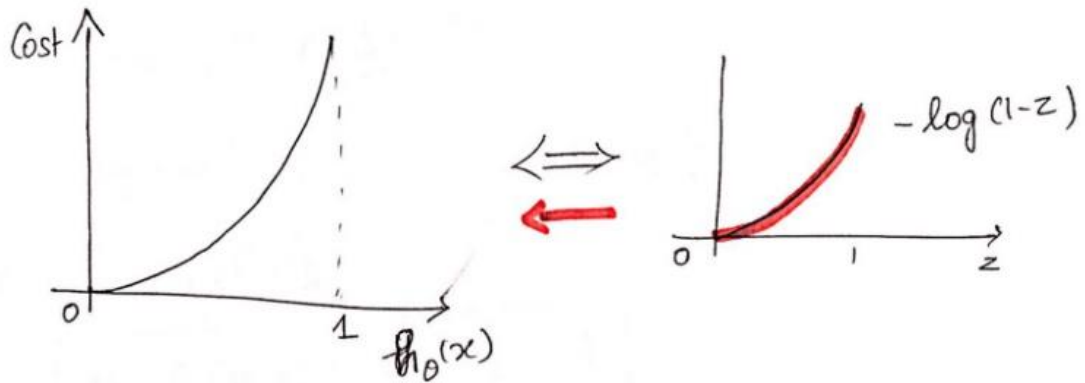


Рисунок 2.6 – Візуалізація (2.10), якщо  $y = 0$

Якщо:

- $Cost = 0 \Rightarrow y = 0 \Rightarrow h_\theta(x) = 0$ ;
- $Cost = \text{нескінченність}$  для  $h_\theta(x) = 1$ .

Якщо  $h_\theta(x) = 1$ , то це те саме, що  $P\{y = 0|x; \theta\} = 0$  із (2.9).

Перепишемо формулу (2.10) для знаходження функції втрат логістичної регресії в іншому вигляді:

$$Cost(h_\theta(x), y) = -y \log(h_\theta(x)) - (1 - y) \log(1 - h_\theta(x)) \quad (2.11)$$

Чому це вірно? Якщо  $y = 1$ , то множник  $(1 - y)$  стає рівним 0 і тільки  $-\log(h_\theta(x))$  буде присутнім у правій частині. Якщо ж  $y = 0$ , то множник  $y$  і тільки  $-\log(1 - h_\theta(x))$  буде присутнім у правій частині.

Чому саме ця функція втрат? Нехай із (2.9) маємо:

$$\hat{y} = P(y = 1|x) \quad (2.12)$$

$\hat{y}$  – імовірність того, що  $y = 1$ , якщо даний якийсь  $x$ . Із (2.12) легко побачити, що:

$$1 - \hat{y} = P(y = 0|x) \quad (2.13)$$

Із (2.12) та (2.13) отримуємо:

$$P(y|x) = \hat{y}^y * (1 - \hat{y})^{(1-y)} \quad (2.14)$$

Якщо  $y = 1$ , то  $P(y|x) = \hat{y}$ .

Візьмемо від обох частин рівності (2.14) натуральний логарифм:

$$\begin{aligned} \log(P(y|x)) &= \log(\hat{y}^y * (1 - \hat{y})^{(1-y)}) = y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}) = \\ &= \log P(y|x) = -Cost(\hat{y}, y) \end{aligned} \quad (2.15)$$

Знак мінус з'являється тому, що, коли ми тренуємо модель, ми маємо максимізувати ймовірність, мінімізуючи функцію втрат. Зменшення вартості збільшить максимальну ймовірність, якщо припустити, що вибірки беруться з однакового незалежного розподілу.

Здійснювати мінімізацію функції витрат ми будемо за допомогою градієнтного спуску, тобто на кожній ітерації будемо рухатися в напрямку глобального мінімуму функції витрат із деяким кроком  $\alpha$ .

Розглянемо приклад використання градієнтного спуску, коли в нас є два признаки  $(x_1, x_2)$ , по яких треба визначити таргет  $(y)$ . Із формул 2.3, 2.6 та 2.7 маємо:  $z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \Rightarrow \hat{y} = a = g(z)$ . Маємо мінімізувати функцію  $Cost(\hat{y}, y) = L(\hat{y}, y)$ . Знайдемо  $\frac{\partial(L)}{\partial \theta_1}$ :

$$\frac{\partial(L)}{\partial \theta_1} = \frac{\partial L}{\partial a} * \frac{\partial a}{\partial z} * \frac{\partial(z)}{\partial \theta_1} \quad (2.16)$$

$$\begin{aligned} \frac{\partial L}{\partial a} &= \frac{\partial}{\partial a} (-y * \log(a) - (1 - y) * \log(1 - a)) = \\ &= -y * \frac{1}{a} - (-1) * \frac{1 - y}{1 - a} = \left(\frac{-y}{a}\right) + \left(\frac{1 - y}{1 - a}\right) \end{aligned} \quad (2.17)$$

$$\frac{\partial a}{\partial z} = a(1 - a) \quad (2.18)$$

$$\frac{\partial z}{\partial \theta_1} = x_1 \quad (2.19)$$

Із 2.16, 2.17, 2.18 та 2.19 отримуємо:

$$\frac{\partial L}{\partial \theta_1} = \left( \left( \frac{-y}{a} + \frac{(1-y)}{1-a} \right) * (a)(1-a) \right) * x_1 = (a-y) * x_1 \quad (2.20)$$

Згідно із градієнтним спуском:

$$\theta_1 = \theta_1 - \alpha \frac{\partial L}{\partial \theta_1} \quad (2.21)$$

Із 2.20 та 2.21 маємо:

$$\theta_1 = \theta_1 - \alpha = \theta_1 - (\hat{y} - y) * x_1 \quad (2.22)$$

У загальному випадку:

$$\theta_i = \theta_i - \alpha \frac{\partial L}{\partial \theta_i}, \quad (2.23)$$

де  $\frac{\partial L}{\partial \theta_i} = \theta_i - (\hat{y} - y) * x_i$  для  $i = \overline{1, m}$ ;

$$\frac{\partial L}{\partial \theta_0} = \theta_0 - (\hat{y} - y).$$

Тепер модифікуємо функцію витрат, щоб вона враховувала всі наші екземпляри:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \text{Cost}(h_{\theta}(x^i), y^i) \quad (2.24)$$

Із 2.11 та 2.24 отримуємо:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m y^i * \log(h_{\theta}(x^i)) + (1 - y^i) * \log(1 - h_{\theta}(x^i)) \quad (2.25)$$

Щоб знайти вектор параметрів  $\theta$  необхідно мінімізувати функцію  $J(\theta)$ . Знову ж таки користуємося градієнтним спуском і всіма викладками, які ми наводили вище. Оскільки (2.24) використовує лінійну комбінацію старих функцій втрат, то й похідна від нової функції втрат буде лінійна комбінація від відповідних похідних старих функцій втрат:

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) = \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^i) - y^i) * x_j^i \quad (2.26)$$

Знайдений вектор параметрів  $\theta$  означає вирішення задачі й фактично закінчення побудови моделі логістичної регресії.

## 2.2 Побудова моделі авторегресії

Авторегресивна модель визначає, що вихідна змінна лінійно залежить від власних попередніх значень та від стохастичного терміна (термін, що недосконало передбачуваний); таким чином модель має форму стохастичного різницевого рівняння (або відношення рекурентності, яке не слід плутати з диференціальним рівнянням). Разом із моделлю ковзного середнього (МА) це особливий випадок і ключовий компонент більш загальних авторегресійно-ковзних середніх (ARMA) та авторегресійних інтегрованих ковзних середніх

(ARIMA) моделей часових рядів, які мають більш складну стохастичну структуру.

У моделі множинної регресії ми прогнозуємо змінну, використовуючи лінійну комбінацію деяких признаков. У моделі авторегресії ми прогнозуємо змінну, використовуючи лінійну комбінацію минулих значень цієї змінної. Термін авторегресія вказує на те, що це регресія є змінною щодо себе [2].

Відтак, авторегресійна модель порядку  $p$  може бути записана, як:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t, \quad (2.27)$$

де  $\varepsilon_t$  – білий шум.

Запис дещо схожий до множинної регресії, але в якості признаков використовуються попередні значення самої змінної  $y_t$ . Будемо називати цю модель AR( $p$ ), тобто авторегресивна модель порядку  $p$ .

Авторегресивні моделі надзвичайно гнучкі при роботі з широким спектром різних моделей часових рядів. Два графіки на рис 2.7 показують серії з моделі AR(1) та AR(2). Зміна параметрів призводить до різних моделей часових рядів. Дисперсія терміну помилки  $\varepsilon_t$  змінить лише масштаб ряду, а не закономірності.

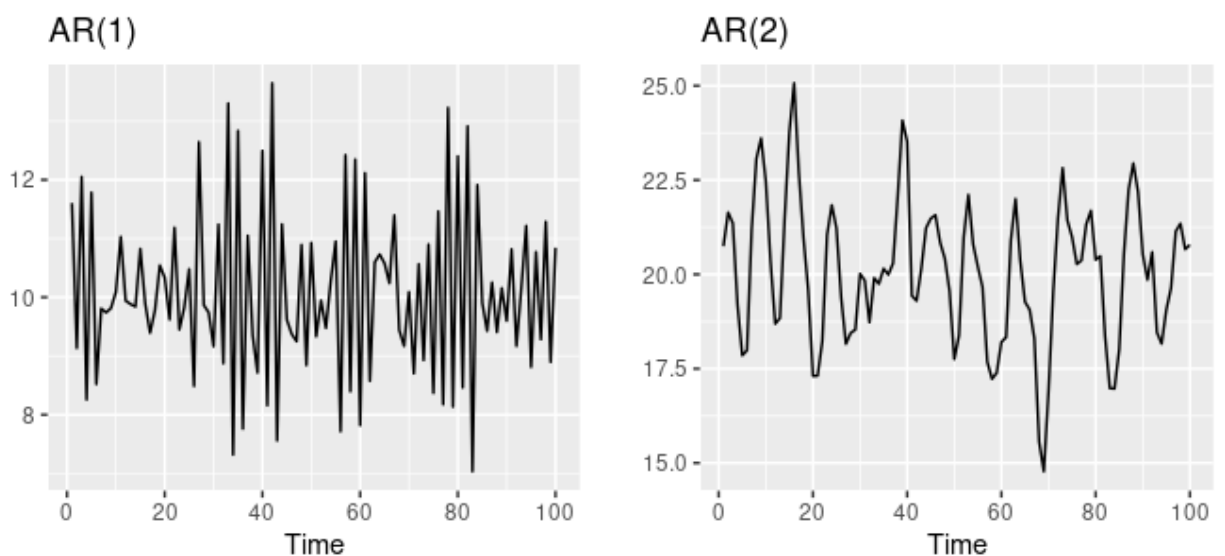


Рисунок 2.7 – Два приклади даних з авторегресивних моделей із різними параметрами

В обох випадках  $\varepsilon_t$  – нормально розподілений білий шум із нульовим математичним сподіванням та одиничною дисперсією (стандартний нормальний розподіл).

Для AR(1) моделі:

- якщо  $\varphi_1 = 0$ , то  $y_t$  еквівалентний білому шуму;
- якщо  $\varphi_1 = 1$  і  $c = 0$ , то  $y_t$  еквівалентний випадковому блуканню;
- якщо  $\varphi_1 = 1$  і  $c \neq 0$ , то  $y_t$  еквівалентний випадковому блуканню із дрейфом;
- якщо  $\varphi_1 < 1$ , то  $y_t$  має тенденцію осцилювати поблизу математичного сподівання.

Зазвичай ми обмежуємо авторегресійні моделі стаціонарними даними, і в цьому випадку потрібні певні обмеження на значення параметрів.

- Для AR(1):  $-1 < \varphi_1 < 1$ .
- Для AR(2):  $-1 < \varphi_2 < 1$ ;  $\varphi_1 + \varphi_2 < 1$ ;  $\varphi_2 - \varphi_1 < 1$ .
- Для  $p \geq 3$  умови набагато складніші.

Тепер розглянемо моделі із ковзним середнім (МА від англ. Moving Average). Замість того, щоб використовувати минулі значення змінної прогнозу в регресії, модель ковзного середнього використовує минулі помилки прогнозу в регресійній моделі [12].

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}, \quad (2.28)$$

де  $\varepsilon_t$  – білий шум.

Будемо називати цю модель MA(q), тобто модель ковзного середнього порядку q. Звичайно, ми на спостерігаємо значення  $\varepsilon_t$ , тож насправді це не регресія у звичному її розумінні.

Зверніть увагу, що кожне значення  $y_t$  ми можемо розглядати як зважене ковзне середнє серед кількох останніх помилок прогнозу. Однак модель ковзного середнього не слід плутати із згладжуванням ковзного середнього. Модель ковзного середнього використовується для прогнозування майбутніх

значень, тоді як згладжування ковзного середнього використовується для оцінки циклу тенденцій минулих значень.

На рис. 2.8 показано деякі дані із моделей MA(1) та MA(2). Зміна параметрів  $\theta_1, \dots, \theta_q$  призведе до різних моделей часових рядів. Як і в авторегресивних моделях, дисперсія терміну помилки  $\varepsilon_t$  змінить лише масштаб ряду, а не закономірності.

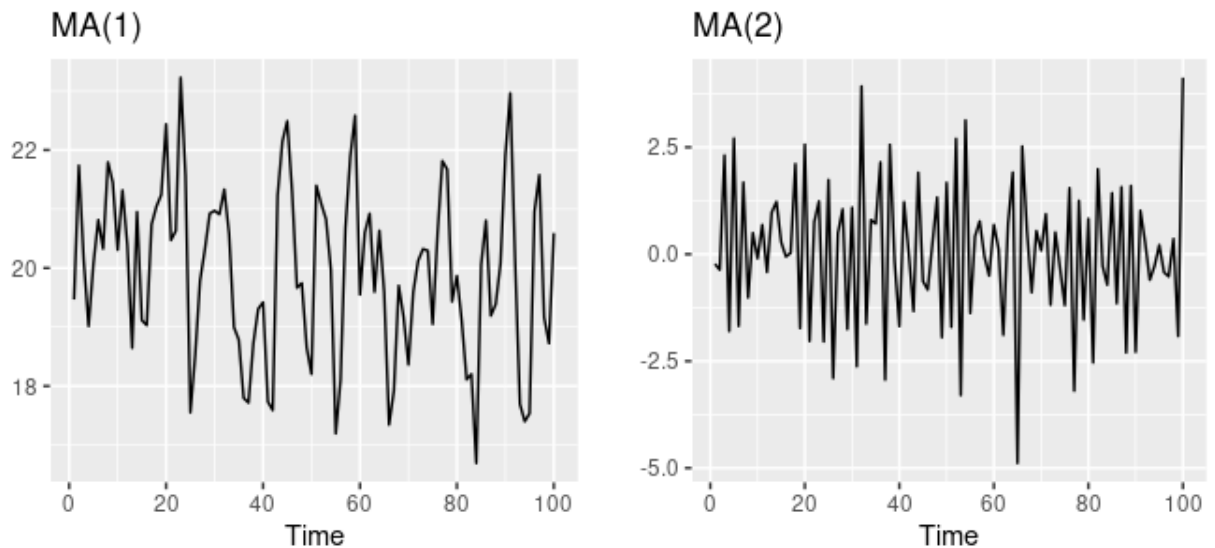


Рисунок 2.8 – Два приклади даних з моделей ковзного середнього із різними параметрами

Можна записати будь-яку стаціонарну AR(p) модель як MA( $\infty$ ) модель:

$$\begin{aligned} y_t &= \varphi_1 y_{t-1} + \varepsilon_t = \varphi_1(\varphi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \varphi_1^2 y_{t-2} + \varphi_1 \varepsilon_{t-1} + \varepsilon_t = \\ &= \varphi_1^3 y_{t-3} + \varphi_1^2 \varepsilon_{t-2} + \varphi_1 \varepsilon_{t-1} + \varepsilon_t = \text{і т. д.} \end{aligned} \quad (2.29)$$

Вважаючи, що  $-1 < \varphi_1 < 1$ , значення  $\varphi_1^k$  спадатиме, якщо  $k$  зростатиме. Тож зрештою ми отримуємо:

$$y_t = \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_1^2 \varepsilon_{t-2} + \varphi_1^3 \varepsilon_{t-3} + \dots \quad (2.30)$$

А ця формула якраз і описує MA( $\infty$ ) модель.

Зворотний результат справедливий, якщо ми накладаємо деякі обмеження на параметри MA. Тоді модель MA називається зворотною. Тобто ми можемо

записати будь-який зворотній  $MA(q)$  процес як  $AR(\infty)$  процес. Зворотні моделі не просто вводяться, щоб дозволити нам перейти з моделей  $MA$  на моделі  $AR$ . Вони також мають деякі бажані математичні властивості.

Наприклад, розглянемо  $MA(1)$  процес,  $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$ . У його  $AR(\infty)$  репрезентації, остання помилка може бути записана як лінійна функція поточних та минулих спостережен:

$$\varepsilon_t = \sum_{j=0}^{\infty} (-\theta)^j y_{t-j} \quad (2.31)$$

- Якщо  $|\theta| > 1$ , то ваги зростають із збільшенням лагів, тому чим віддаленіші спостереження, тим більший їх вплив на поточну похибку.
- Якщо  $|\theta| = 1$ , то ваги постійні за розміром, а віддалені спостереження мають той самий вплив, що й останні спостереження.
- Якщо  $|\theta| < 1$ , то найновіші спостереження мають більшу вагу, ніж спостереження з більш далекого минулого.

Перші дві умови не несуть ніякого сенсу, тож ми хочемо, щоб виконувалася остання третя умова, тобто процес є зворотнім, якщо  $|\theta| < 1$ .

Обмеження зворотності для інших моделей подібні до обмежень стаціонарності, які ми накладала для моделей авторегресії.

- Для  $MA(1)$ :  $-1 < \theta_1 < 1$ .
- Для  $MA(2)$ :  $-1 < \theta_2 < 1$ ;  $\theta_1 + \theta_2 > -1$ ;  $\theta_1 - \theta_2 < 1$ .
- Для  $q \geq 3$  умови набагато складніші.

Якщо ми поєднаємо авторегресійну модель із моделлю ковзного середнього, ми отримаємо нову модель, яка носить назву  $ARMA$  (від англ. *AutoRegressive Moving Average*).

$$y_t = c + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (2.32)$$

У якості признаков у правій частині рівності використовуються як попередні значення  $y_t$ , так і відповідні помилки. Ми називаємо таку модель ARMA(p,q), де p – порядок авторегресії, а q – порядок ковзного середнього.

Пошук відповідних значень p та q в моделі ARMA(p,q) може бути полегшено шляхом побудови частинних автокореляційних функцій задля оцінки p, а також використання автокореляційних функцій задля оцінки q. Додаткову інформацію можливо підбирати, розглядаючи ті ж функції для залишків моделі, пристосованої початковим вибором p та q.

Моделі ARMA після вибору p та q загалом можуть бути пристосовані за допомогою регресії найменших квадратів задля знаходження значень параметрів, які мінімізують член похибки. Загалом доброю практикою вважають знаходити найменші значення p та q, які забезпечують прийнятну пристосованість до даних [3].

## Висновки

Моделі дискретного вибору використовуються для пояснення або прогнозування вибору з множини двох або більше дискретних, тобто взаємовиключних, альтернатив.

Лінійна регресія дає нам неперервний результат, у той час як логістична регресія забезпечує дискретний. Прикладом неперервного виходу є ціна на оренду приміщення. Прикладом ж дискретного виходу – це прогнозування того, чи буде дане приміщення користуватися попитом (те, що ми й розглядали раніше й будемо розглядати більш детально на реальних даних у 3 розділі).

Для побудови логістичної регресії, ми вводимо екземпляри із деякою кількістю признаков і таргетом (те, що ми маємо прогнозувати). За допомогою градієнтного спуску ми мінімізуємо функцію втрат і знаходимо параметри нашої моделі для того, щоб робити подальші передбачення.

У моделі множинної регресії ми прогнозуємо змінну, використовуючи лінійну комбінацію деяких признаков. У моделі авторегресії ми прогнозуємо

змінну, використовуючи лінійну комбінацію минулих значень цієї змінної. Термін авторегресія вказує на те, що це регресія є змінною щодо себе. Моделі із ковзним середнім замість того, щоб використовувати минулі значення змінної прогнозу в регресії, використовують минулі помилки прогнозу в регресійній моделі. Якщо ми поєднаємо авторегресійну модель із моделлю ковзного середнього, ми отримаємо нову модель, яка носить назву ARMA.

## РОЗДІЛ 3 ЗАСТОСУВАННЯ МОДЕЛЕЙ НА РЕАЛЬНИХ ДАНИХ ТРЦ “ВЕЛЕС” ДЛЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ЦІН НА ОРЕНДУ

### 3.1 Опис використовуваних даних ТРЦ “Велес”

Торгово-розважальний центр “Велес” – один із найбільших торгово-розважальних центрів Івано-Франківська, який знаходиться в одному із великих спальних районів, що за 2,5 км від центра міста, за адресою вулиця Вовчинецька, 225а. Загальна площа ТРЦ Велес складає близько 20000 кв. м. Для зручності покупців автомобільний паркінг вміщує 350 паркомісць. Бутікова галерея розташована на першому й другому поверхах, а різного роду розваги, включаючи кінотеатр, боулінг, роллердром, настільний теніс, дитячий лабіринт, перегони на мотоциклах та спортивних авто, альпін-парк, на третьому. Також на іншому куті третього поверху розташовуються офісні приміщення [30].

Використовуючи дані про оренду приміщень за 2012-2020 рр., у даному розділі я спробую побудувати модель дискретного вибору для прогнозування попиту на приміщення, відштовхуючись від типу приміщення, його розташування в ТРЦ, площі та ціни за 1 кв. м., а також модель ARIMA для прогнозування ціни на оренду для певного приміщення, враховуючи поквартальну ціну на оренду цього приміщення протягом останніх 9 років.

### 3.2 Застосування моделі дискретного вибору для аналізу цін на оренду й прогнозування попиту приміщень

У цьому підрозділі я буду використовувати дані, які містять інформацію про кожне приміщення ТРЦ, враховуючи його лот (деякий порядковий номер, який ми потім використаємо для того, що дістати нові признаки для нашої моделі), контрагент (тип приміщення), площу, середню ціну на оренду за 1 кв.м. за 2020 рік та найголовніше – попит (бінарна змінна, яка показує, чи є попит на дане приміщення, чи ні).

Одразу ж варто зазначити, що саме ми розуміємо під “попитом”? У даному випадку попит – це змінна, яка показує, чи є попит на дане приміщення серед орендарів на ринку торгівельної нерухомості. Для отримання цієї змінної я передав дані про середню вартість оренди за 2020 рік експерту із ТРЦ “Велес”, який за допомогою свого досвіду й обізнаності проставив ці значення для кожного приміщення (лоту). Попит – це саме те, що ми хочемо прогнозувати, натренувавшись на даних, які нам передав експерт. Для чого нам це? Перш за все, маючи готову модель, яку ми натренуємо на даних за 2020 рік, ми зможемо прогнозувати попит для деякого приміщення, враховуючи його дані, скажімо, за наступний 2021 рік. Ціна на оренду в більшості випадків буде змінюватися. Також може змінитися орендар приміщення, а відповідно й змінитися тип цього приміщення. Наприклад, до цього певний лот орендувала маловідома ювелірна компанія й попит на це приміщення був доволі великий, а в наступному році це приміщення почав орендувати магазин брендового одягу “Nike”, і попит на це приміщення різко падає, оскільки мало хто може конкурувати із брендами світового класу. Також ми можемо використати дану модель для того, щоб порівнювати два приміщення за попитом між собою. Як відомо модель бінарного дискретного вибору, окрім самого повернутого класу, повертає ще й імовірність приналежності до цього класу. Тобто навіть якщо два приміщення користуються попитом (змінна попиту має значення 1), то значення ймовірностей приналежності до класу “1” можуть бути різними, скажімо, 0,55 та 0,93. А це означає, що приміщення із імовірністю 0,93 користується набагато більшим попитом, ніж приміщення із імовірністю 0,55.

Лот, контрагент, площа й ціна – це й будуть наші основні признаки, які ми будемо використовувати для прогнозування попиту на приміщення. Таких змінних трішки малувато й хотілося б якось підв’язати розташування самого приміщення для прогнозування того, чи є на нього попит, чи ні. Для цього введемо нову змінну “Біля входу”. Ця змінна є бінарною із значеннями 0 та 1, де 0 характеризує те, що даний лот знаходиться не біля входу, а 1 – навпаки, біля входу. Чому саме така змінна? Інтуїтивно здається, що попит на



входу, як близькість до ескалатору, оскільки піднятися на 2 поверх відвідувачі ТРЦ можуть тільки так. На схемі ескалатор знаходиться посередині ближче до правого краю, поблизу лотів 203 (магазин “Мегаспорт”), 209 та 204/2. Тому вважатимемо, що лоти 203, 204/1, 204/2, 205, 206, 209, 211, 216, 218 та 219 знаходяться близько до входу (у даному випадку до ескалатору) й відповідно змінна “Біля входу” для даних лотів матиме значення 1, для решти лотів – значення 0.

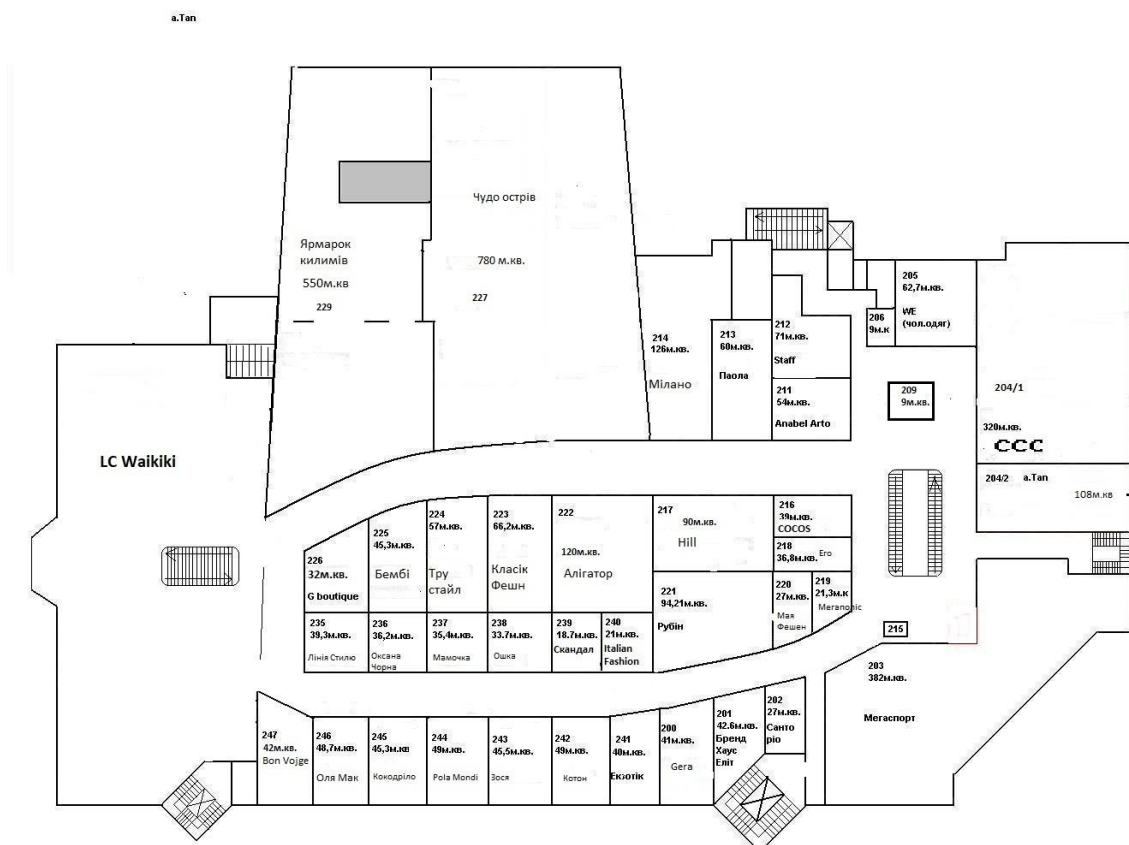


Рисунок 3.2 – Схема 2 поверху ТРЦ “Велес”

На рис. 3.3 зображена схема 3 поверху ТРЦ “Велес”. Аналогічно до 2 поверху, близькість до входу інтерпретуватимемо, як близькість до ескалатору, який розташований прямо по центру даної схеми поблизу лотів 307 (ролер клуб) та 314 (морозиво). Тому вважатимемо, що лоти 305-307, 313 та 314 знаходяться близько до входу (у даному випадку до ескалатору) й відповідно змінна “Біля входу” для даних лотів матиме значення 1, для решти лотів – значення 0.

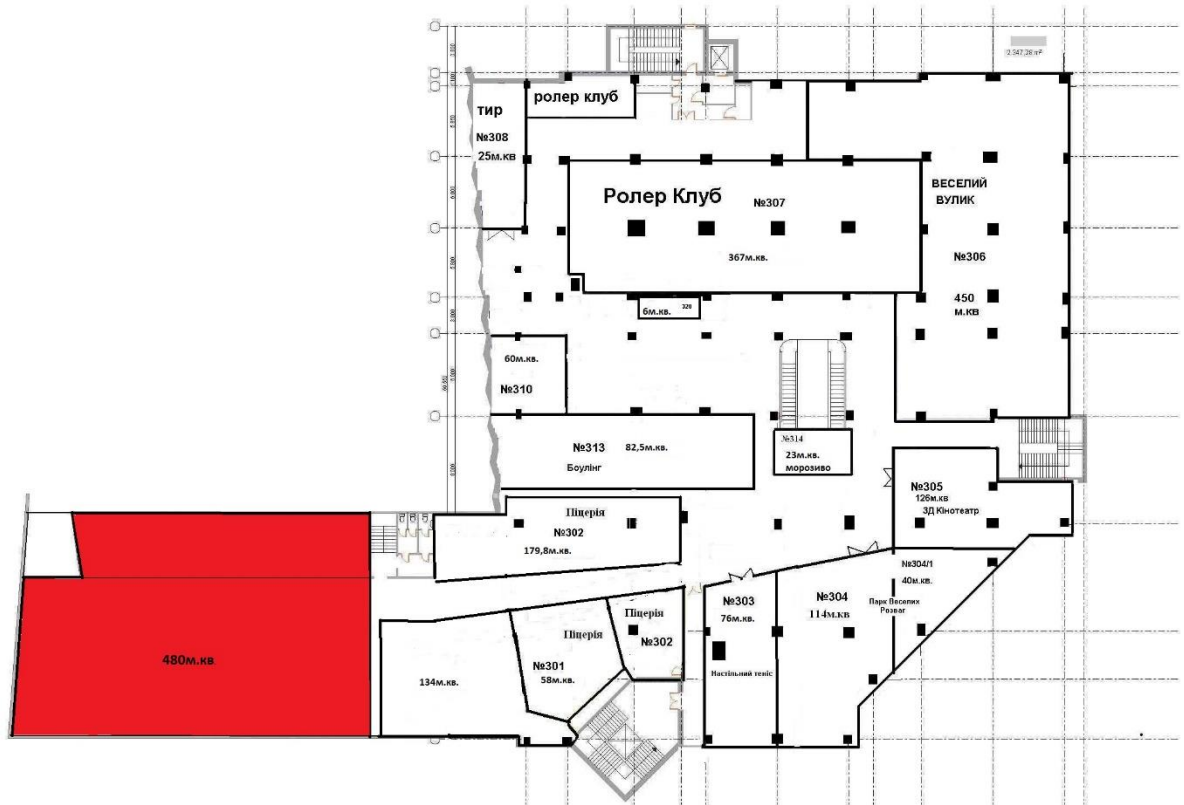


Рисунок 3.3 – Схема 3 поверху ТРЦ “Велес”

На рис. 3.4 зображена схема офісних приміщень. Тут важко щось судити, оскільки цих приміщень доволі небагато й інтуїтивно здається, що близькість до входу для цього типу приміщень не мала б сильно корелювати із попитом на це приміщення. Тому вважатимемо, що всі офісні приміщення знаходяться далеко від входу, тобто всі лоти мають значення 0 для змінної “Біля входу”.



```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 138 entries, 0 to 137
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  ---            -
0   Лот              138 non-null   object
1   Контрагент      138 non-null   object
2   Площа           138 non-null   float64
3   Ціна            138 non-null   float64
4   Біля входу     138 non-null   int64
5   Попит          138 non-null   int64
dtypes: float64(2), int64(2), object(2)
memory usage: 6.6+ KB

```

Рисунок 3.5 – Коротка інформація про дані.

Розглянемо детальніше признак “Лот”. Як уже зазначалося вище, дана змінна має тип “object”, що в більшості випадків означає текстовий тип даних. Взагалі, грубо кажучи, це індекс, який не мав би впливати на попит і яким у більшості аналогічних задач просто нехтують. Ми вже й так трішки раніше із цієї змінної витягнули ще одну змінну “Біля входу”. Проте якщо ми поглянемо детальніше на рисунки 3.1, 3.2 та 3.3, стане зрозуміло, що перша цифра кожного лоту характеризує поверх, на якому розташовується даний лот (дане приміщення). А це цілком підходить для того, щоб перетворити текстову змінну “Лот” у числову змінну, яка характеризуватиме на якому поверсі знаходиться дане приміщення. Варто відзначити, що правило, наведене вище, підходить на для всіх лотів. Є лоти, які починаються із слова “Офіс”. Проте ми знаємо, що всі офіси знаходяться на 3 поверсі на іншому крилі ТРЦ. Також варто відзначити окремі випадки: “3G інтернет” та “парковка”. “3G інтернет” також знаходиться на 3 поверсі, як і офіси, а от “парковка” знаходиться надворі, тобто ні на якому поверсі, тому значення цієї змінної конкретно для парковки становитиме 0.

Тепер розглянемо детальніше змінну “Контрагент”. На рис. 3.6 та 3.7 зображені частоти зустрічань унікальних значень цієї змінної. Як ми бачимо, унікальних значень доволі багато й дуже багато унікальних значень із частотою зустрічань, що дорівнює 1, тому було прийнято рішення закодувати (перевести

змінну із текстового типу даних у числовий) за допомогою лейбльованого енкодингу. Усе доволі просто: кожному унікальному текстовому значенню ми ставимо у відповідність деяке унікальне числове значення. Оскільки унікальних значень 51, то алгоритм закодує текстові дані числами від 0 до 50.

одяг	31
офіс	31
взуття	5
білизна	4
банкомат	4
Термінал поповнення	4
ювелірка	4
піцерія	3
моб.тел.	3
відділення банку	2
кафе	2
обмін валют	2
сумки	2
аксесуари	2
Оренда торгобладнання	2
посуд	2
годинники	1
Аптека	1
сигари, кальяни	1
оптика	1
одяг для вагітних	1
спорт товари	1
тир	1
Прод.супермаркет	1
настільний теніс	1
інтернет	1
дит.білизна	1
дит.розваги	1
моб.аксесуари	1
парфумерія	1
ролердром	1
косметика	1
ломбард	1
7Д кіно	1
боулінг	1

Рисунок 3.6 – Частоти зустрічань унікальних значень змінної “Контрагент”

(частина 1)

поштомат	1
морозиво	1
зоотовари	1
Парк веселих розваг	1
бістро	1
електрозаправка	1
хімчистка	1
побутова хімія	1
тур.фірма	1
продаж води	1
дом.текстиль	1
держ.лотереї	1
кінотеатр	1
дитячий супермаркет	1
біжутерія	1
килими	1

Рисунок 3.7 – Частоти зустрічань унікальних значень змінної “Контрагент”  
(частина 2)

Звісно, це не універсальний спосіб кодування даної текстової змінної. Можна було б якось пооб’єднювати деякі схожі типи в один, щоб зменшити кількість унікальних типів, скажімо, до 10. Наприклад, об’єднати типи “піцерія”, “морозиво” і “бістро” в один тип із назвою “їжа”, або типи “кінотеатр”, “дит. розваги”, “ролердром”, “настільний теніс” і т.д. у новий тип “розваги”. А потім можна було б використати OneHotEncoding – такий тип кодування, якому у відповідність ставиться не одне число, а вектор розмірності унікальних значень кодуваної змінної, де на певному місці даного вектора 1, а на інших – 0. Проте забігаючи наперед, тип енкодингу, який ми запропонували й використали вище теж непогано знаходить закономірності із змінною “Попит”, а модель показує доволі непогані результати.

Отож ми перетворили дві текстові змінні, які ми мали на початку, у числовий формат, із яким модель може дуже легко працювати. На рис. 3.8 зображений кінцевий датафрейм, який ми будемо подавати на вхід моделі і яка прогнозуватиме попит для приміщень.

	Лот	Контрагент	Площа	Ціна	Біля входу	Попит	
	0	1	38	41.00	488.29	0	0
	1	1	32	46.00	550.00	0	0
	2	1	10	33.00	650.00	0	0
	3	1	24	22.70	750.00	0	0
	4	1	38	35.15	510.00	0	0
	...	...	...	...	...	...	...
	133	3	35	32.00	80.00	0	1
	134	3	35	30.00	167.00	0	1
	135	3	35	32.00	100.00	0	1
	136	3	35	32.00	100.00	0	1
	137	0	20	2.00	50.00	0	0
138 rows × 6 columns							

Рисунок 3.8 – Кінцевий датафрейм після перетворення текстових змінних у числові

Тепер проаналізуємо, як наші признаки впливають на “Попит”. На рис. 3.9 зображена матриця кореляцій усіх змінних. Як ми бачимо, найбільше на “Попит” впливає “Лот”, тобто в даному випадку поверх. Їх кореляція становить приблизно 0,73, що означає, чим вищий поверх, тим більший попит на дане приміщення. Також велика кореляція між “Ціною” та “Попитом”, проте в даному випадку кореляція від’ємна, що означає, що вони обернено пропорційні, тобто із збільшенням ціни, попит буде меншим.

	Лот	Контрагент	Площа	Ціна	Біля входу	Попит
<b>Лот</b>	1.000000	0.311184	-0.037874	-0.538654	-0.219960	0.728155
<b>Контрагент</b>	0.311184	1.000000	-0.186089	-0.131815	-0.075975	0.283399
<b>Площа</b>	-0.037874	-0.186089	1.000000	-0.156599	0.172275	0.126910
<b>Ціна</b>	-0.538654	-0.131815	-0.156599	1.000000	0.289543	-0.567976
<b>Біля входу</b>	-0.219960	-0.075975	0.172275	0.289543	1.000000	-0.184831
<b>Попит</b>	0.728155	0.283399	0.126910	-0.567976	-0.184831	1.000000

Рисунок 3.9 – Матриця кореляцій кінцевого датафрейму

Наступний крок – розділити наші дані на тренувальні й тестові. Для тестового набору використовуватимемо 33% усіх даних, що в нас є. Після цієї операції ми отримуємо дві матриці із признаками: тренувальну ( $X_{train}$ ) та тестову ( $X_{test}$ ) та два вектори із таргетом, тобто те, що ми маємо передбачати: також тренувальний ( $y_{train}$ ) та тестовий ( $y_{test}$ ). При 33%, що йдуть на тестову вибірку, на рис. 3.10 зображені розмірності відповідних тренувальних та тестових матриць та векторів.

```
X_train shape: (92, 5)
X_test shape: (46, 5)
y_train shape: (92,)
y_test shape: (46,)
```

Рисунок 3.10 – Розмірності тренувальних та тестових матриць признаков та таргет-векторів

За допомогою тренувальних матриць признаков і таргет-векторів тренуємо модель дискретного вибору, а саме модель логістичної регресії. Варто також зазначити, що перед тим, як тренувати саму модель, ми стандартизуємо всі наші признаками, тобто віднімаємо середнє й ділимо на середньоквадратичне відхилення. Це робиться для того, щоб представити всі змінні на одному проміжку й для швидшої збіжності градієнтного спуску, який буде використовуватися для мінімізації функції втрат так, як описано в Розділі 2 цієї дипломної роботи.

На рис. 3.11 наведені метрики результатів натренованої моделі. Як ми бачимо, метрики представлені для тестового набору, який ми оприділили вище, і для деякого “cv”, про який ми нічого не говорили до цього. Отож метрики cv (cross-validation) – це середні метрики по валідаційних вибірках, ґрунтуючись на K-Fold розділенні [17]. Як це працює? Ми передаємо деяку навчальну вибірку для тренування моделі, проте замість того, щоб тренувати нашу модель на всіх цих даних, ми знову ділимо її на вже реально тренувальну й валідаційну, тренуємося на першій і перевіряємо роботу моделі на другій, підраховуючи метрики. Проводимо ми таку операцію K разів, саме тому метод і дістав назву K-Fold розділення. У нашому випадку  $K = 5$ . Варто відзначити, що

за цим алгоритмом кожне спостереження (у нашому випадку лот) має хоча б одного разу попасти у валідаційну вибірку. Метрики крос-валідації показують реальну спроможність моделі прогнозувати деяку змінну, на відміну від тестової, яка дуже залежить від того, як ми її відділимо на початку.

data	metrics	
cv	accuracy	0.859064
	roc_auc	0.931563
	precision	0.927778
	recall	0.853846
	f1	0.887019
test	accuracy	0.934783
	roc_auc	0.972917
	precision	0.965517
	recall	0.933333
	f1	0.949153

Рисунок 3.11 – Метрики натренованої логістичної регресії

Тепер трішки детальніше про кожну із метрик. Отож точність (accuracy) – це відношення кількості правильно розподілених таргетів до кількості таргетів всього [1]. Розраховується за наступною формулою:

$$accuracy(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \mathbb{I}(\hat{y}_i = y_i) \quad (3.1)$$

Площа під ROC-кривою (ROC-AUC) – площа, обмежена ROC-кривою і віссю частки помилкових позитивних класифікацій. ROC-крива – графік, що дозволяє оцінити якість бінарної класифікації, відображає співвідношення між часткою об'єктів від загальної кількості носіїв ознаки, правильно класифікованих до загальної кількості об'єктів, що не несуть ознаки, помилково класифікованих, як такі, що мають ознаку [14]. На рис 3.12 зображено ROC-криву для тестової вибірки натренованої логістичної регресії.

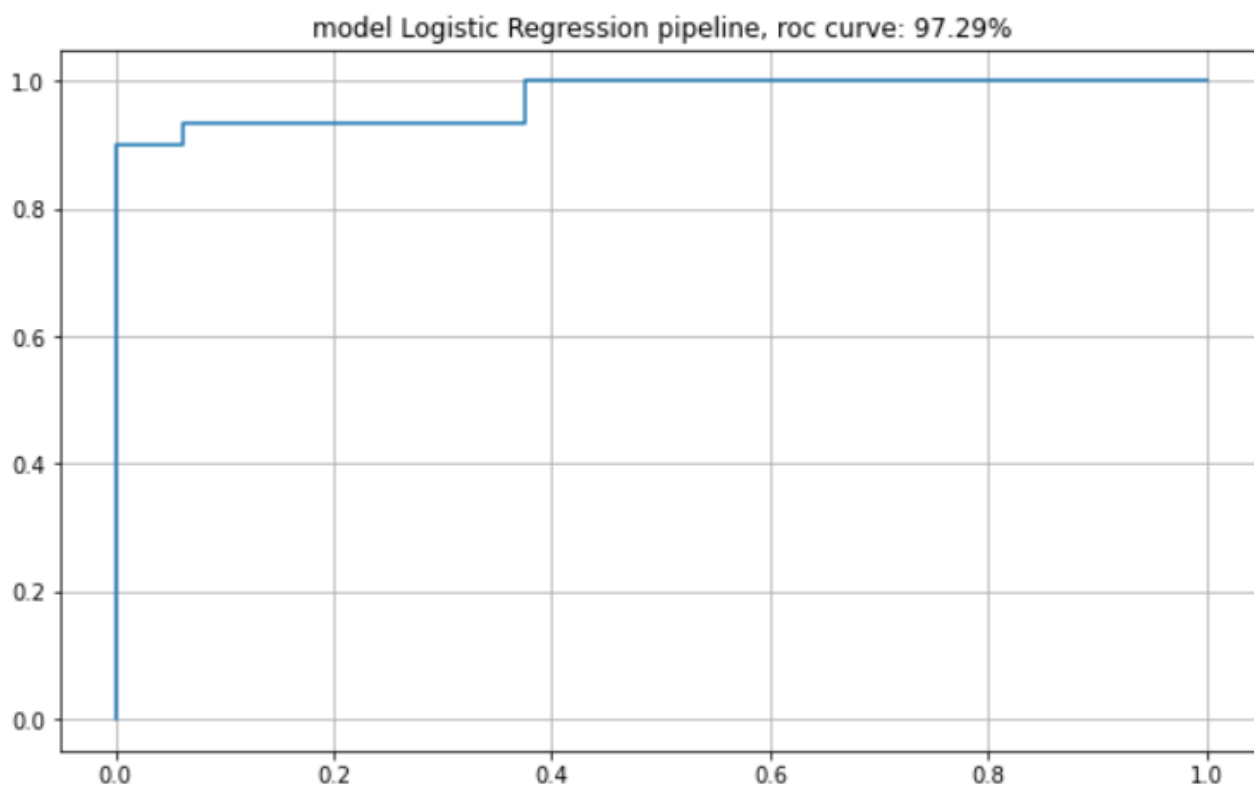


Рисунок 3.12 – ROC-крива для тестової вибірки

Перш, ніж перейти до precision, recall та F1 метрик, доцільно було б розповісти про матрицю невідповідностей. Отож матриця невідповідностей – це така матриця, яка складається із кількості істинно позитивних (TP від True Positive), істинно негативних (TN від True Negative), хибно позитивних (FP від False Positive) та хибно негативних (FN від False Negative) прокласифікованих таргетів [28]. У табл. 3.1 дане твердження представлено таблично.

Таблиця 3.1 – Представлення матриці невідповідності

		Справжній стан	
		Позитивний стан	Негативний стан
Загальна сукупність			
Прогнозований стан	Позитивний прогнозований стан	Істинно позитивний	Хибно позитивний, помилка 1 роду
	Негативний прогнозований стан	Хибно негативний, помилка 2 роду	Істинно негативний

На рис. 3.13 зображено матрицю невідповідності для тестової вибірки для натренованої логістичної регресії.

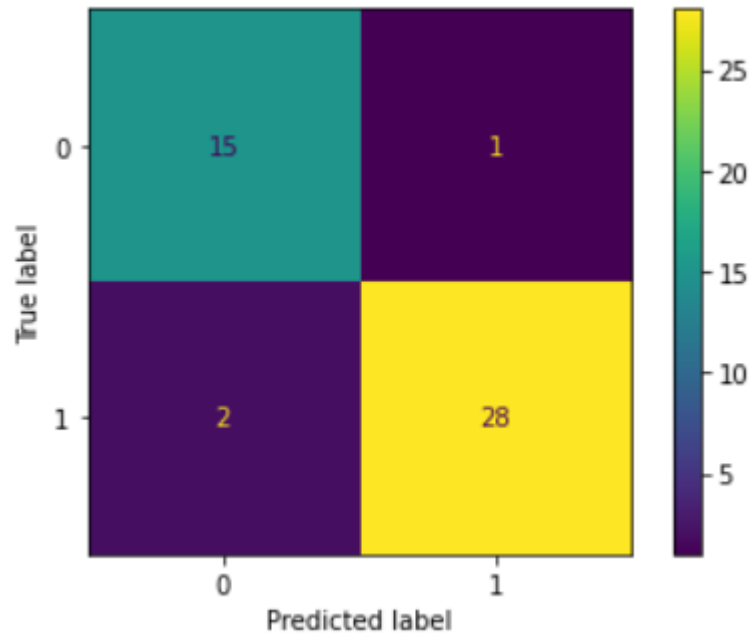


Рисунок 3.13 – Матриця невідповідностей для тестової вибірки

Відповідно, *precision* – це відношення істинно позитивних (TP) до всіх позитивно прокласифікованих [13]:

$$precision = \frac{TP}{TP + FP} \quad (3.2)$$

А *recall* – це відношення істинно позитивних (TP) до всіх дійсно позитивних [13]:

$$recall = \frac{TP}{TP + FN} \quad (3.3)$$

І F1 – середнє гармонійне [7] метрик *precision* та *recall* [5]:

$$F_1 = 2 * \frac{precision * recall}{precision + recall} = \frac{2TP}{2TP + FP + FN} \quad (3.4)$$

Повертуючись до рис. 3.11, ми бачимо, що для крос-валідаційної вибірки всі метрики більші, ніж 85%, що доволі непоганий результат, враховуючи невеликий початковий датасет (лише 138 лотів). Для тестової вибірки всі метрики й взагалі більші, ніж 93%, що означає доволі хорошу спроможність натренованої моделі прогнозувати попит для нових лотів.

Як уже описувалося вище, дана модель може бути використана для того, щоб порівнювати попити різних приміщень між собою, взявши до уваги ймовірності приналежності, які видає модель для класу 1. На рис. 3.14 зображено всі початкові дані із новим стовпчиком “Попит %”, який показує відсоток того, на скільки сильний попит існує на те, чи інше приміщення. Як ми бачимо, модель вважає, що лот №1 має набагато більший попит, ніж лот №103, проте обидва ці лоти характеризуються 0 у колонці “Попит”.

	Лот	Контрагент	Площа	Ціна	Біля входу	Попит	Попит %
0	101	посуд	41.00	488.29	0	0	47.72
1	102	одяг	46.00	550.00	0	0	38.74
2	103	білизна	33.00	650.00	0	0	16.50
3	104	косметика	22.70	750.00	0	0	23.01
4	104/1	посуд	35.15	510.00	0	0	46.51
...	...	...	...	...	...	...	...
133	Офіс №8/1	офіс	32.00	80.00	0	1	99.08
134	Офіс №9	офіс	30.00	167.00	0	1	98.93
135	Офіс №9/1	офіс	32.00	100.00	0	1	99.05
136	Офіс №9/2	офіс	32.00	100.00	0	1	99.05
137	парковка	електрозаправка	2.00	50.00	0	0	9.74

138 rows × 7 columns

Рисунок 3.14 – Початкові дані із добавленою колонкою “Попит %”

Отже, ми змогли побудувати доволі хорошу модель дискретного вибору, яку можна використовувати як для прогнозування того, чи є попит на дане приміщення для нових лотів, так і для того, щоб порівнювати значення попитів для різних приміщень між собою.

### 3.3 Застосування моделі ARIMA для прогнозування цін на оренду

У цьому підрозділі я використовуватиму дані, які містять інформацію щодо квартальних цін на оренду різних приміщень (лотів) протягом останніх 9 років (2012-2020 рр.). Також для кожного приміщення вказаний контрагент (тип приміщення) та площа. Варто відзначити, що будувати ARIMA моделі ми будемо для кожного приміщення окремо, базуючись на конкретно його цінах за попередні 9 років.

Проведемо детальний аналіз для лоту №236. Цей лот – магазин одягу, який розташовується на 2 поверсі ТРЦ “Велес” досить далеко від основного ескалатору. На рис. 3.15 зображено графік ціни в залежності від дати для даного лоту.

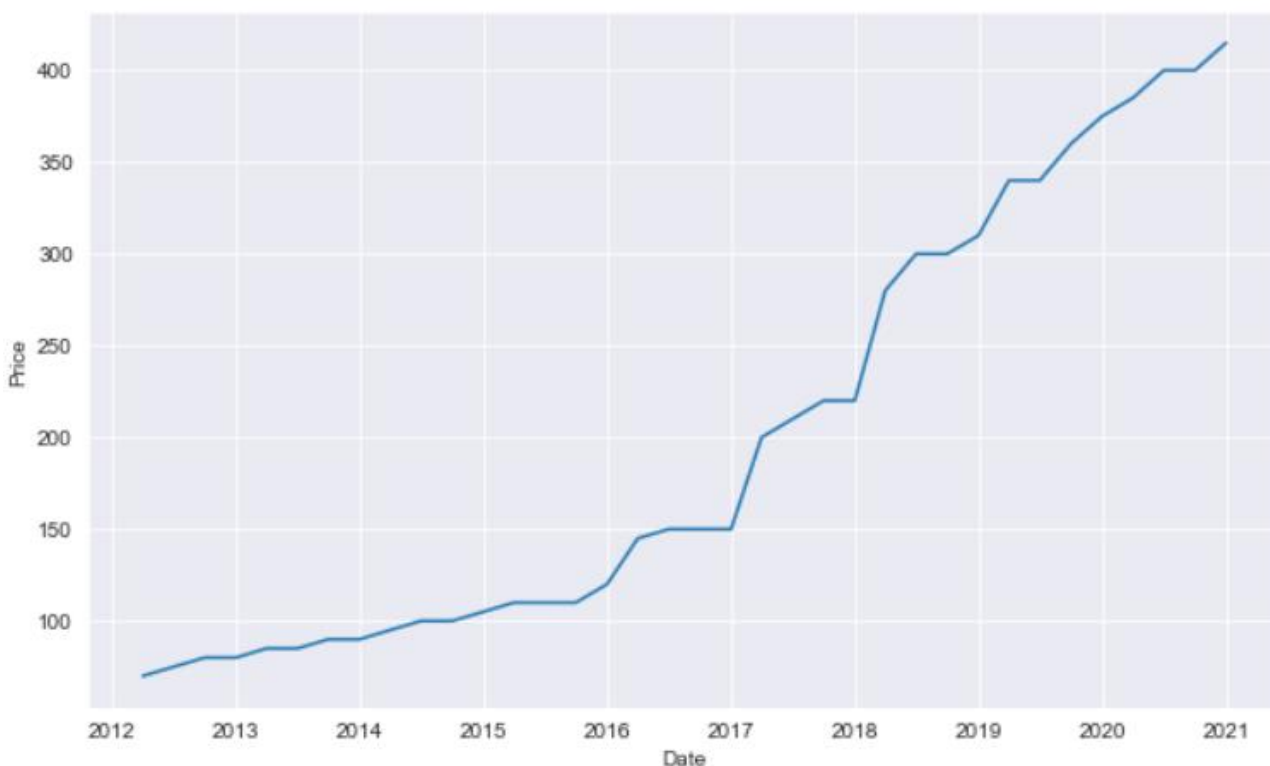


Рисунок 3.15 – Графік залежності ціни від дати для лоту №236

Не складно замітити, що даний процес – процес із висхідним трендом, тобто він є нестационарним. Що взагалі означає, що процес – стаціонарний? По-перше, середнє значення ряду не повинно бути функцією від часу. На рис. 3.16 червоний графік праворуч не є стаціонарним, тому що його середнє збільшується із плином часу. По-друге, дисперсія ряду також не повинна бути

функцією від часу. Ця властивість відома як гомоскедантичність. На рис. 3.17 червоний графік має різний розподіл даних у часі, що означає його нестаціонарність. По-третє, коваріація  $i$ -того та  $(i+m)$ -того члена не повинна бути функцією від часу. На рис. 3.18 червоний графік праворуч не є стаціонарним саме через те, що його коваріація не є сталою з плином часу.

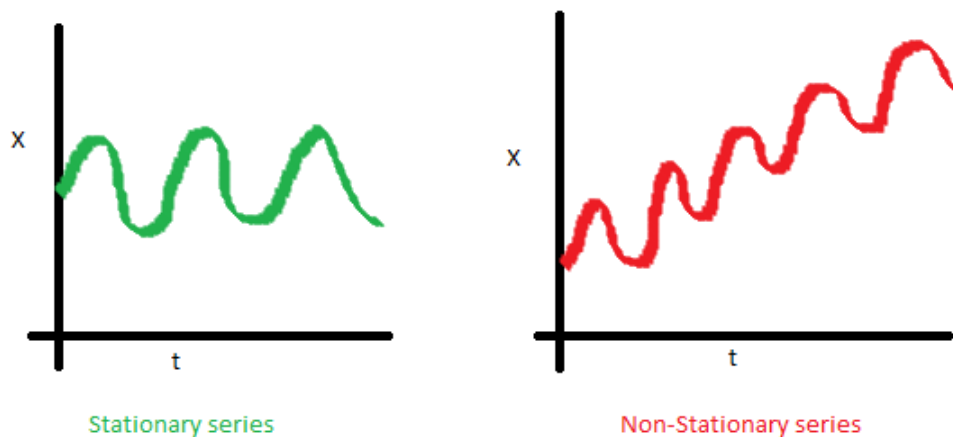


Рисунок 3.16 – Приклад ряду із не сталим середнім значенням

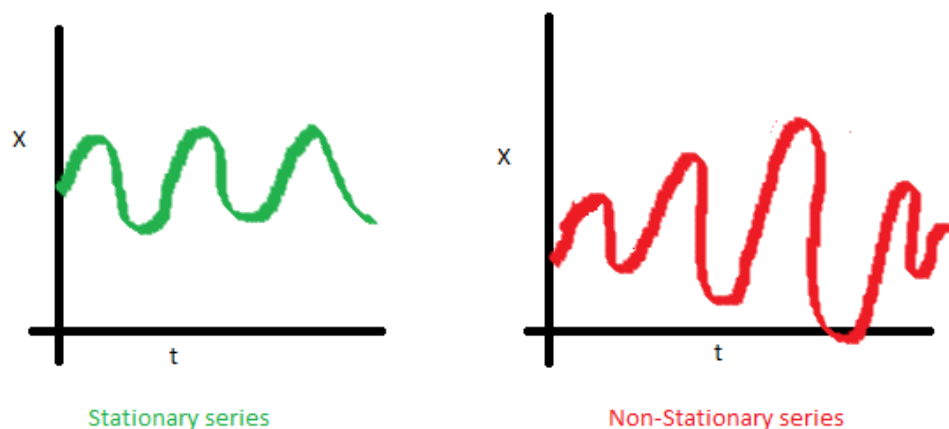


Рисунок 3.17 – Приклад ряду із не сталою дисперсією

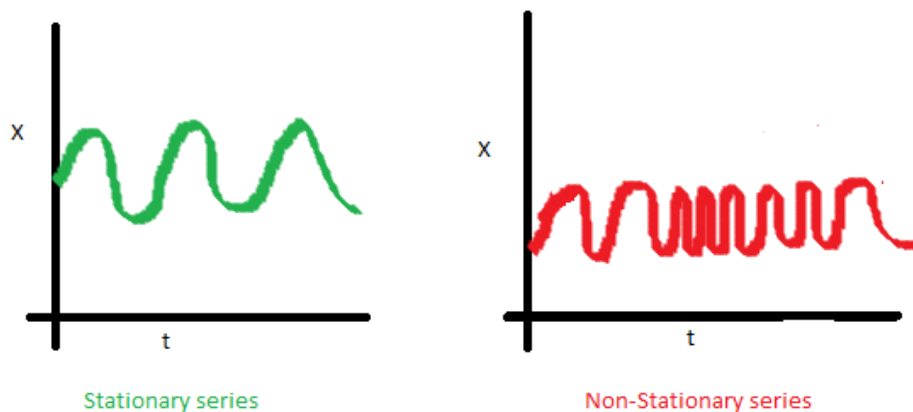


Рисунок 3.18 – Приклад ряду із не сталою коваріацією

Чому це важливо? При проведенні лінійної регресії припускають, що всі спостереження не залежать одне від одного. Однак у часових рядах ми знаємо, що спостереження залежать від часу. Виявляється, багато результатів, які виконуються для незалежних випадкових величин (закон великих чисел і центральна гранична), мають місце для стаціонарних випадкових величин. Отже, роблячи дані нерухомими, ми можемо фактично застосувати методи регресії до цієї залежної від часу змінної. Існує два способи перевірити стаціонарність часового ряду. Перший – це перегляд даних. Візуалізуючи дані, було б легко визначити зміну середнього значення або варіацію даних. Для більш точної оцінки існує тест Дікі-Фуллера. Я не буду вдаватися до особливостей цього тесту, але якщо "Статистика тесту" перевищує деяке "Критичне значення", то такий часовий ряд буде стаціонарним. На рис. 3.19 зображено візуальну перевірку початкового часового ряду на стаціонарність. На рис. 3.20 зображено результати тесту Дікі-Фуллера. Як ми бачимо, даний часовий ряд є нестаціонарним за обома способами: для першого ми бачимо середнє значення, яке зростає із плином часу, для другого ж "Статистика тесту" близька до 1, що явно більше "Критичного значення", яке ми взяли рівним 0.01 для даних задач.

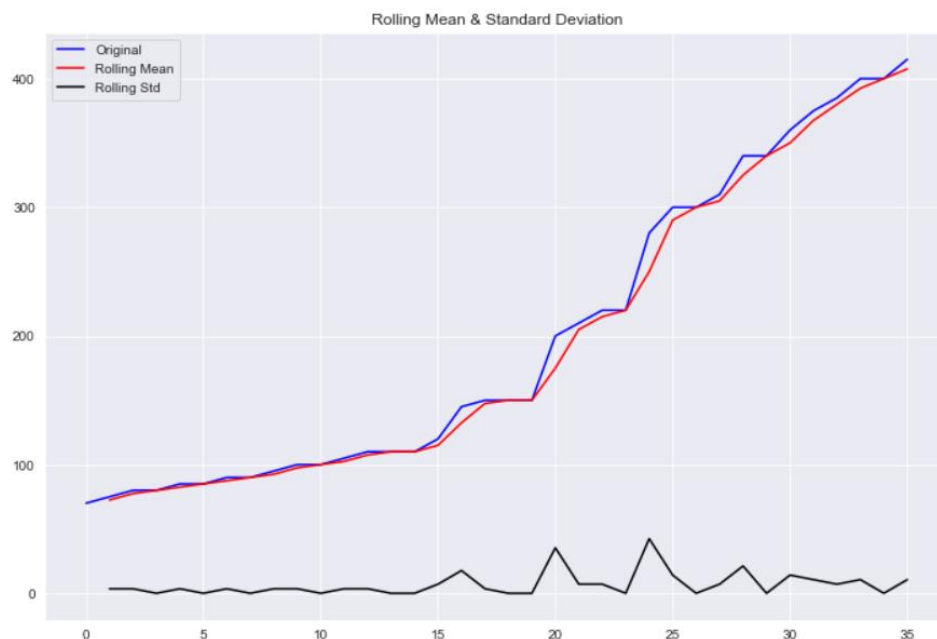


Рисунок 3.19 – Візуальний спосіб перевірки стаціонарності початкового часового ряду

```

Results of Dickey-Fuller Test:
p-value = 0.9970. The series is likely non-stationary.
Test Statistic           1.366959
p-value                   0.996959
#Lags Used                0.000000
Number of Observations Used 35.000000
Critical Value (1%)      -3.632743
Critical Value (5%)     -2.948510
Critical Value (10%)    -2.613017
dtype: float64

```

Рисунок 3.20 – Результати тесту Дікі-Фуллера для початкового часового ряду

На даному етапі для нас є дуже важливим отримати стаціонарний часовий ряд. Для цього візьмемо першу різницю (першу похідну) ряду й проведемо тести на стаціонарність ще раз. На рис. 3.21 зображено візуальний спосіб перевірки на стаціонарність. На рис. 3.22 зображено результати тесту Дікі-Фуллера для продиференційованого часового ряду. Як ми бачимо, за обома способами продиференційований часовий ряд є стаціонарним: середнє значення й дисперсія є незмінними із плином часу й “Статистика тесту” є набагато меншою, ніж “Критичне значення”.



Рисунок 3.21 – Візуальний спосіб перевірки стаціонарності продиференційованого часового ряду

```

Results of Dickey-Fuller Test:
p-value = 0.0000. The series is likely stationary.
Test Statistic          -5.789758e+00
p-value                 4.903475e-07
#Lags Used              0.000000e+00
Number of Observations Used  3.400000e+01
Critical Value (1%)      -3.639224e+00
Critical Value (5%)      -2.951230e+00
Critical Value (10%)     -2.614447e+00
dtype: float64

```

Рисунок 3.22 – Результати тесту Дікі-Фуллера для продиференційованого часового ряду

Наступний крок – побудова АКФ (автокореляційної функції) та ЧАКФ (частково автокореляційної функції). Часткова автокореляція при лагу  $k$  – це кореляція, яка виникає після усунення ефекту будь-яких кореляцій у термінах коротших лагів. Розглянемо часовий ряд, що був породжений процесом авторегресії (AR) із лагом  $k$ . АКФ описує автокореляцію між спостереженням та іншим спостереженням на попередньому етапі часу, який включає пряму та непряму інформацію про залежність. Це означає, що ми очікуємо, що АКФ для часового ряду  $AR(k)$  буде набувати великого значення до лагу  $k$ , а інерція цього зв'язку буде переходити до наступних значень відставання, відстаючи в якийсь момент у міру послаблення ефекту. Ми знаємо, що ЧАКФ описує лише прямий зв'язок між спостереженням та його відставанням. Це означає, що не буде кореляції для значень лагів понад  $k$ . Це саме те, що ми очікуємо від графіків АКФ та ЧАКФ для процесу  $AR(k)$ . Розглянемо часовий ряд, який був породжений процесом ковзного середнього (MA) із лагом  $k$ . Процес ковзного середнього – це модель авторегресії часового ряду залишкових помилок із попередніх прогнозів. Інший спосіб думати про модель ковзного середнього – це коригування майбутніх прогнозів на основі помилок, допущених за останніми прогнозами. Ми могли б очікувати, що показник АКФ для процесу  $MA(k)$  покаже сильну кореляцію з останніми значеннями до лагу  $k$ , потім різке зниження до низької або взагалі відсутності кореляції. За визначенням, саме так генерувався процес. Що стосується ЧАКФ, ми могли б очікувати, що сюжет покаже сильний зв'язок із відставанням та відставання кореляції від відставання

вперед. Знову ж, це саме те, що ми очікуємо від графіків АКФ та ЧАКФ для процесу  $MA(k)$ . На рис. 3.23 зображено графіки АКФ та ЧАКФ для початкового часового ряду. На рис 3.24 зображено графіки АКФ та ЧАКФ для продиферинційованого часового ряду.

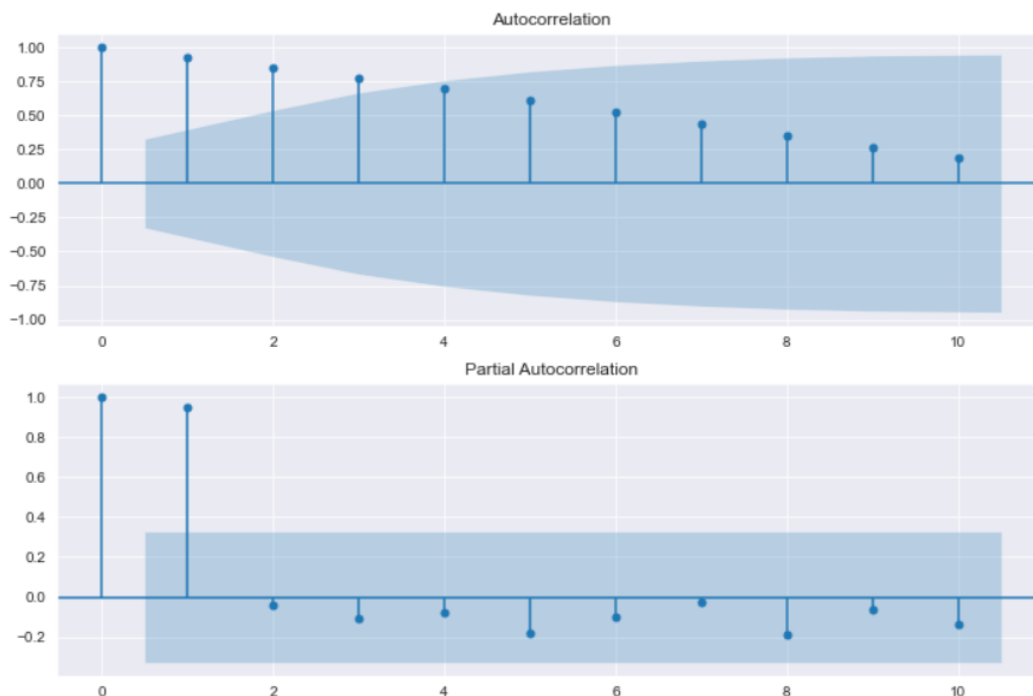


Рисунок 3.23 – АКФ та ЧАКФ для початкового часового ряду

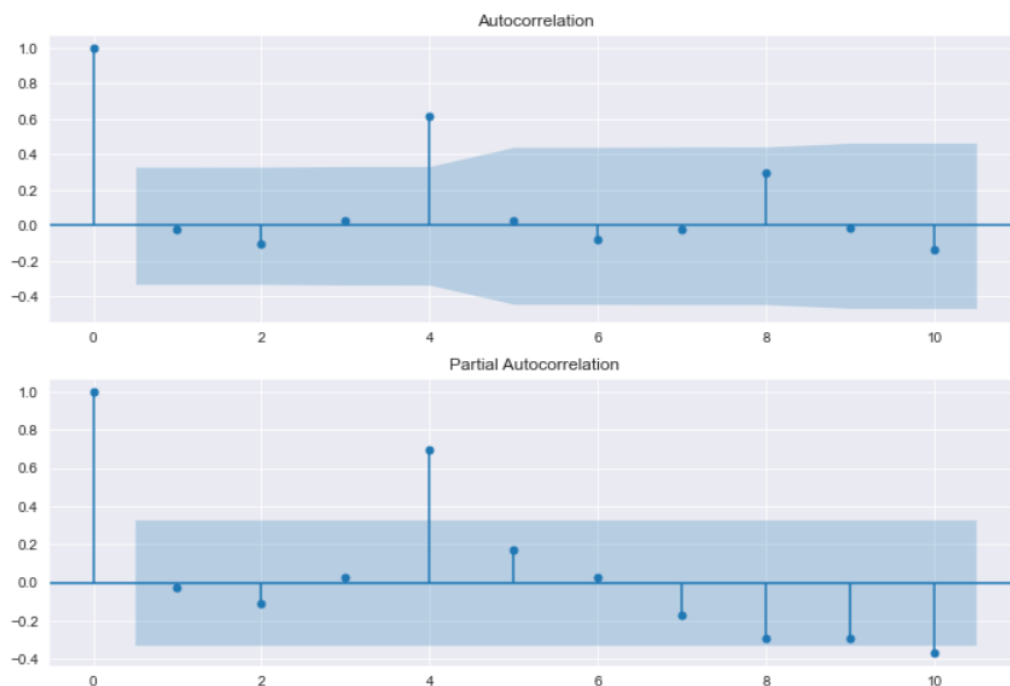


Рисунок 3.24 – АКФ та ЧАКФ для продиферинційованого часового ряду

Тепер можемо перейти безпосередньо до побудови ARIMA моделі, яка характеризується трьома параметрами:  $p$  – параметр авторегресійної (AR)

складової моделі,  $d$  – параметр диференціювання і  $q$  – параметр ковзного середнього (MA). Трішки вище ми перевіряли, що для першої різниці процес стає стаціонарним, тому покладемо  $d = 1$ . Параметр  $p$  можна дослідити із ЧАКФ, або шляхом емпіричного дослідження. У нашому випадку очевидно, що протягом 4 лагів AR є значним. Це означає, що ми можемо використати  $p = 4$ . Для даного випадку покладемо  $q$  рівним 1. Тобто побудуємо модель ARIMA(4,1,1). На рис. 3.25 зображений графік реальних значень ціни та спрогнозованих для останніх 7 входжень.



Рисунок 3.25 – Графік реальної й спрогнозованої ціни оренди

Як ми бачимо, модель доволі непогано відтворює динаміку зміни ціни й фінальну ціну передбачає майже точно. Стосовно метрик: MAPE = 3,35% і RMSE = 14,3.

MAPE визначається наступною формулою [11]:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3.5)$$

RMSE визначається наступною формулою [32]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (3.6)$$

Щоб побачити, як працює наша модель, ми можемо побудувати графік залишкового розподілу, а також АКФ та ЧАКФ залишкового розподілу. Для хорошої моделі графік залишкового розподілу має нагадувати криву нормального розподілу, а АКФ та ЧАКФ не мають мати значущих термінів. На рис. 3.26 зображений графік залишкового розподілу. На рис. 3.27 зображені графіки АКФ та ЧАКФ. Графік залишкового розподілу дійсно приблизно нагадує нормальний розподіл. Щодо АКФ та ЧАКФ, то вони не мають значущих термінів.

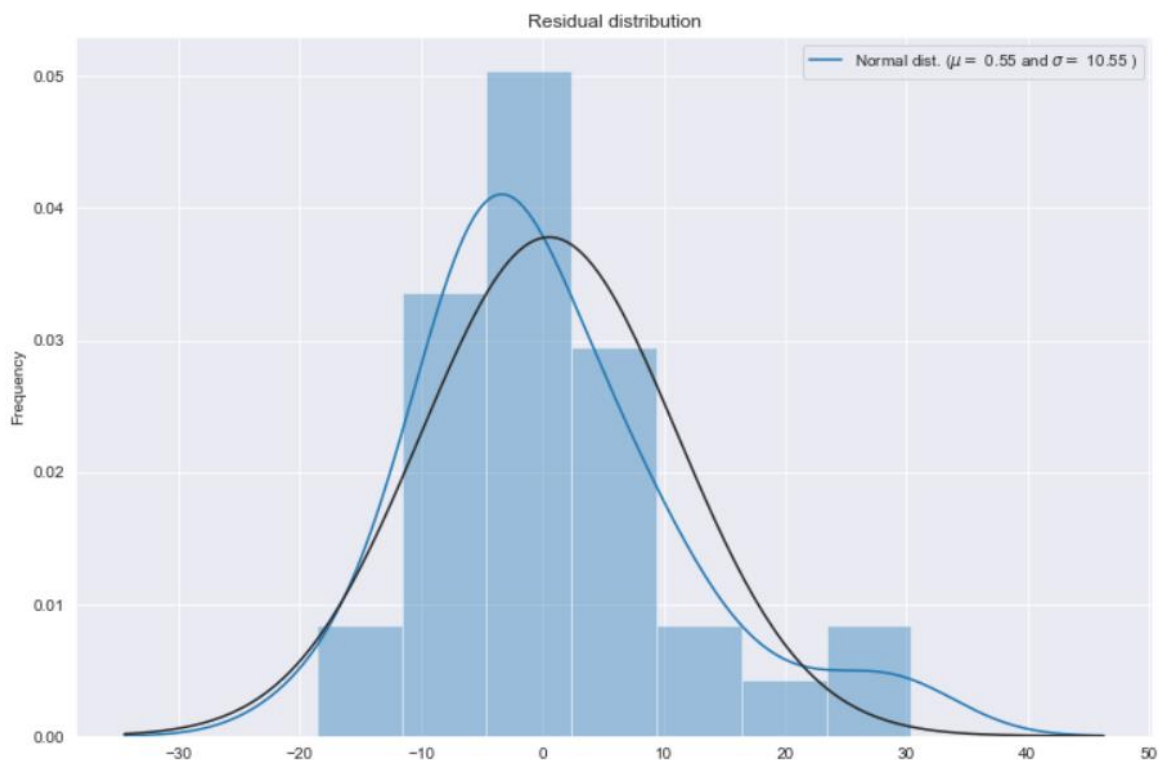


Рисунок 3.26 – Графік залишкового розподілу

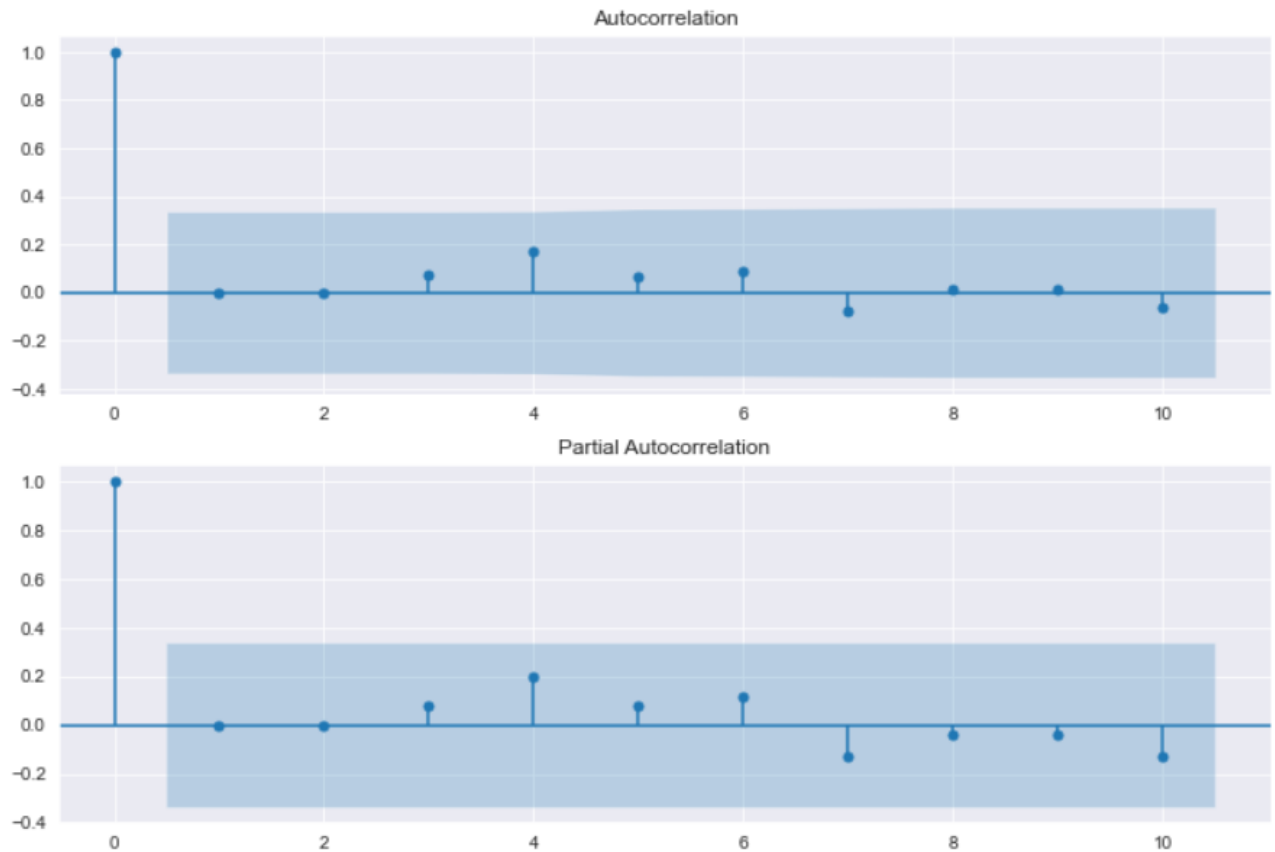


Рисунок 3.27 – Графіки АКФ та ЧАКФ для залишкового розподілу моделі

Отже, ми змогли побудувати модель, за допомогою якої можна прогнозувати ціну оренди деякого приміщення, базуючись на його цінах за попередні 9 років. Цікаво те, що, провівши низку експериментів для інших лотів, у більшості випадків найкращі результати дає модель ARIMA із параметрами  $p = 4$ ,  $d = 1$  та  $q = 1$ . На мою думку, це не співпадіння й таким чином простежується те, що ціна на оренду міняється в залежності від макроекономічних чинників і міняється одразу ж для всіх типів приміщень, не залежно від того, де вони розташовуються в ТРЦ.

### 3.4 Опис програмного продукту

Для обох моделей було розроблено кінцевий модуль мовою програмування Python у середовищі розробки Visual Studio Code і протестовано в середовищі розробки Jupyter Notebook.

Для моделі дискретного вибору (логістичної регресії) було розроблено доволі універсальний метод, за допомогою якого можна побудувати будь-яку модель (не тільки модель логістичної регресії, яку ми використовували в даній роботі) і отримати по ній метрики для початкового датасету із цінами на оренду, щоб прогнозувати попит на це приміщення.

Для прогнозування цін на оренду, використовуючи ціни за попередні 9 років, також було розроблено доволі універсальний модуль, де користувач сам може провести всі етапи побудови моделі й вибору параметрів, або ж скористатися готовою моделлю, передати дані й отримати прогноз на наступний рік.

## Висновки

У результаті цього розділу ми змогли побудувати дві моделі для аналізу й прогнозування вартості оренди торгівельних приміщень, базуючись на даних ТРЦ “Велес”. Варто відзначити, що модель дискретного вибору використовувалася саме для аналізу вартості, щоб прогнозувати попит на це приміщення. Модель ARIMA ж у свою чергу прогнозувала вартість орендної ціни деякого приміщення, базуючись на даних ціни за останні 9 років. Обидві моделі дають доволі непоганий результат: 93,5% точності моделі дискретного вибору на тестових даних та 3,34% MAPE для моделі ARIMA. Програмним продуктом обох моделей стали кінцеві модулі із різними корисними функціями, які користувач Python зможе із легкістю використовувати для своїх потреб.

## РОЗДІЛ 4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ

### 4.1 Постановка завдання

У даному розділі проводиться оцінка основних характеристик програмного продукту, розробленого для вирішення задач аналізу й прогнозування цін на оренду за допомогою моделі дискретного вибору та моделі ARIMA.

Нижче наведено аналіз різних варіантів реалізації модулю з метою вибору оптимальної, з огляду при цьому як на економічні фактори, так і на характеристики продукту, що впливають на продуктивність роботи і на його сумісність з апаратним забезпеченням. Для цього було використано апарат функціонально-вартісного аналізу.

Функціонально-вартісний аналіз (ФВА) – це технологія, яка дозволяє оцінити реальну вартість продукту або послуги незалежно від організаційної структури компанії. ФВА проводиться з метою виявлення резервів зниження витрат за рахунок ефективніших варіантів виробництва, кращого співвідношення між споживчою вартістю виробу та витратами на його виготовлення. Для проведення аналізу використовується економічна, технічна та конструкторська інформація.

Алгоритм функціонально-вартісного аналізу включає в себе визначення послідовності етапів розробки продукту, визначення повних витрат (річних) та кількості робочих часів, визначення джерел витрат та кінцевий розрахунок вартості програмного продукту.

У роботі застосовується метод ФВА для проведення техніко-економічного аналізу розробки системи аналізу й прогнозування цін на оренду. Оскільки рішення стосовно проектування та реалізації компонентів, що розробляється, впливають на всю систему, кожна окрема підсистема має її задовольняти. Тому фактичний аналіз представляє собою аналіз функцій

програмного продукту, призначеного для збору, обробки та проведення аналізу даних по цінах на оренду торгівельних приміщень.

Технічні вимоги до програмного продукту є наступні:

- функціонування на персональних комп'ютерах із стандартним набором компонентів;
- зручність та зрозумілість для користувача;
- швидкість обробки даних та доступ до інформації в реальному часі;
- можливість зручного масштабування та обслуговування;
- мінімальні витрати на впровадження програмного продукту.

#### 4.2 Обґрунтування функцій програмного продукту

Головна функція  $F_0$  – розробка програмного продукту, який вирішує задачу аналізу й прогнозування цін на оренду торгівельних приміщень. Беручи за основу цю функцію, можна виділити наступні:

$F_1$  – вибір мови програмування;

$F_2$  – вибір фреймворку машинного навчання;

$F_3$  – вибір середовища розробки.

Кожна з цих функцій має декілька варіантів реалізації:

Функція  $F_1$ :

а) Python

б) C++

Функція  $F_2$ :

а) Tensorflow;

б) Caffe

Функція  $F_3$ :

а) Jupyter Notebook;

б) Visual Studio.

Варіанти реалізації основних функцій наведені у морфологічній карті системи (рис. 4.1).

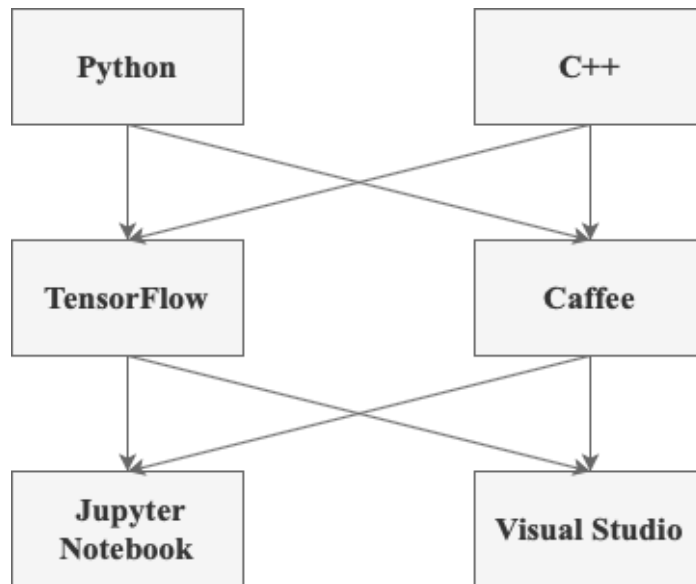


Рисунок 4.1 – Морфологічна карта

Морфологічна карта відображає множину всіх можливих варіанти основних функцій.

Таблиця 4.1 – Позитивно-негативна матриця

Функції	Варіанти реалізації	Переваги	Недоліки
$F_1$	<i>A</i>	Швидка розробка програми, доступність бібліотек, кросплатформеність	Низька швидкість роботи, особливо, якщо потрібно обробляти велику кількість даних
	<i>B</i>	Код швидко виконується	Іде багато часу на розробку програми
$F_2$	<i>A</i>	Надійно працює з складними проектами	Не підтримується багатьма мовами
	<i>B</i>	Надійність	Додатковий час на інсталяцію та вивчення

Продовження таблиці 4.1

$F_3$	<i>A</i>	Підтримується багатьма мовами програмування, легко запускається на будь-якому сервері	Відсутня можливість роботи без інтернету
	<i>B</i>	Багато інструментів, безпечна	Підтримує одночасно лише одну мову програмування

На основі цієї карти будемо позитивно-негативну матрицю варіантів основних функцій (Таб. 4.1). Робимо висновок, що при розробці програмного продукту деякі варіанти реалізації функцій варто відкинути, тому що вони не відповідають поставленим перед програмним продуктом задачам. Ці варіанти відзначені у морфологічній карті.

Функція  $F_1$ :

Перевагу віддаємо швидкості вивчення, простоті використання та наявності стандартних бібліотек для обчислення. Для спрощення роботи по написанню коду варіант Б має бути відкинтий.

Функція  $F_2$ :

Обидва варіанти можна використовувати в розробці.

Функція  $F_3$ :

Віддаємо перевагу варіанту А в разі вибору мови програмування Python.

Таким чином, будемо розглядати такі варіанти реалізації ПП:

$F_{1a} - F_{2a} - F_{3a}$

$F_{1a} - F_{2b} - F_{3a}$

Для оцінювання якості розглянутих функцій обрана система параметрів, описана нижче.

### 4.3 Обґрунтування системи параметрів ПП

На основі даних, розглянутих вище, визначаються основні параметри вибору, які будуть використані для розрахунку коефіцієнта технічного рівня.

Для того, щоб охарактеризувати програмний продукт, будемо використовувати наступні параметри:

- X1 – швидкодія мови програмування;
- X2 – об'єм пам'яті для обчислень та збереження даних;
- X3 – час навчання даних;
- X4 – потенційний об'єм програмного коду.

Гірші, середні і кращі значення параметрів вибираються на основі вимог замовника й умов, що характеризують експлуатацію ПП як показано у таблиці 4.2.

Таблиця 4.2 – Основні параметри ПП

Назва Параметра	Умовні позначення	Одиниці виміру	Значення параметра		
			гірші	середні	кращі
Швидкодія мови програмування	X1	оп/мс	10000	14000	19000
Об'єм пам'яті	X2	Мб	420	128	64
Час попередньої обробки даних	X3	мс	4	3	2
Потенційний об'єм програмного коду	X4	кількість рядків коду	4000	2500	1000

За даними таблиці 4.2 будуються графічні характеристики параметрів – рис. 4.2 – рис. 4.5.

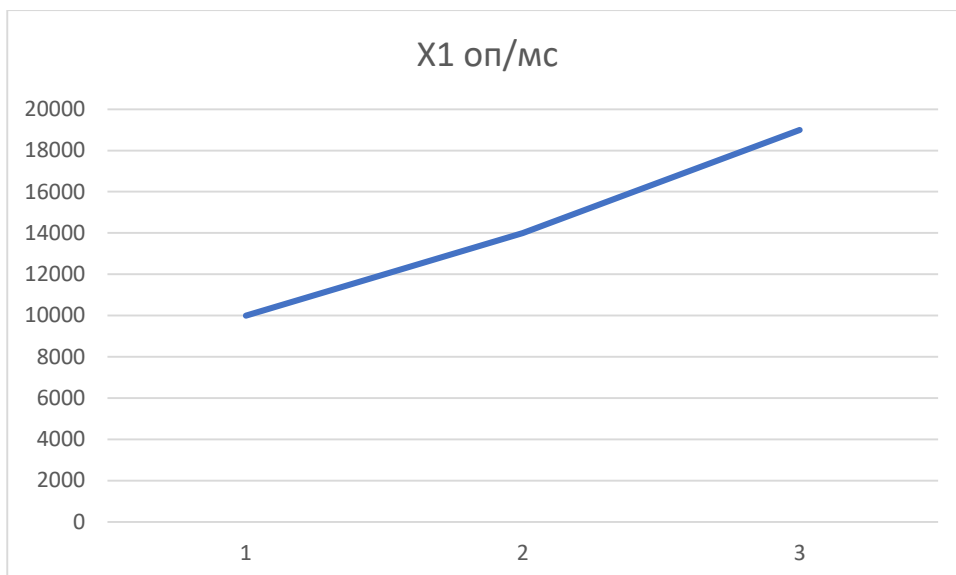


Рисунок 4.2 – X1, швидкодія мови програмування

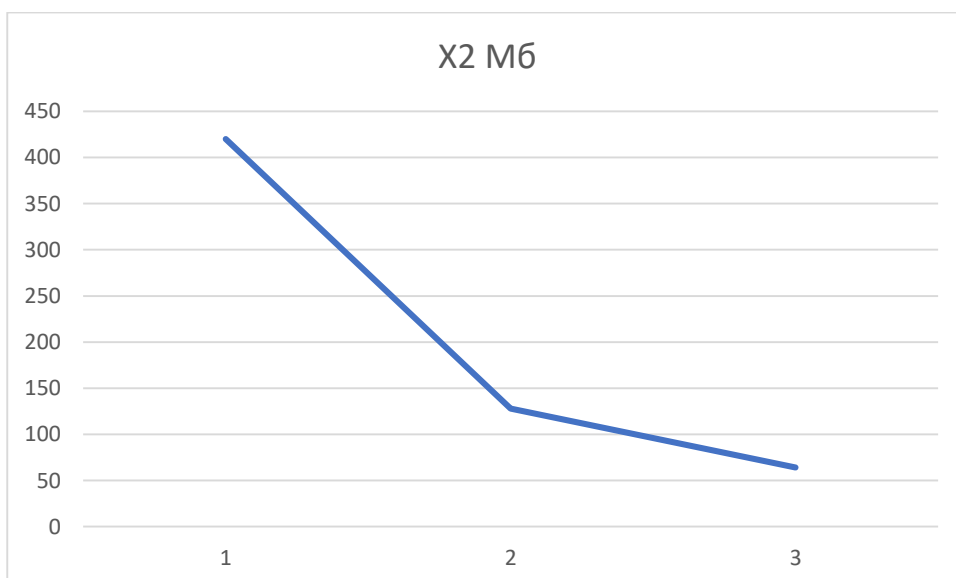


Рисунок 4.3 – X2, об'єм пам'яті

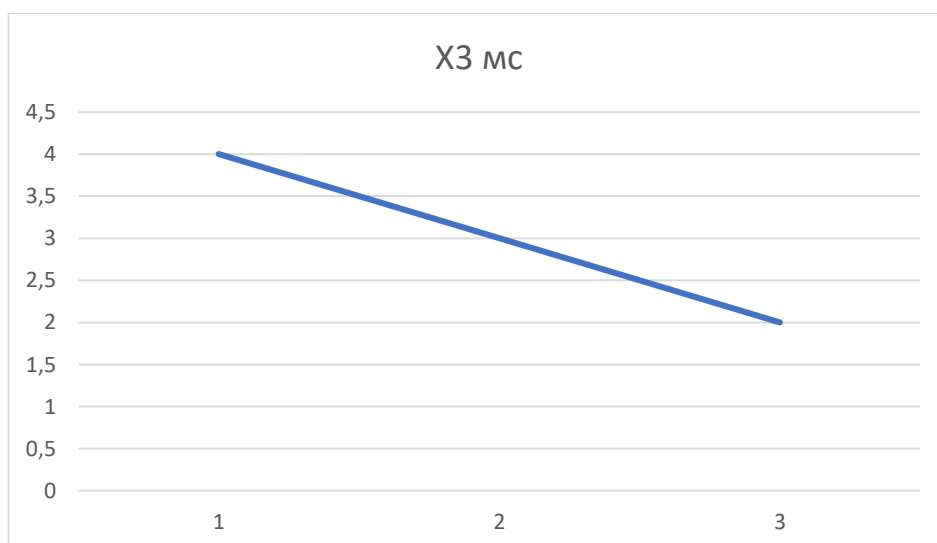


Рисунок 4.4 – X3, час попередньої обробки даних

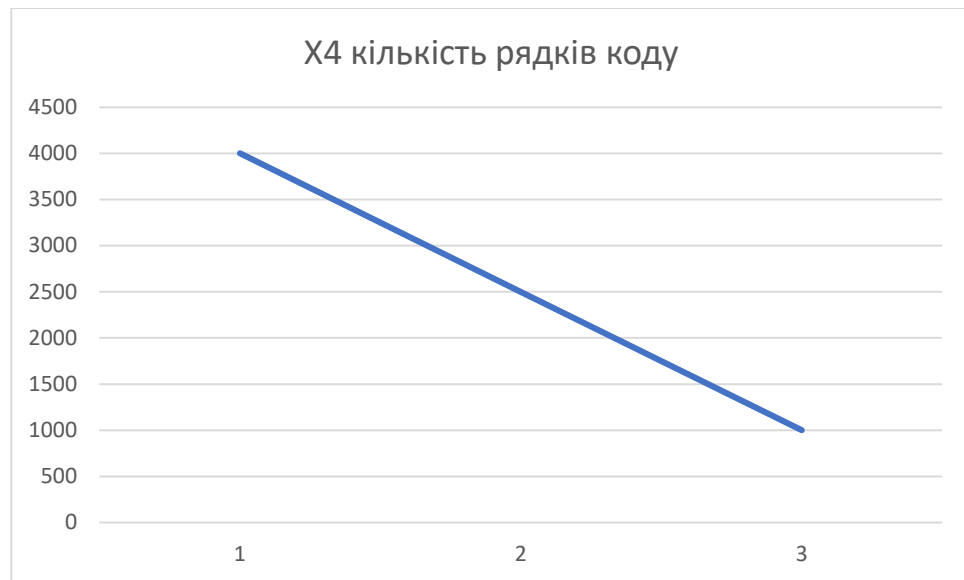


Рисунок 4.5 – X4, потенційний об'єм програмного коду

#### 4.4 Аналіз експертного оцінювання параметрів

Після детального обговорення й аналізу кожний експерт оцінює ступінь важливості кожного параметру для конкретно поставленої цілі – розробка програмного продукту, який дає найбільш точні результати при знаходженні параметрів моделей адаптивного прогнозування і обчислення прогнозних значень.

Значимість кожного параметра визначається методом попарного порівняння. Оцінку проводить експертна комісія із 7 людей. Визначення коефіцієнтів значимості передбачає:

- визначення рівня значимості параметра шляхом присвоєння різних рангів;
- перевірку придатності експертних оцінок для подальшого використання;
- визначення оцінки попарного пріоритету параметрів;
- обробку результатів та визначення коефіцієнту значимості.

Результати експертного ранжування наведені у таблиці 4.3.

Таблиця 4.3 – Результати ранжування параметрів

Позначення параметра	Назва параметра	Одиниці виміру	Ранг параметра за оцінкою експерта							Сума рангів $R_i$	Відхилення $\Delta_i$	$\Delta_i^2$
			1	2	3	4	5	6	7			
X1	Швидкодія мови програмування	Оп/мс	4	5	2	5	3	4	5	28	3,5	12,25
X2	Об'єм пам'яті	Мб	2	1	3	1	2	1	2	12	-12,5	156,25
X3	Час попередньої обробки даних	мс	5	3	5	5	4	5	3	30	5,5	30,25
X4	Потенційний об'єм програмного коду	Кількість рядків коду	3	5	4	3	5	4	4	28	3,5	12,25
	Разом		14	14	14	14	14	14	14	98	0	211

Для перевірки степені достовірності експертних оцінок, визначимо наступні параметри:

а) сума рангів кожного з параметрів і загальна сума рангів:

$$R_i = \sum_{j=1}^N r_{ij} R_{ij} = \frac{Nn(n+1)}{2} = 98, \quad (4.1)$$

де  $N$  – число експертів,

$n$  – кількість параметрів;

б) середня сума рангів:

$$T = \frac{1}{n} R_{ij} = 24,5 \quad (4.2)$$

в) відхилення суми рангів кожного параметра від середньої суми рангів:

$$\Delta_i = R_i - T. \quad (4.3)$$

Сума відхилень по всіх параметрах повинна дорівнювати 0;

г) загальна сума квадратів відхилення:

$$S = \sum_{i=1}^N \Delta_i^2 = 211. \quad (4.4)$$

Порахуємо коефіцієнт узгодженості:

$$W = \frac{12S}{N^2(n^3 - n)} = \frac{12 \cdot 211}{7^2(4^3 - 4)} = 0,86 > W_k = 0,67. \quad (4.5)$$

Ранжування можна вважати достовірним, тому що знайдений коефіцієнт узгодженості перевищує нормативний, котрий дорівнює 0,67.

Скориставшись результатами ранжирування, проведемо попарне порівняння всіх параметрів і результати занесемо у таблицю 4.4.

Таблиця 4.4 – Попарне порівняння параметрів

Параметри	Експерти							Кінцева оцінка	Числове значення
	1	2	3	4	5	6	7		
X1 і X2	<	=	<	<	<	>	>	<	0,5
X1 і X3	<	>	<	=	>	<	<	<	0,5
X1 і X4	>	<	<	=	<	=	>	<	0,5
X2 і X3	>	<	>	<	>	>	>	<	1,5
X2 і X4	>	>	<	=	>	>	<	<	1,5
X3 і X4	>	>	<	<	>	>	>	>	1,5

Числове значення, що визначає ступінь переваги  $i$ -го параметра над  $j$ -тим,  $a_{ij}$  визначається по формулі:

$$a_{ij} = \begin{cases} 1.5 \text{ при } X_i > X_j \\ 1.0 \text{ при } X_i = X_j \\ 0.5 \text{ при } X_i < X_j \end{cases} \quad (4.6)$$

З отриманих числових оцінок переваги складемо матрицю  $A = \| a_{ij} \|$ .

Для кожного параметра зробимо розрахунок вагомості  $K_{ei}$  за наступними формулами:

$$K_{Vi} = \frac{b_i}{\sum_{i=1}^n b_i} \quad (4.7)$$

$$b_i = \sum_{i=1}^N a_{ij} \quad (4.8)$$

Відносні оцінки розраховуються декілька разів доти, поки наступні значення не будуть незначно відрізнятись від попередніх (менше 2%). На другому і наступних кроках відносні оцінки розраховуються за наступними формулами:

$$K_{Vi} = \frac{b'_i}{\sum_{i=1}^n b'_i}, \quad (4.9)$$

$$b'_i = \sum_{i=1}^N a_{ij} b_j \quad (4.10)$$

Як видно з таблиці 4.5, різниця значень коефіцієнтів вагомості не перевищує 2%, тому більшої кількості ітерацій не потрібно.

Таблиця 4.5 – Розрахунок вагомості параметрів

Параметри $x_i$	Параметри $x_j$				Перша ітер.		Друга ітер.		Третя ітер.	
	X1	X2	X3	X4	$b_i$	$K_{Bi}$	$b_i^1$	$K_{Bi}^1$	$b_i^2$	$K_{Bi}^2$
X1	1,0	0,5	0,5	0,5	2,5	0,16	6,25	0,1	24,625	0,09
X2	0,5	1,0	1,5	1,5	4,5	0,28	20,25	0,3	88,125	0,3
X3	0,5	1,5	1,0	1,5	4,5	0,28	20,25	0,3	88,125	0,31
X4	0,5	1,5	1,5	1,0	4,5	0,28	20,25	0,3	88,125	0,3
Всього:					16	1	67	1	289	1

#### 4.5 Аналіз рівня якості варіантів реалізації функцій

Визначаємо рівень якості кожного варіанту виконання основних функцій окремо.

Абсолютні значення параметрів X2 (Об'єм пам'яті), X3 (час попередньої обробки даних) та X4 (потенційний об'єм програмного коду) відповідають технічним вимогам умов функціонування даного ПП.

Абсолютне значення параметра X1 (швидкість роботи мови програмування) обрано не найгіршим.

Коефіцієнт технічного рівня для кожного варіанта реалізації ПП розраховується так (таблиця 4.6):

$$K_K(j) = \sum_{i=1}^n K_{ei,j} B_{i,j}, \quad (4.11)$$

де  $n$  – кількість параметрів;

$K_{ei}$  – коефіцієнт вагомості  $i$ -го параметра;

$B_i$  – оцінка  $i$ -го параметра в балах.

Таблиця 4.6 – Розрахунок показників рівня якості варіантів реалізації основних функцій ПП

Основні функції	Варіант реалізації функції	Параметри	Абсолютне значення параметра	Бальна оцінка параметра	Коефіцієнт вагомості параметра	Коефіцієнт рівня якості
F1	A	X1	10000	6	0,09	0,54
F2	A	X2	64	5	0,3	1,5
	Б	X2	128	2	0,31	0,62
F3	A	X3	1000	8	0,3	2,4

За даними з таблиці 5.7 за формулою:

$$K_K = K_{TY}[F_{1k}] + K_{TY}[F_{2k}] + \dots + K_{TY}[F_{zk}], \quad (4.12)$$

Визначаємо рівень якості кожного з варіантів:

$$K_{K1} = 0,54 + 1,5 + 2,4 = 4,44,$$

$$K_{K2} = 0,54 + 0,62 + 2,4 = 3,56.$$

Як видно з розрахунків, кращим є перший варіант, для якого коефіцієнт технічного рівня має найбільше значення.

#### 4.6 Економічний аналіз варіантів розробки ПП

Для визначення вартості розробки ПП спочатку проведемо розрахунок трудомісткості.

Всі варіанти включають в себе два окремих завдання:

1. Розробка проекту програмного продукту;
2. Розробка програмної оболонки;

Завдання 1 за ступенем новизни відноситься до групи А, завдання 2 – до групи Б. За складністю алгоритми, які використовуються в завданні 1 належать до групи 1; а в завданні 2 – до групи 3.

Для реалізації завдання 1 використовується довідкова інформація, а завдання 2 використовує інформацію у вигляді даних.

Проведемо розрахунок норм часу на розробку та програмування для кожного з завдань.

Загальна трудомісткість обчислюється як

$$T_0 = T_P \cdot K_{\Pi} \cdot K_{СК} \cdot K_M \cdot K_{СТ} \cdot K_{СТ.М}, \quad (4.13)$$

де  $T_P$  – трудомісткість розробки ПП;

$K_{\Pi}$  – поправочний коефіцієнт;

$K_{СК}$  – коефіцієнт на складність вхідної інформації;

$K_M$  – коефіцієнт рівня мови програмування;

$K_{СТ}$  – коефіцієнт використання стандартних модулів і прикладних програм;

$K_{СТ.М}$  – коефіцієнт стандартного математичного забезпечення.

Для першого завдання, виходячи із норм часу для завдань розрахункового характеру ступеню новизни А та групи складності алгоритму 1, трудомісткість дорівнює:  $T_P = 90$  людино-днів. Поправочний коефіцієнт, який враховує вид нормативно-довідкової інформації для першого завдання:  $K_{\Pi} = 1.7$ . Поправочний коефіцієнт, який враховує складність контролю вхідної та вихідної інформації для всіх семи завдань рівний 1:  $K_{СК} = 1$ . Оскільки при розробці першого завдання використовуються стандартні модулі, врахуємо це за допомогою коефіцієнта  $K_{СТ} = 0.8$ . Тоді загальна трудомісткість програмування першого завдання дорівнює:

$$T_1 = 90 \cdot 1.7 \cdot 0.8 = 122.4 \text{ людино-днів.}$$

Проведемо аналогічні розрахунки для подальших завдань.

Для другого завдання (використовується алгоритм третьої групи складності, степінь новизни Б), тобто  $T_p = 27$  людино-днів,  $K_{II} = 0.9$ ,  $K_{СК} = 1$ ,  $K_{СТ} = 0.8$ :

$$T_2 = 27 \cdot 0.9 \cdot 0.8 = 19.44 \text{ людино-днів.}$$

Складаємо трудомісткість відповідних завдань для кожного з обраних варіантів реалізації програми, щоб отримати їх трудомісткість:

$$T_I = (122.4 + 19.44 + 4.8 + 19.44) \cdot 8 = 1328,64 \text{ людино-годин.}$$

$$T_{II} = (122.4 + 19.44 + 6.91 + 19.44) \cdot 8 = 1345,52 \text{ людино-годин.}$$

Найбільш високу трудомісткість має варіант II.

У розробці беруть участь два програмісти з окладом 9000 грн., один аналітик в області даних з окладом 14000. Визначимо середню зарплату за годину за формулою:

$$C_{ч} = \frac{M}{T_m \cdot t} \text{ грн.,} \quad (4.14)$$

де  $M$  – місячний оклад працівників;

$T_m$  – кількість робочих днів тиждень;

$t$  – кількість робочих годин в день.

$$C_{ч} = \frac{9000 + 9000 + 14000}{3 \cdot 21 \cdot 8} = 63,5 \text{ грн.} \quad (4.15)$$

Тоді, розрахуємо заробітну плату за формулою:

$$C_{зп} = C_{ч} \cdot T_i \cdot K_d, \quad (4.16)$$

де  $C_{ч}$  – величина погодинної оплати праці програміста;

$T_i$  – трудомісткість відповідного завдання;

$K_d$  – норматив, який враховує додаткову заробітну плату.

Зарплата розробників за варіантами становить:

$$I. \quad C_{зп} = 63,5 \cdot 1328,64 \cdot 1,2 = 101242,37 \text{ грн.}$$

$$II. \quad C_{зп} = 63,5 \cdot 1345,52 \cdot 1,2 = 102528,62 \text{ грн.}$$

Відрахування на єдиний соціальний внесок становить 22%

$$I. \quad C_{від} = C_{зп} \cdot 0,22 = 101242,37 \cdot 0,22 = 22273,32 \text{ грн.}$$

$$II. \quad C_{від} = C_{зп} \cdot 0,22 = 102528,62 \cdot 0,22 = 22556,3 \text{ грн.}$$

Тепер визначимо витрати на оплату однієї машино-години. ( $C_M$ )

Так як одна ЕОМ обслуговує одного програміста з окладом 9000 грн., з коефіцієнтом зайнятості 0,2 то для однієї машини отримаємо:

$$C_T = 12 \cdot M \cdot K_3 = 12 \cdot 9000 \cdot 0,2 = 21600 \text{ грн.}$$

З урахуванням додаткової заробітної плати:

$$C_{зп} = C_T \cdot (1 + K_3) = 26400 \cdot (1 + 0,2) = 25920 \text{ грн.}$$

Відрахування на соціальний внесок:

$$C_{від} = C_{зп} \cdot 0,22 = 31680 \cdot 0,22 = 5702,4 \text{ грн.}$$

Амортизаційні відрахування розраховуємо при амортизації 25% та вартості ЕОМ – 20000 грн.

$$C_A = K_{TM} \cdot K_A \cdot C_{ПР} = 1,15 \cdot 0,25 \cdot 20000 = 5750 \text{ грн.,}$$

де  $K_{TM}$  – коефіцієнт, який враховує витрати на транспортування та монтаж приладу у користувача;

$K_A$  – річна норма амортизації;

$C_{ПР}$  – договірна ціна приладу.

Витрати на ремонт та профілактику розраховуємо як:

$$C_P = K_{TM} \cdot C_{ПР} \cdot K_P = 1,15 \cdot 20000 \cdot 0,05 = 1150 \text{ грн.,}$$

де  $K_P$  – відсоток витрат на поточні ремонти.

Ефективний годинний фонд часу ПК за рік розраховуємо за формулою:

$$T_{ЕФ} = (D_K - D_B - D_C - D_P) \cdot t_3 \cdot K_B = (365 - 104 - 12 - 16) \cdot 8 \cdot 0,9 = \\ = 1677,6 \text{ годин,}$$

де  $D_K$  – календарна кількість днів у році;

$D_B, D_C$  – відповідно кількість вихідних та святкових днів;

$D_P$  – кількість днів планових ремонтів устаткування;

$t$  – кількість робочих годин в день;

$K_B$  – коефіцієнт використання приладу у часі протягом зміни.

Витрати на оплату електроенергії розраховуємо за формулою:

$$C_{\text{ЕЛ}} = T_{\text{ЕФ}} \cdot N_C \cdot K_3 \cdot C_{\text{ЕН}} = 1677,6 \cdot 0,3 \cdot 2,58 \cdot 3,52 = 3543,09 \text{ грн.},$$

де  $N_C$  – середньо-споживча потужність приладу;

$K_3$  – коефіцієнтом зайнятості приладу;

$C_{\text{ЕН}}$  – тариф за 1 кВт-годин електроенергії.

Накладні витрати розраховуємо за формулою:

$$C_{\text{Н}} = C_{\text{ПР}} \cdot 0,67 = 20000 \cdot 0,67 = 13400 \text{ грн.}$$

Тоді, річні експлуатаційні витрати будуть:

$$C_{\text{ЕКС}} = C_{\text{ЗП}} + C_{\text{ВІД}} + C_{\text{А}} + C_{\text{Р}} + C_{\text{ЕЛ}} + C_{\text{Н}}, \quad (4.17)$$

$$C_{\text{ЕКС}} = 25920 + 5702,4 + 5750 + 1150 + 4570,59 + 13400 = 55465,49 \text{ грн.}$$

Собівартість однієї машино-години ЕОМ дорівнюватиме:

$$C_{\text{М-Г}} = C_{\text{ЕКС}} / T_{\text{ЕФ}} = 55465,49 / 1677,6 = 33,06 \text{ грн/год.}$$

Оскільки в даному випадку всі роботи, які пов'язані з розробкою програмного продукту ведуться на ЕОМ, витрати на оплату машинного часу, в залежності від обраного варіанта реалізації, складає:

$$C_{\text{М}} = C_{\text{М-Г}} \cdot T, \quad (4.18)$$

$$\text{I. } C_{\text{М}} = 33,06 \cdot 1328,64 = 43924,84 \text{ грн.}$$

$$\text{II. } C_{\text{М}} = 33,06 \cdot 1345,52 = 44482,89 \text{ грн.}$$

Накладні витрати складають 67% від заробітної плати:

$$C_{\text{Н}} = C_{\text{ЗП}} \cdot 0,67, \quad (4.19)$$

$$\text{I. } C_{\text{Н}} = 101242,37 \cdot 0,67 = 67832,39 \text{ грн.}$$

$$\text{II. } C_{\text{Н}} = 102528,62 \cdot 0,67 = 68694,18 \text{ грн.}$$

Отже, вартість розробки ПП за варіантами становить:

$$C_{\text{ПП}} = C_{\text{ЗП}} + C_{\text{ВІД}} + C_{\text{М}} + C_{\text{Н}}, \quad (4.20)$$

I.  $C_{\text{ПП}} = 101242,37 + 22273,32 + 43924,84 + 67832,39 = 235272,92$  грн.

II.  $C_{\text{ПП}} = 102528,62 + 22556,3 + 44482,89 + 68694,18 = 237961,99$  грн.

#### 4.7 Вибір кращого варіанту ПП техніко-економічного рівня

Розрахуємо коефіцієнт техніко-економічного рівня за формулою:

$$K_{\text{ТЕР}j} = K_{\text{К}j} / C_{\text{Ф}j}, \quad (4.21)$$

$$K_{\text{ТЕР}1} = 4,44 / 235272,92 = 1,89 \cdot 10^{-5},$$

$$K_{\text{ТЕР}2} = 3,56 / 237961,99 = 1,6 \cdot 10^{-5}.$$

Як бачимо, найбільш ефективним є перший варіант реалізації програми з коефіцієнтом техніко-економічного рівня  $K_{\text{ТЕР}1} = 1,89 \cdot 10^{-5}$ .

Після виконання функціонально-вартісного аналізу програмного комплексу що розроблюється, можна зробити висновок, що з альтернатив, що залишилися після першого відбору двох варіантів виконання програмного комплексу оптимальним є перший варіант реалізації програмного продукту. У нього виявився найкращий показник техніко-економічного рівня якості  $K_{\text{ТЕР}} = 1,89 \cdot 10^{-5}$ .

Цей варіант реалізації програмного продукту має такі параметри:

- мова програмування – Python;
- використання моделей з великою ємністю;
- використання стандартного інтерфейсу візуалізації, швидкість розробки.

Даний варіант виконання програмного комплексу дає користувачу зручний інтерфейс, непоганий функціонал і швидкодію.

## Висновки

Проведено повний функціонально-вартісний аналіз програмного продукту. Визначено та проведено оцінку основних функцій програмного продукту. Визначено параметри, які характеризують програмний продукт. Проведено експертне оцінювання параметрів та аналіз якості варіантів реалізації функцій.

Проведено економічний аналіз варіантів розробки – трудомісткість, витрати на заробітну плату та інші витрати.

На основі аналізу вибрано варіант реалізації програмного продукту.

## ВИСНОВКИ

У даній роботі зроблено аналіз цін на оренду реальних торгівельних приміщень і розроблено власні моделі для прогнозування попиту на приміщення та ціни на оренду у зручному для користувача вигляді.

У першому розділі даної дипломної роботи розглянуто поняття нерухомості, ринку комерційної нерухомості та орендної плати; проаналізовано методики розрахунку орендної плати в Україні.

У другому розділі розглянуто теоретичні відомості використовуваних моделей, а саме моделі дискретного вибору та моделі ARIMA.

У третьому розділі побудовано модель дискретного вибору для прогнозування попиту на приміщення та модель ARIMA для прогнозування ціни деякого приміщення на реальних даних ТРЦ “Велес”. Створено програмний продукт в середовищі Jupyter Notebook мовою Python у вигляді кінцевих модулів.

У четвертому розділі описано результати виконання функціонально-вартісного аналізу розробленого програмного продукту для виконання обчислювальних експериментів.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Accuracy score [Електронний ресурс]. – Режим доступу: [https://scikit-learn.org/stable/modules/model\\_evaluation.html#accuracy-score](https://scikit-learn.org/stable/modules/model_evaluation.html#accuracy-score)
2. Autoregressive models [Електронний ресурс]. – Режим доступу: <https://otexts.com/fpp2/AR.html>
3. Autoregressive-moving-average model [Електронний ресурс]. – Режим доступу: [https://en.wikipedia.org/wiki/Autoregressive%E2%80%93moving-average\\_model#ARMA\\_model](https://en.wikipedia.org/wiki/Autoregressive%E2%80%93moving-average_model#ARMA_model)
4. Discrete Choice Model and Analysis [Електронний ресурс]. – Режим доступу: <https://www.publichealth.columbia.edu/research/population-health-methods/discrete-choice-model-and-analysis>
5. F-score [Електронний ресурс]. – Режим доступу: <https://en.wikipedia.org/wiki/F-score>
6. Friedman, J., Ordway, Nick (1995): Analysis and evaluation of income-generating real estate. TRANS. from English., – М.: McGraw-hill.
7. Harmonic mean [Електронний ресурс]. – Режим доступу: [https://en.wikipedia.org/wiki/Harmonic\\_mean](https://en.wikipedia.org/wiki/Harmonic_mean)
8. Hensher DA, Rose J, Greene WH. 2005a. Applied Choice Analysis: A Primer. Cambridge University Press, 2005.
9. Logistic Regression – Detailed Overview [Електронний ресурс]. – Режим доступу: <https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc>
10. M. Ben-Akiva, M. Lerman. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, 1985.
11. Mean absolute percentage error [Електронний ресурс]. – Режим доступу: [https://en.wikipedia.org/wiki/Mean\\_absolute\\_percentage\\_error](https://en.wikipedia.org/wiki/Mean_absolute_percentage_error)
12. Moving average models [Електронний ресурс]. – Режим доступу: <https://otexts.com/fpp2/MA.html>

13. Precision and recall [Електронний ресурс]. – Режим доступу: [https://en.wikipedia.org/wiki/Precision\\_and\\_recall](https://en.wikipedia.org/wiki/Precision_and_recall)
14. ROC-крива [Електронний ресурс]. – Режим доступу: <https://uk.wikipedia.org/wiki/ROC-крива>
15. Train, K. Discrete Choice Methods with Simulation. Cambridge University Press, 2009.
16. Understanding Logistic Regression in Python [Електронний ресурс]. – Режим доступу: <https://www.datacamp.com/community/tutorials/understanding-logistic-regression-python>
17. What does Kfold in python exactly do? [Електронний ресурс]. – Режим доступу: <https://stackoverflow.com/questions/36063014/what-does-kfold-in-python-exactly-do>
18. Бусов, В. И. Управление недвижимостью: теория и практика: учебник для академического бакалаврата / В. И. Бусов, А. А. Поляков. – М.: Издательство Юрайт, 2014. – 517 с.
19. Господарський кодекс України від 16.01.2003 // Відомості Верховної Ради, 2003. – № 18-22
20. Дзера О.В. Цивільне право України: Підручник: У 2-х кн. / О. В. Дзери (кер. авт. кол.), Д. В. Боброва, А. С. Довгерт та ін.; За ред. О. В. Дзери, Н. С. Кузнецової. – К.: «Юрінком Інтер», 2004. – Кн. 1. – С. 147.
21. Економіка нерухомості: Підручник. А.М. Асаул, І.А.Брижань, В.Я.Чевганова. – К.: «Лібра», 2012. – 304с.
22. Закон України "Про оподаткування прибутку підприємств" від 28.12.94 // ВВР України. – 1994. – № 334/94.
23. Закон України «Про збір на обов'язкове державне пенсійне страхування». [Електронний ресурс]: Оф. Сайт Верховної Ради України. – Режим доступу: <http://zakon1.rada.gov.ua/cgi-bin/laws/main.cgi>
24. Закон України «Про інвестиційну діяльність» від 18.09.1991р. // ВВР України. – 1991. – № 47. – 646 с.

25. Закон України «Про іпотеку» від 05.06.2003 р. № 898-ІУ // Відомості Верховної Ради України. – 2003. – № 38. – С. 313.

26. Закон України «Про оренду державного та комунального майна» від 10.04.1992р. // ВВР України. – 1992. – № 30. – 416 с.

27. Козак С.В. Статистичне оцінювання ринку житлової нерухомості України. Дис. канд. екон. наук: 08.00.10, Нац. акад. статиститки, обліку та аудиту. – Київ, 2014. – 200 с.

28. Матриця невідповідностей [Електронний ресурс]. – Режим доступу: [https://uk.wikipedia.org/wiki/Матриця\\_невідповідностей](https://uk.wikipedia.org/wiki/Матриця_невідповідностей)

29. Мочерний С.В. Економічна енциклопедія: у трьох томах. Т.3. / С. В. Мочерний (відп. ред.) та ін. – К.: Видавничий центр «Академія», 2002. – 952 с.

30. Про Велес [Електронний ресурс]. – Режим доступу: <https://veles.in.ua/about/>

31. Сергеева И.Г. Экономика недвижимости: Учеб. пособие. / И. Г. Сергеева. – СПб.: НИУ ИТМО, 2013. – 51 с.

32. Среднеквадратическая ошибка модели [Електронний ресурс]. – Режим доступу: <http://statistica.ru/glossary/general/srednekvadraticheskaya-oshibka/>

33. Тацій Ю.О. Розвиток ринків комерційної нерухомості у великих містах / Ю. О. Тацій // Містобудування та територіальне планування. – 2014. – № 52. – С. 403-410.

34. Цивільний кодекс України від 16.01.2003р. // ВВР України. – 2003. – № 40-44. – 356 с.

## ДОДАТОК А ЛІСТИНГ ПРОГРАМНОГО ПРОДУКТУ

Модуль demand\_prediction.py

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from copy import deepcopy
import re
import pickle

from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.model_selection import train_test_split, cross_validate
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import plot_confusion_matrix, get_scorer, roc_auc_score,
roc_curve
from sklearn.pipeline import make_pipeline, Pipeline

class ColumnLabelEncoder:
    def __init__(self, column='Контрагент'):
        self.column = column

    def fit(self, X, y=None):
        return self

    def transform(self, X):
        output = X.copy()
        output[self.column] = LabelEncoder().fit_transform(output[self.column])
        return output
```

```
def fit_transform(self, X, y=None):  
    return self.fit(X, y).transform(X)
```

```
class GetFloor:
```

```
    def __init__(self, column='Лот'):  
        self.column = column
```

```
    def fit(self, X, y=None):  
        return self
```

```
    def transform(self, X):  
        output = X.copy()
```

```
    def get_floor(lot):  
        if lot == 'парковка':  
            return 0  
        try:  
            floor = lot // 100  
        except:  
            try:  
                floor = int(lot.split('/')[0]) // 100  
            except:  
                floor = 3  
        return floor
```

```
    output[self.column] = output[self.column].apply(get_floor)  
    return output
```

```

def fit_transform(self, X, y=None):
    return self.fit(X, y).transform(X)

def plot_classification(estimator, X_train, X_test, y_train, y_test):
    """Plots predicted and actual value"""

    estimator.fit(X_train, y_train)
    y_predicted = estimator.predict_proba(X_test)[:, 1]
    score = roc_auc_score(y_test, y_predicted) * 100

    fpr, tpr, thresholds = roc_curve(y_test, y_predicted)

    plt.figure(figsize=(10, 6))
    plt.plot(fpr, tpr)
    plt.axis('tight')
    model_name = re.findall(r'^[^(.]+', str(estimator))[0]
    plt.title(f"model {model_name}, roc curve: %.2f" % score + "%")
    plt.grid(True)
    plt.show()

def cv_scores(model, X_train, y_train, scoring, cv=5, verbose=0):
    """Realizes cross_validate function for scorings"""

    scores = cross_validate(model, X_train, y_train, cv=cv, n_jobs=-
1,scoring=scoring)

    cv_results = {}
    for score in scoring:

```

```

cv_scores = np.mean(scores[f'test_{score}'])
cv_results[f'{score}'] = cv_scores

```

```

if verbose > 2:
    print(f"fit_time: {cv_results['fit_time']}")
    print(f"score_time: {cv_results['score_time']}\n")
    for score in scoring:
        print(f"mean cv {score} scores: {cv_results[f'{score}']}")

return cv_results

```

```

def test_scores(model, X_train, X_test, y_train, y_test, scoring, verbose=0):
    """Returns trained model and different scorings"""

    model.fit(X_train, y_train)

    test_results = {}
    for score in scoring:
        score_func = get_scorer(score)
        test_scores = score_func(model, X_test, y_test)
        test_results[f'{score}'] = test_scores

        if verbose > 2:
            print(f'test {score} scores:', test_scores)

    return model, test_results

```

```

def combine_model_results(cv, test, model_name):

```

```
"""Processes and merges cv and test scores results"""
```

```
indexs = []
```

```
for idx in range(len(cv)):
```

```
    indexs.append(('cv', list(cv.keys())[idx]))
```

```
index = pd.MultiIndex.from_tuples(indexs, names=["data", "metrics"])
```

```
cv_series = pd.Series(cv.values(), index=index, name=model_name)
```

```
indexs = []
```

```
for idx in range(len(test)):
```

```
    indexs.append(('test', list(test.keys())[idx]))
```

```
index = pd.MultiIndex.from_tuples(indexs, names=["data", "metrics"])
```

```
test_series = pd.Series(test.values(), index=index, name=model_name)
```

```
return pd.concat([cv_series, test_series])
```

```
def baseline(model, X_train, X_test, y_train, y_test, scorings=['accuracy'],
verbose=False):
```

```
    """Returns pretrained model and cv score with test score as series"""
```

```
    cv_result = cv_scores(deepcopy(model), X_train, y_train, scorings, cv=5,
verbose=verbose)
```

```
    model, test_result = test_scores(model, X_train, X_test, y_train, y_test,
scoring=scorings, verbose=verbose)
```

```
    model_name = re.findall(r'^[^(.]+', str(model))[0]
```

```
    results = combine_model_results(cv_result, test_result, model_name)
```

```
return model, results
```

```
def add_demand_percent_column(model, X):
```

```
    """
```

```
    Add 'Попит %' column to dataframe.
```

```
    Arguments:
```

- model: trained model;
- X: data to transform.

```
    Returns:
```

- data: transformed X with added 'Попит %' column.

```
    """
```

```
    data = X.copy()
```

```
    y_pred_proba = model.predict_proba(X)
```

```
    data["Попит %"] = np.round(y_pred_proba[:, 1] * 100, 2)
```

```
    return data
```

```
def save_model(model, filename):
```

```
    filename = filename
```

```
    pickle.dump(model, open(filename, 'wb'))
```

```
def main(data_file, model=LogisticRegression()):
```

```
    """
```

```
    Run pipeline on data from data_file.
```

```
    Arguments:
```

- data\_file: str, path for data file to read data from (excel format);
- model: model to use for training.

```

"""

data = pd.read_excel(data_file)

y, X = data['Попит'], data.drop(['Попит'], axis=1)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33,
random_state=42)

model_name = model.__class__.__name__

pipe = Pipeline([('GetFloor', GetFloor()),
                 ('ColumnLabelEncoder', ColumnLabelEncoder()),
                 ('StandardScaler', StandardScaler()),
                 (model_name, model)])
pipe.__class__.__name__ = model_name + "pipeline"

scorings = ['accuracy', 'roc_auc', 'precision', 'recall', 'f1']
pipe, results = baseline(pipe, X_train, X_test, y_train, y_test, scorings=scorings)

plot_classification(pipe, X_train, X_test, y_train, y_test)
plot_confusion_matrix(pipe, X_test, y_test)
print(results)

return pipe, results

```

Модуль price\_prediction.py

```

import warnings
warnings.filterwarnings('ignore')

```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
from statsmodels.tsa.stattools import adfuller
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from sklearn.metrics import mean_squared_error
from scipy import stats

color = sns.color_palette()
sns.set_style('darkgrid')

def plot_price(data):
    plt.figure(figsize=(10,6))
    sns.lineplot(x="Date", y="Price", legend='full', data=data)

def test_stationarity(timeseries, window=2, cutoff=0.01):

    #Determining rolling statistics
    rolmean = timeseries.rolling(window).mean()
    rolstd = timeseries.rolling(window).std()

    #Plot rolling statistics:
    fig = plt.figure(figsize=(12, 8))
    orig = plt.plot(timeseries, color='blue',label='Original')
    mean = plt.plot(rolmean, color='red', label='Rolling Mean')
    std = plt.plot(rolstd, color='black', label = 'Rolling Std')
```

```

plt.legend(loc='best')
plt.title('Rolling Mean & Standard Deviation')
plt.show()

#Perform Dickey-Fuller test:
print('Results of Dickey-Fuller Test:')
dftest = adfuller(timeseries, autolag='AIC', maxlag=2)
dfoutput = pd.Series(dftest[0:4], index=['Test Statistic','p-value','#Lags
Used','Number of Observations Used'])
for key,value in dftest[4].items():
    dfoutput['Critical Value (%s)'%key] = value
pvalue = dftest[1]
if pvalue < cutoff:
    print('p-value = %.4f. The series is likely stationary.' % pvalue)
else:
    print('p-value = %.4f. The series is likely non-stationary.' % pvalue)

print(dfoutput)

def get_first_diff(train_df):
    first_diff = train_df['Price'] - train_df['Price'].shift(1)
    first_diff = first_diff.dropna(inplace = False)
    return first_diff

def plot_acf_pacf(data, lags=10):
    fig = plt.figure(figsize=(12,8))
    ax1 = fig.add_subplot(211)
    fig = plot_acf(data, lags=10, ax=ax1)

```

```
ax2 = fig.add_subplot(212)
fig = plot_pacf(data, lags=10, ax=ax2)
```

```
def check_prediction(data, size=30, p=4, d=1, q=1):
    train, test = data['Price'][0:size], data['Price'][size:len(data)]
    date = data['Date'][size:len(data)]
    history = [x for x in train]
    predictions = list()

    # walk-forward validation
    for t in test:
        model = sm.tsa.ARIMA(history, order=(p,d,q))
        model_fit = model.fit()
        output = model_fit.forecast()
        yhat = output[0][0]
        predictions.append(yhat)
        history.append(t)
        print('predicted=%f, expected=%f' % (yhat, t))

plt.figure(figsize=(10,6))
plt.plot(date, test, label='true')
plt.plot(date, predictions, 'red', label='prediction')
plt.legend()
plt.ylabel("Price")
plt.xlabel("Date")
plt.title("Price Prediction")

predictions = np.array(predictions)
```

```

mape = np.mean(abs((test-predictions)/test))*100
print("\nMAPE: %.2f% mape, \"%")

rmse = np.sqrt(mean_squared_error(test, predictions))
print('RMSE: %.4f% rmse)

return model_fit, predictions

def check_model(model):
    resid = model.resid
    print(stats.normaltest(resid))
    # returns a 2-tuple of the chi-squared statistic, and the associated p-value. the p-
value is very small, meaning
    # the residual is not a normal distribution

    fig = plt.figure(figsize=(12,8))
    ax0 = fig.add_subplot(111)

    sns.distplot(resid, fit=stats.norm, ax=ax0) # need to import scipy.stats

    # Get the fitted parameters used by the function
    (mu, sigma) = stats.norm.fit(resid)

    #Now plot the distribution using
    plt.legend(['Normal dist. ( $\mu$ = $\{:.2f\}$  and  $\sigma$ = $\{:.2f\}$  )'.format(mu,
sigma)], loc='best')
    plt.ylabel('Frequency')
    plt.title('Residual distribution')

```

```
# ACF and PACF
fig = plt.figure(figsize=(12,8))
ax1 = fig.add_subplot(211)
fig = plot_acf(model.resid, lags=10, ax=ax1)
ax2 = fig.add_subplot(212)
fig = plot_pacf(model.resid, lags=10, ax=ax2)

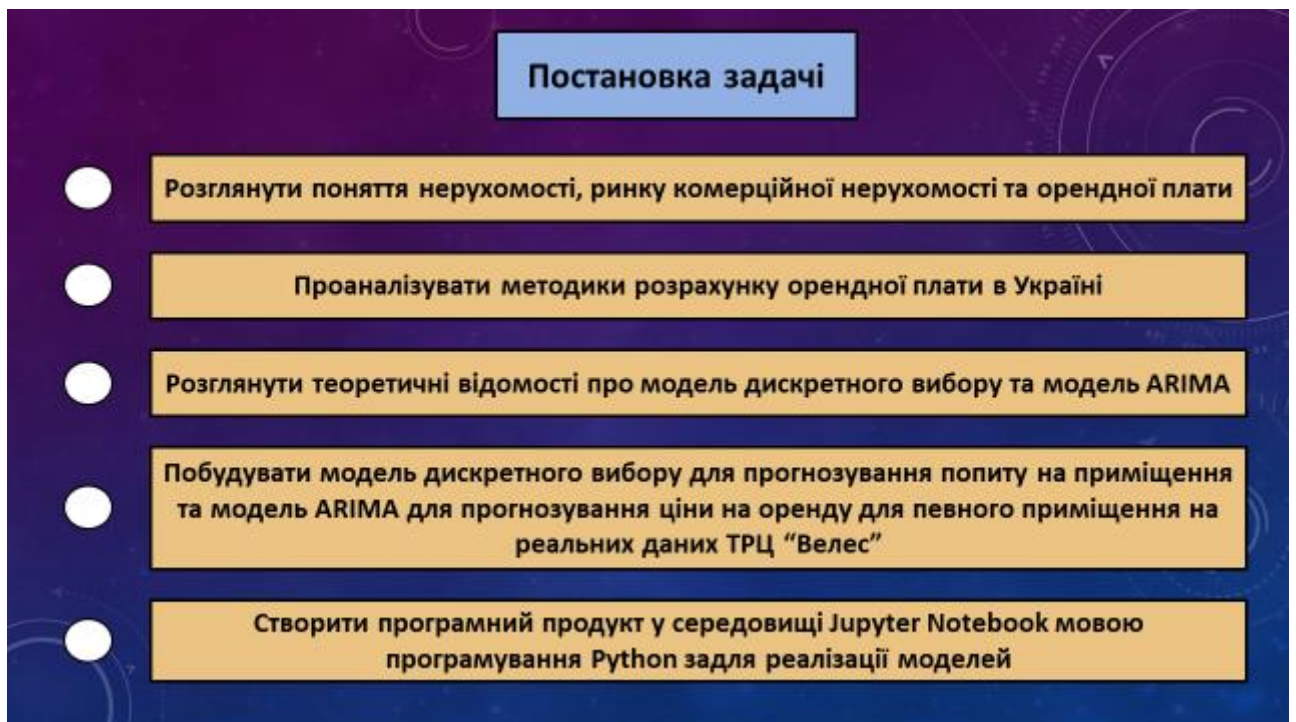
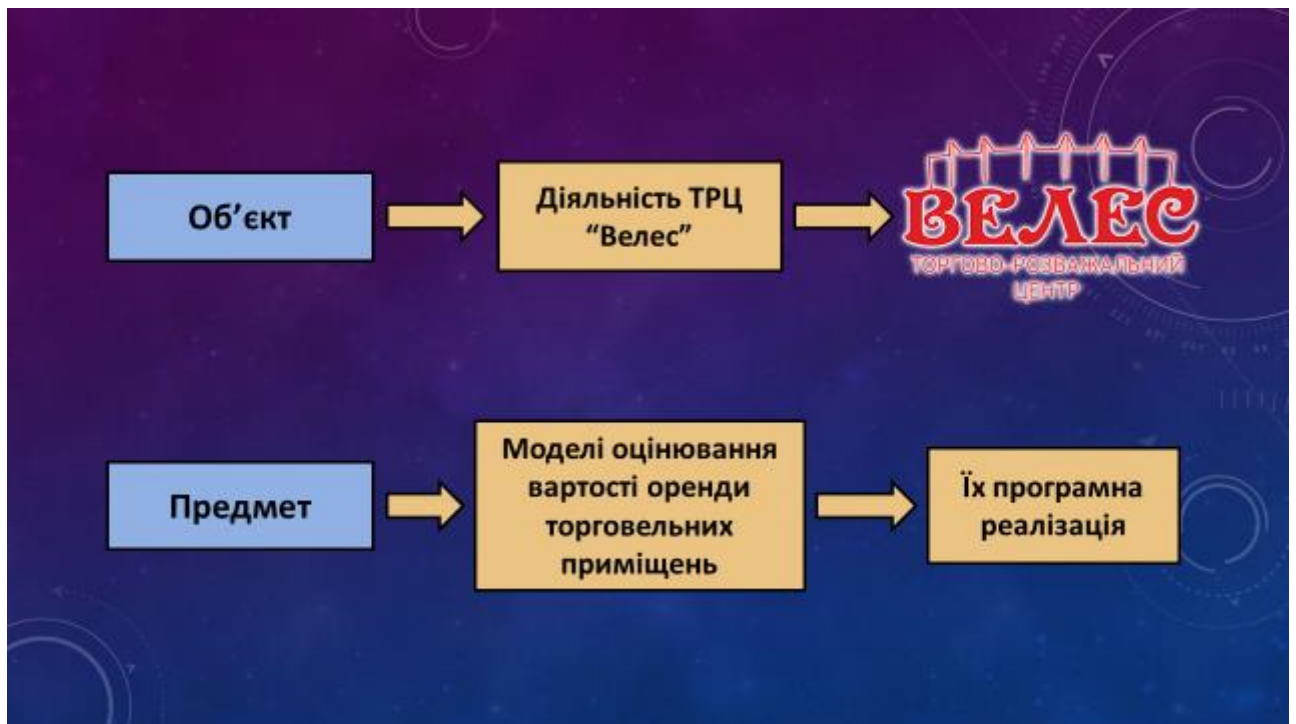
def predict(data, p=4, d=1, q=1):
    model = sm.tsa.ARIMA(data, order=(p,d,q))
    model_fit = model.fit()
    output = model_fit.forecast()
    return output[0][0]
```

## ДОДАТОК Б ПРЕЗЕНТАЦІЯ

# СИСТЕМА ДЛЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ВАРТОСТІ ТОРГОВЕЛЬНИХ ПРИМІЩЕНЬ

ВОЗНЯК ВОЛОДИМИР ЗІНОВІЙОВИЧ  
КЕРІВНИК: БІДЮК ПЕТРО ІВАНОВИЧ







**Модель дискретного вибору**

$$h_{\theta}(x^i) = \theta^T x^i, \text{ де } x^i = \begin{pmatrix} x_0^i \\ x_1^i \\ x_2^i \\ \dots \\ x_n^i \end{pmatrix} \in \mathbb{R}^{n+1}; \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \dots \\ \theta_n \end{pmatrix} \in \mathbb{R}^{n+1}$$

$$\text{Cost}(h_{\theta}(x), y) = -y \log(h_{\theta}(x)) - (1 - y) \log(1 - h_{\theta}(x))$$

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m y^i * \log(h_{\theta}(x^i)) + (1 - y^i) * \log(1 - h_{\theta}(x^i))$$

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) = \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^i) - y^i) * x_j^i$$

	Лот	Контрагент	Площа	Ціна
0	101	посуд	41.00	488.29
1	102	одяг	46.00	550.00
2	103	білизна	33.00	650.00
3	104	косметика	22.70	750.00
4	104/1	посуд	35.15	510.00
...	...	...	...	...
133	Офіс №8/1	офіс	32.00	80.00
134	Офіс №9	офіс	30.00	167.00
135	Офіс №9/1	офіс	32.00	100.00
136	Офіс №9/2	офіс	32.00	100.00
137	парковка	електрозаправка	2.00	50.00

138 rows x 4 columns

**Початкові дані**



**Схема 1 поверху ТРЦ "Велес"**

	Лот	Контрагент	Площа	Ціна
0	101	посуд	41.00	488.29
1	102	одяг	46.00	550.00
2	103	білизна	33.00	650.00
3	104	косметика	22.70	750.00
4	104/1	посуд	35.15	510.00
...	...	...	...	...
133	Офіс №8/1	офіс	32.00	80.00
134	Офіс №9	офіс	30.00	167.00
135	Офіс №9/1	офіс	32.00	100.00
136	Офіс №9/2	офіс	32.00	100.00
137	парковка	електрозаправка	2.00	50.00

138 rows x 4 columns

**Початкові дані**



	Лот	Контрагент	Площа	Ціна	Біля входу	Попит
0	1	38	41.00	488.29	0	0
1	1	32	46.00	550.00	0	0
2	1	10	33.00	650.00	0	0
3	1	24	22.70	750.00	0	0
4	1	38	35.15	510.00	0	0
...	...	...	...	...	...	...
133	3	35	32.00	80.00	0	1
134	3	35	30.00	167.00	0	1
135	3	35	32.00	100.00	0	1
136	3	35	32.00	100.00	0	1
137	0	20	2.00	50.00	0	0

138 rows x 6 columns

**Дані після обробки**

```
X_train shape: (92, 5)
X_test shape: (46, 5)
y_train shape: (92,)
y_test shape: (46,)
```

**Розмірності тренувальних та тестових матриць признаков та таргет-векторів**



```
data metrics
cv accuracy 0.859064
roc_auc 0.931563
precision 0.927778
recall 0.853846
f1 0.887019
test accuracy 0.934783
roc_auc 0.972917
precision 0.965517
recall 0.933333
f1 0.949153
```

**Отримані метрики**

	Лот	Контрагент	Площа	Ціна	Біля входу	Попит	Попит %
0	101	посуд	41.00	488.29	0	0	47.72
1	102	одяг	46.00	550.00	0	0	38.74
2	103	білизна	33.00	650.00	0	0	16.50
3	104	косметика	22.70	750.00	0	0	23.01
4	104/1	посуд	35.15	510.00	0	0	46.51
...	...	...	...	...	...	...	...
133	Офіс №8/1	офіс	32.00	80.00	0	1	99.08
134	Офіс №9	офіс	30.00	167.00	0	1	98.93
135	Офіс №9/1	офіс	32.00	100.00	0	1	99.05
136	Офіс №9/2	офіс	32.00	100.00	0	1	99.05
137	парковка	електрозаправка	2.00	50.00	0	0	9.74

138 rows × 7 columns

Початкові дані із доданою колонкою "Попит %"

## Модель ARIMA

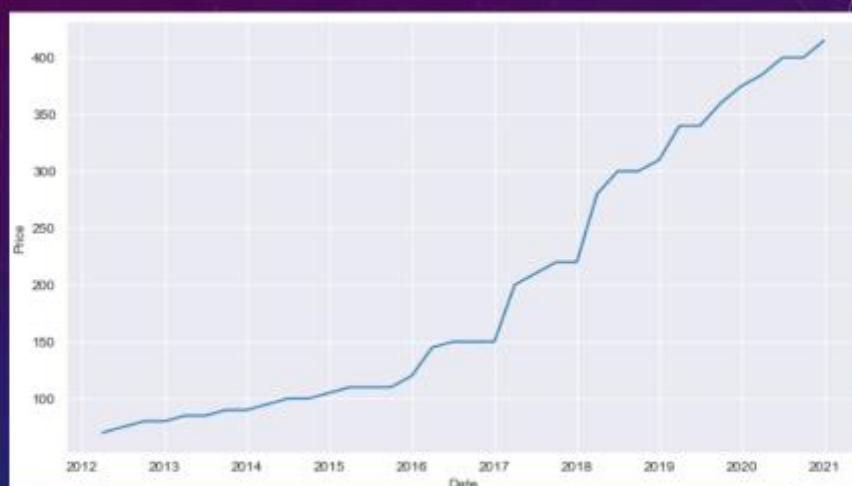
ARIMA(p,d,q):

$$\Delta^d X_t = c + \sum_{i=1}^p a_i \Delta^d X_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t,$$

де  $\varepsilon_t$  – "білий шум";

$c, a_i, b_j$  – параметри моделі;

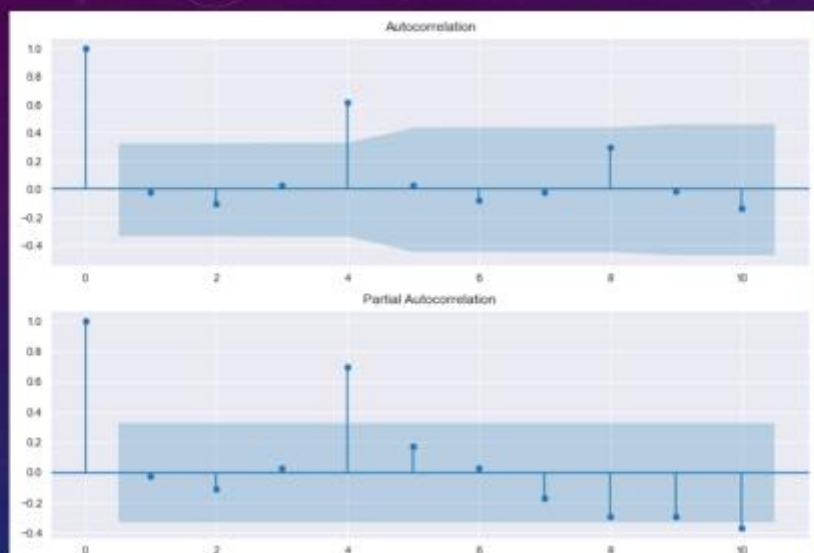
$\Delta^d$  – оператор різниці часового ряду порядку d.



Графік залежності ціни від дати для лоту №236



Візуальний спосіб перевірки стаціонарності  
продиференційованого часового ряду



АКФ та ЧАКФ для продиференційованого часового ряду



Графік реальної й спрогнозованої ціни оренди

## Висновки

- ✓ Розглянуто поняття нерухомості, ринку комерційної нерухомості та орендної плати
- ✓ Проаналізовано методики розрахунку орендної плати в Україні
- ✓ Розглянуто теоретичні відомості про модель дискретного вибору та модель ARIMA
- ✓ Побудовано модель дискретного вибору для прогнозування попиту на приміщення та модель ARIMA для прогнозування ціни на оренду для певного приміщення на реальних даних ТРЦ "Велес"
- ✓ Створено програмний продукт у середовищі Jupyter Notebook мовою програмування Python задля реалізації моделей

## Можливі подальші дослідження

- Протестувати поведінку моделей на даних інших ТРЦ
- Зробити моделі універсальними для будь-яких вхідних даних

**Дякую за увагу!**

