

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

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

До захисту допущено:

Завідувач кафедри

\_\_\_\_\_ Оксана ТИМОЩУК

«\_\_» \_\_\_\_\_ 2023 р.

**Дипломна робота**

**на здобуття ступеня бакалавра**

**за освітньо-професійною програмою «Системний аналіз і управління»**

**спеціальності 124 «Системний аналіз»**

**на тему: «Прогнозування вартості оренди житла**

**з використанням кількісних та якісних методів»**

Виконала: студентка IV курсу, групи КА-93

Личковаха Анастасія Олексіївна

\_\_\_\_\_

Керівник: доцент, к.т.н.

Савченко Ілля Олександрович

\_\_\_\_\_

Консультант з нормоконтролю: к.ф.-м.н.

Статкевич Віталій Михайлович

\_\_\_\_\_

Консультант з економічного розділу: доцент, к.е.н.

Рощина Надія Василівна

\_\_\_\_\_

Рецензент: | доцент, к.т.н.

Булах Богдан Вікторович

\_\_\_\_\_

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

Студентка \_\_\_\_\_

Київ – 2023 року

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

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 124 «Системний аналіз»

Освітня програма «Системний аналіз і управління»

ЗАТВЕРДЖУЮ

Завідувач кафедри

\_\_\_\_\_ Оксана ТИМОЩУК

«\_\_» травня 2023 р.

**ЗАВДАННЯ**  
**на дипломну роботу студентці**  
**Личковасі Анастасії Олексіївни**

1. Тема роботи «Прогнозування вартості оренди житла з використанням кількісних та якісних методів», керівник роботи доцент, к.т.н. Савченко Ілля Олександрович, затверджені наказом по університету від «30» травня 2023 р. №2065-с.
2. Термін подання студентом роботи – «13» травня 2023 р.
3. Вихідні дані до роботи: зібрані реальні дані про оренду житла у Києві в табличному форматі.
4. Зміст роботи: 1. Дослідження теоретичних складових вартості оренди житла. 2. Дослідження кількісних та якісних методів для прогнозування вартості оренди житла. 3. Побудова моделі лнійної регресії та проведення морфологічного аналізу. 4. Функціонально-вартісний аналіз програмного продукту.

5. Перелік ілюстративного матеріалу графіки попиту на оренду не рухомого майна, малюнки залежностей у моделі лінійної регресії, рисунки прикладів таблиць морфологічного аналізу, графіки розподілу результатів відносно вихідних даних, презентація.

#### 6. Консультанти розділів роботи

| Розділ      | Прізвище, ініціали та посада консультанта | Підпис, дата   |                  |
|-------------|-------------------------------------------|----------------|------------------|
|             |                                           | завдання видав | завдання прийняв |
| Економічний | Рощина Н.В., доцент, к.е.н.               |                |                  |

#### 7. Дата видачі завдання «17» квітня 2023 р.

#### Календарний план

| № з/п | Назва етапів виконання дипломної роботи                                                                               | Термін виконання етапів роботи | Примітка |
|-------|-----------------------------------------------------------------------------------------------------------------------|--------------------------------|----------|
| 1.    | Формування тематики дослідження                                                                                       | 23.04.2023                     | Виконано |
| 2.    | Попередня підготовка до написання роботи: огляд літератури за обраною тематикою, постановка задачі                    |                                |          |
| 3.    | Підготовка першого розділу роботи: збір інформації, вибір даних, аналіз можливостей побудови програми                 | 07.05.2023                     | Виконано |
| 4.    | Підготовка другого розділу роботи: математичний опис використаних методів та програмних алгоритмів, побудова програми | 14.05.2023                     | Виконано |

Студентка

\_\_\_\_\_

Анастасія ЛИЧКОВАХА

Керівник

\_\_\_\_\_

Ілля САВЧЕНКО

## РЕФЕРАТ

Дипломна робота: 107 с., 21 рис., 16 табл., 2 додатки, 23 джерела.

ПРОГНОЗУВАННЯ, ЛІНІЙНА РЕГРЕСІЯ, МОРФОЛОГІЧНИЙ  
АНАЛІЗ, АНАЛІЗ РИНКУ НЕРУХОМОСТІ.

Дана робота присвячена дослідження ринку нерухомості в Україні, а саме оренди нерухомого житла (квартири) у місті Києві. За результатами дослідження очікується отримати прогноз вартості оренди зазначеного житла. За вихідні дані було взято самостійно зібрані дані про вартість оренди квартири у Києві у реальному часі.

Об'єкт дослідження – прогнозування ціни вартості оренди житла із використанням кількісних та якісних методів на основі самостійно зібраних реальних даних.

Мета дослідження – розробка за використання мови програмування Python та програми Excel алгоритму, що дозволяє точно та чітко спрогнозувати дані використовуючи різнопланові методи, які враховують вплив параметрів, не пов'язаних між собою. Ефективність побудованого прогнозу була оцінена обрана метриками та скорельована обраним методом якісного аналізу.

Ключовою особливістю розробленого алгоритму є використання двох методів прогнозування та універсальність його використання, тобто спроможність адаптування під інші вихідні дані.

## ABSTRACT

Bachelor thesis: 107 p., 21 fig., 16 tabl., 2 append., 34 sources.

FORECASTING, LINEAR REGRESSION, MORPHOLOGICAL ANALYSIS, REAL ESTATE MARKET ANALYSIS.

This work is dedicated to the research of the real estate market in Ukraine, specifically the rental of residential properties (apartments) in the city of Kyiv. The aim of the research is to obtain a forecast of the rental cost of the mentioned properties. The initial data consists of independently collected real-time rental price data for apartments in Kyiv.

The object of the research is the prediction of the rental price using quantitative and qualitative methods based on independently gathered real data.

The objective of the study is to develop an algorithm using the Python programming language and Excel software, which allows for accurate and precise prediction using diverse methods that consider the influence of unrelated parameters. The effectiveness of the developed forecast was evaluated using selected metrics and correlated with the chosen qualitative analysis method.

The key feature of the developed algorithm is the utilization of two forecasting methods and its universality, meaning its ability to adapt to different input data.

## ЗМІСТ

|                                                                                                      |    |
|------------------------------------------------------------------------------------------------------|----|
| ВСТУП.....                                                                                           | 9  |
| РОЗДІЛ 1. ДОСЛІДЖЕННЯ ТЕОРЕТИЧНИХ СКЛАДОВИХ ВАРТОСТІ<br>ОРЕНДИ ЖИТЛА.....                            | 11 |
| 1.1 Теоретичні відомості щодо нерухомого майна та його класифікація.....                             | 11 |
| 1.2 Поняття ринку нерухомості та його динаміка в Україні.....                                        | 15 |
| 1.3 Методи розрахунку вартості оренди нерухомого майна.....                                          | 21 |
| 1.4 Висновки до розділу 2 .....                                                                      | 23 |
| РОЗДІЛ 2. ДОСЛІДЖЕННЯ КІЛЬКІСНИХ ТА ЯКІСНИХ МЕТОДІВ ДЛЯ<br>ПРОГНОЗУВАННЯ ВАРТОСТІ ОРЕНДИ ЖИТЛА ..... | 24 |
| 2.1 Методи та моделі прогнозування.....                                                              | 24 |
| 2.2 Методи кількісного аналізу. Модель лінійної регресії.....                                        | 26 |
| 2.3 Методи якісного аналізу. Метод морфологічного аналізу.....                                       | 38 |
| 2.4 Висновки до розділу 2 .....                                                                      | 46 |
| РОЗДІЛ 3. ПОБУДОВА МОДЕЛІ ЛІНІЙНОЇ РЕГРЕСІЇ ТА ПРОВЕДЕННЯ<br>МОРФОЛОГІЧНОГО АНАЛІЗУ .....            | 48 |
| 3.1 Прогнозування за побудованою моделлю лінійної регресії.....                                      | 48 |
| 3.2 Прогнозування застосуванням методу морфологічного аналізу.....                                   | 53 |
| 3.3 Висновки на основі отриманих результатів.....                                                    | 64 |
| РОЗДІЛ 4. ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО<br>ПРОДУКТУ .....                               | 69 |
| 4.1 Постановка задачі проектування.....                                                              | 70 |
| 4.2 Обґрунтування функцій програмного продукту.....                                                  | 70 |
| 4.3 Обґрунтування системи параметрів програмного продукту.....                                       | 73 |
| 4.4 Аналіз експертного оцінювання параметрів.....                                                    | 77 |
| 4.5 Аналіз рівня якості варіантів реалізації функцій.....                                            | 80 |
| 4.6 Економічний аналіз варіантів розробки ПП.....                                                    | 82 |

|                                                               |     |
|---------------------------------------------------------------|-----|
|                                                               | 8   |
| 4.7 Вибір кращого варіанту ПП техніко-економічного рівня..... | 88  |
| 4.8 Висновки до розділу.....                                  | 89  |
| ВИСНОВКИ.....                                                 | 90  |
| ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ.....                                 | 92  |
| ДОДАТОК А ЛІСТИНГ ПРОГРАМИ.....                               | 94  |
| ДОДАТОК В ПРЕЗЕНТАЦІЯ .....                                   | 102 |



## ВСТУП

Ринок нерухомості в Україні за різних часів потерпав від різного роду змін за впливу багатьох зовнішніх факторів. Наразі, ринок оренди нерухомого майна є дуже нестабільною ланкою, враховуючи, зовнішні фактори впливу, які несе за собою повномасштабне вторгнення країни-агресора на території нашої держави.

Метою дослідження є побудова алгоритму що дозволяє точно та чітко спрогнозувати дані, використовуючи кількісні та якісні методи, що дозволило б усебічно розглянути ситуацію та оцінити вплив параметрів, які не пов'язані між собою, або важно описуються математично.

Завдання полягає у створенні алгоритму для побудови чіткого математичного прогнозу кількісним методом, що базується на реальних даних та коригування цього прогнозу за допомогою отриманих якісним методом висновків, що залежать від чинників, які не могли бути розглянуті у вихідних даних або оцінені математично.

Об'єктом дослідження є власноруч зібрані реальні дані про орендоване житло (квартири) у місті Києві впродовж певного проміжку часу. У процесі побудови прогнозу задля досягнення поставленої у роботі мети, на думку авторів, є украй важливим спиратися саме на реальні дані для досягнення найбільш точних та не теоретичних, а наближених до реалій результатів.

Предметом дослідження визначено якісні та кількісні методи прогнозування, такі як лінійна регресія та модифікований метод морфологічного аналізу.

Пояснювальна записка, складена до роботи, складається із чотирьох розділів. У першому досліджено теоретичну основу ринку нерухомості, проаналізовано взаємозв'язки, що його формують та розглянуто фактори впливу та процес формування вартості оренди житла. У другому розділі глибоко розглянуто математичний підхід, можливості, переваги та суть методів, що

використовуються для прогнозування. У третьому розділі наведено їх реалізацію та проаналізовано отримані результати. Четвертий розділ присвячено функціонально-вартісному аналізу розробленого продукту.

# 1 ДОСЛІДЖЕННЯ ТЕОРЕТИЧНИХ СКЛАДОВИХ ТА ФАКТОРІВ ВПЛИВУ ВАРТОСТІ ОРЕНДИ ЖИТЛА

## 1.1. Теоретичні відомості щодо нерухомого майна та його класифікація

Нерухомість – це вид майна, який законодавчо визнано нерухомим. Тобто, це земельні ділянки, об'єкти, пов'язані із землею, будівлі та такі споруди, переміщення котрих неможливе без нанесення їм непорівняної шкоди [12].

Поняття «нерухомість» на законодавчому рівні України визначається Законом України «Про іпотеку»: нерухоме майно - 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) [13]. Також 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) [14].

Враховуючи наведені вище законодавчі акти можна виділити наступні ознаки нерухомості.

1. Об'єкти нерухомості є земельними ділянками або будівлями чи спорудами, нерозривно із нею пов'язаними.

2. Переміщення об'єктів нерухомості є неможливим або таким, що знецінює їх або змінює їх призначення.

Із 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.

У Державному реєстрі прав зазначаються права на такі об'єкти нерухомого майна:

1. Житлові будинки.
2. Квартири.
3. Будівлі, у яких розташовані приміщення, призначені для перебування людини, розміщення рухомого майна, збереження матеріальних цінностей, здійснення виробництва тощо.
4. Приміщення – частини внутрішнього об'єму житлових будинків, квартир та будівель.
5. Споруди, призначені для здійснення спеціальних технічних функцій, що не належать до будівель та приміщень.

Класифікація таких об'єктів із точки зору практичного управління нерухомим майном може бути розподілена за наступними ознаками:

1. Спеціалізація:
  - a. спеціалізована нерухомість – нерухомість, що має спеціальний характер, у зв'язку з чим рідко продається на відкритому ринку. Часто для такого майна із його конструктивними особливостями, розмірами, специфікою та характеристиками відсутній ринок у даній місцевості або відсутній взагалі. Прикладами таких об'єктів можуть слугувати заводи (нафтопереробні, хімічні, металургійні) або гідро- и електростанції;
  - b. неспеціалізована нерухомість – усі інша нерухомість, на яку існує попит на загальному ринку для використання у існуючих або аналогічних цілях.
2. Ступінь готовності до експлуатації:
  - a. об'єкти, що введені в експлуатацію;
  - b. об'єкти, що вимагають реконструкції або капітального ремонту перед уведенням в експлуатацію;

- с. Об'єкти, будівництво яких не завершене.
- 3. Походження:
  - а. природні об'єкти – земельні ділянки, насадження, відокремлені водні об'єкти, ділянки надр. Ці об'єкти нерухомості називають ще й нерухомістю за природою;
  - б. штучні об'єкти – об'єкти, що належать до «нерухомості за законом».
- 4. Відтворюваність у натуральній формі:
  - а. відтворювані об'єкти – будівлі, споруди, насадження;
  - б. невідтворювані об'єкти – земельні ділянки, виокремлені водні об'єкти, ділянки надр.

У практичному застосуванні найчастіше об'єкти нерухомості класифікують за сегментами ринку, що базується на споживчих вподобаннях учасників ринку. Таким чином, класифікація об'єктів нерухомості може бути розділена на три великих сегменти – земельні ділянки, житлова і нежитлова нерухомість.

1. Житлова нерухомість – об'єкти нерухомості, призначені для проживання людини або людей. Із комерційної точки зору поняття тягне за собою такі види економічних відносин як оренда, купівля-продаж, а також будівництво та експлуатація житлового фонду. Серед житлової нерухомості найчастіше виділяють малоповерхові (до трьох поверхів), багатоповерхові (від 4 до 9 поверхів), підвищеної поверховості (від 10 до 20 поверхів), висотні (понад 20 поверхів) будинки. Також об'єктами житлової нерухомості, вочевидь, вважається квартира, кімната, секція, поверх під'їзду, дачний будинок.

2. Нежитлова нерухомість являє собою об'єкти комерційної, промислової, інституційної та рекреаційної нерухомості, готелі та мотелі.

- а. Комерційна нерухомість – офісні будівлі, споруди для роздрібної торгівлі, автомобільні стоянки, бензоколонки та інші об'єкти, призначені для використання у сфері обслуговування.
- б. Промислова нерухомість – промислові будівлі та споруди, а також будівлі, що відносяться до їх інфраструктури, діючі заводи, технопарки, спеціалізовані офіси, складські приміщення.

- с. Інституційна нерухомість – будівлі для урядових і муніципальних органів влади, навчальних закладів, лікарень, санаторіїв та інших функцій спеціального призначення.
- d. Рекреаційна промисловість – споруди та будівлі, призначені для забезпечення відпочинку. Це об'єкти на території курортних зон, клубів за інтересами (наприклад, яхт-клуб або гольф-клуб), стадіони, спортивні майданчики, басейни тощо.
- е. Готелі та мотелі є особливим класом нерухомості, так як по факту являють собою житлову нерухомість, але призначену для тимчасового, а не постійного перебування людини у її межах.

Варто зазначити, що наведені класифікації, хоч і є достатньо чіткими, але можуть не охоплювати усі об'єкти або тільки-но виникаючі. Тому, крім зазначених вище ознак класифікації, прийнято додатково виділяти нерухомість змішаного типу використання, що комбінує у собі властивості із декількох зазначених категорій.

У рамках даної роботи буде розглянуто житлову нерухомість, а точніше, квартири, не враховуючи при цьому поверховість будинку, у якому дана квартира знаходиться, а лиш зважаючи на неї, як числову характеристику та чинник, що впливає на висновки лінійно. Із точки зору комерційних відносин розглядатимемо саме оренду квартир, обмежуючись при цьому містом Києвом та не заглиблюючись у правові та економічні відносини при такому процесі (тобто, опустивши деталі участі рієлтора, його винагороди, наявності документів на квартиру, виду договору оренди тощо), акцентуючи увагу виключно на математичній стороні питання формування ціни.

## 1.2. Поняття ринку нерухомості та його динаміка в Україні

Включаючи усе згадане у попередньому підрозділі, перейдемо до розгляду комерційних відносин щодо нерухомого майна. Поняття «ринок нерухомості» означає сукупність комерційних та правових стосунків навколо угод із об'єктами нерухомості, а саме: оренди, купівлі-продажу, іпотеки тощо.

Інакше кажучи, ринок нерухомості – це взаємозалежна система ринкових механізмів, що забезпечують створення, передачу та експлуатацію об'єктів нерухомості.

Серед визначень поняття «ринок нерухомості», що наводяться у наукових працях, ми можемо виділити визначення, наведене у роботі Дж. Фрідмана та Н. Ордуей, за якими, ринок нерухомості – це певний набір механізмів, за допомогою яких передаються права на власність і пов'язані з нею інтереси, встановлюються ціни і розподіляється простір між різними конкуруючими варіантами землекористування [2].

Економічними суб'єктами ринку нерухомості є:

1. Покупці (орендарі) – головні суб'єкти ринку нерухомості. Може виступати у вигляді юридичної або фізичної особи, а також як орган державного управління, що має право на дану операцію.
2. Продавці (орендодавці) – будь-яка юридична або фізична особа, що має право власності на нерухоме майно.
3. Професійні учасники ринку нерухомості – розгалужене поняття, яке регулюється переліком необхідних процесів у кожному окремому випадку.

Задля подальшого більш детального розгляду ринку нерухомості варто увести поняття попиту і пропозиції. У загальному випадку прийнято розглядати попит і пропозицію як єдину економічну модель, що відображає процес ціноутворення на ринку. При цьому попит і пропозиція є універсальними характеристиками будь-якого ринку. Попит – наявна на ринку потреба у товарах,

а пропозиція – кількість розглядаємого ринком товару або продукту, що є на цьому ринку або може бути на нього доставлена. Уперше термін «попит і пропозиція» використав Дж. Стюарт у 1767 році у своїй роботі «Принципи політичної економії» [8].

Тобто, попит відображає потребу та бажання покупця придбати певний товар чи послугу, а також, його спроможність сплатити за цей товар чи послугу за тою ціною, що знаходиться у межах доступного діапазону. З більш математичної точки зору величина попиту характеризує кількість товару чи продукту, який покупці здатні та бажають купити за можливими цінами. Також варт розглянути закон попиту. Закон попиту стверджує, що величина попиту зменшується по мірі збільшення ціни на товар, а значить, демонструє обернено пропорційну залежність між величиною попиту и ціною на товар.

Пропозиція, у свою чергу, характеризує бажання та можливість продавця пропонувати свої товари чи послуги на ринку за певними цінами. Знов таки, говорячи більш кількісно, величина пропозиції відображає кількість товарів чи послуг, які продавець бажає та здатен запропонувати на ринку для продажу за певними цінами протягом деякого проміжку часу. Як і величина попиту, величина пропозиції залежить не тільки від ціни, а від цілого ряду нецінових показників, серед яких можна виділити виробничі можливості, рівень цін на інші товари, споживчі смаки, мода, можливість заміщення товару іншими, інфляційні очікування тощо. Закон пропозиції стверджує, що за інших незмінних чинників величина пропозиції збільшується по мірі збільшення ціни на товар [10].

Проте, варто враховувати, що теорія попиту та пропозиції відображає теоретичну модель, та розраховані за нею ціни, є «розрахованими у вакуумі», тобто ідеальними та можливо недосяжними.

Ринок нерухомості також можна піддати класифікації. Наприклад, у залежності від юридичних прав на нерухомість, ринок нерухомості можна поділити на ринок купівлі та ринок продажу. За способами здійснення угод ринок



можна поділити на первинний та вторинний. Первинний ринок розглядає об'єкти, що вводимуться в експлуатацію вперше, тобто такими, що є новоствореними або приватизованими. Вторинний ринок розглядає об'єкти, що вже перебувають у експлуатації та є такими, що будуть переброданими або передані за допомогою іншої форми угоди до іншого власника. Варто зазначити, що первинний і вторинний ринки, являючись складовими загального поняття ринку нерухомості є взаємопов'язаними та прямо впливають один на одного. Наприклад, ціни вторинного ринку є важливим показником, що відображає наскільки рентабельним є будівництво нових споруд та будівель [22].

Ринок нерухомості України на сьогодні виділяє найважливішими сегментами такі як ринок житлової нерухомості та ринок нежитлової нерухомості й землі. До складу об'єктів, що є основними на ринку нерухомості України зараз входять квартири і кімнати багатоквартирних житлових будинків. Варто враховувати, що ціни на нерухомість постійно зростають із різних причин, чи то політичний стан чи штучне підвищення зі сторони посередників ринку.

Конкретно у цій роботі надалі буде розглянуто ринок житлової нерухомості для оренди, але для загального погляду на ситуації ринку в цілому розглянемо також і нерухоме житлове майно для купівлі та продажу. Експерти з оцінювання нерухомого майна у Києві стверджують, що за останній рік вартість квадратного метру житла на вторинному ринку столиці зросла з 46 700 грн. до 49 880 грн. (у середньому, +3 180 грн.). Що цікаво, то це підвищення яскраво відображається саме у гривні і підпорядковане неціновому чиннику (інфляція) збільшення ціни, бо у доларовому еквіваленті ціна за квадратний метр майже не змінилася. Загалом, експерти також говорять про те, що наразі найбільший попит на ринку купівлі-продажу саме на однокімнатні квартири, після них – на трикімнатні. Також варто зазначити, що у зв'язку із ситуацією, дуже зріс попит на заміські будинки.

Також варто окремо розглянути вплив повномасштабного вторгнення на ринок нерухомості в Україні в цілому та у Києві. Із початком війни практично усі забудовники призупинили свою діяльність, таким чином, у 2022 році загально площа нововведеного в експлуатацію житла становила 7,1 млн квадратних метрів, що на 38% менше, ніж у попередньому році. Крім того, надважливо врахувати кількість руйнувань унаслідок ракетних та артилерійських обстрілів. Це впливає на зміну вектору прикладення сили зі сторони забудовників, тепер замість побудови нових житлових будинків, їм треба відновлювати пошкоджені. За наявними даними, об'єми руйнувань перевищують кількість усього нововведеного в експлуатацію житла за останні сім років [19].

Попит наразі на первинному ринку становить лише 30% від довоєнних обсягів та приблизно 15% у Києві. Схожа ситуація проглядається і на вторинному ринку. Динаміка кількості укладених угод купівлі-продажу житла на вторинному ринку зображена на малюнку нижче.

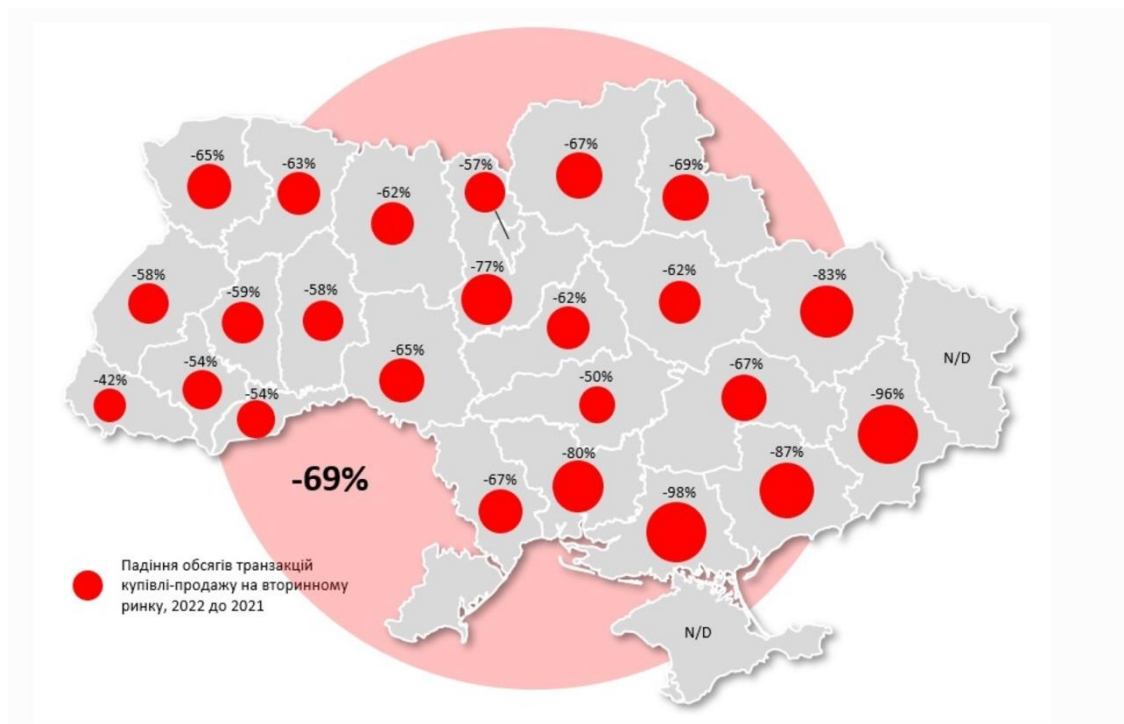


Рис. 1.1 – Динаміка кількості укладених договорів купівлі-продажу [19]

За дослідженням ЛУН, одного із найбільших сайтів для пошуку житла для купівлі-продажу та оренди у всій Україні, можемо побачити наступну статистику щодо зростання середньої вартості найдешевшої квартири у Києві за рік [20]:

### Зростання середньої вартості найдешевшої квартири у житловому комплексі у Києві за рік

|            |              |      |
|------------|--------------|------|
| 1-кімнатна | 2.33 млн грн | +66% |
| 2-кімнатна | 3.26 млн грн | +70% |
| 3-кімнатна | 4.4 млн грн  | +76% |

 лун ціни від забудовників при 100% оплаті

Рис. 1.2 – Динаміка цін на квартири у Києві [21]

Іншою є ситуація з орендою житла. На ринку оренди в Києві, середня ціна зданих у листопаді однокімнатних квартир менша на 2000 грн за середню по ринку, про що свідчать статистичні дані ЛУН:

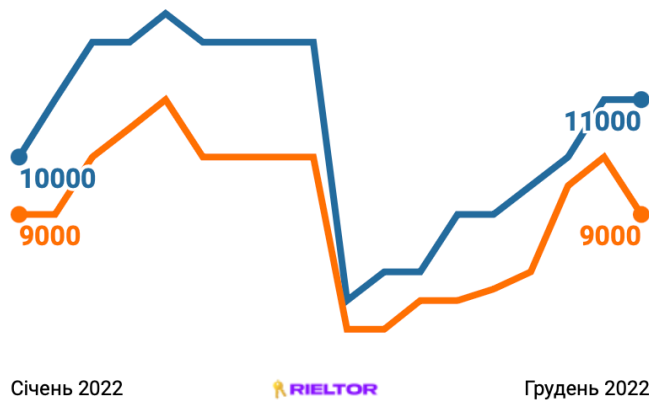


Рис 1.3 – Графік змінення ціни на оренду житла у Києві [20]

При цьому інтерес до довгострокової оренди повністю відновився у липні 2022:

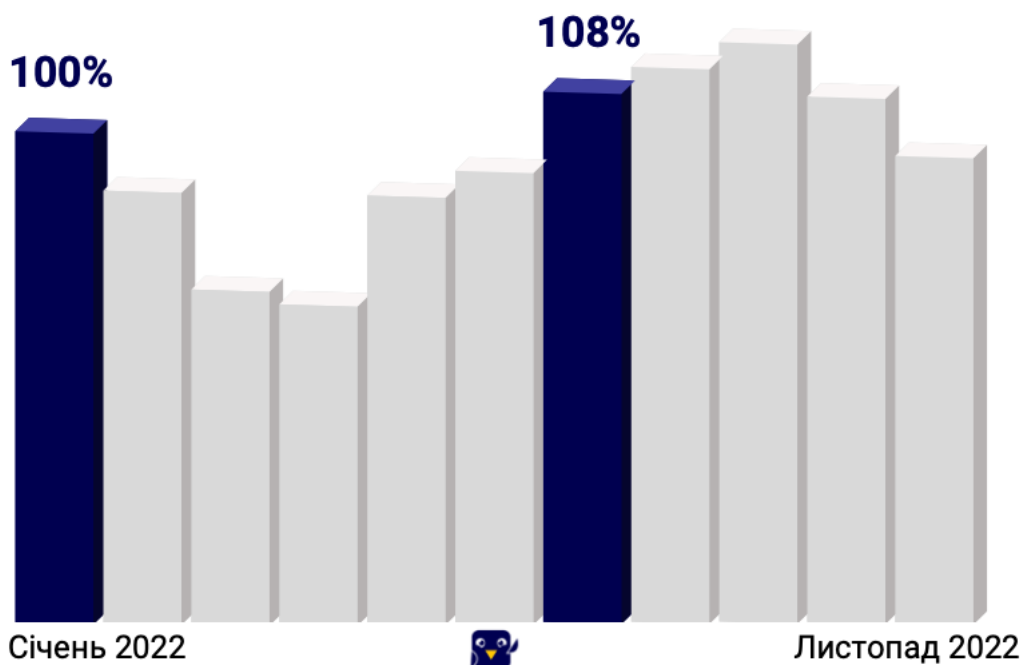


Рис. 1.4 – Діаграма зацікавленості людей щодо довгострокової оренди [20]

Загалом, можна зробити висновок про значне зменшення об'ємів ринку нерухомості по всій Україні та у місті Києві. Наразі житло стало менш доступним у той час, як потреба у ньому зросла. Враховуючи, майже повне зупинення введення в експлуатацію новобудов, маємо значне зменшення інтересу до купівлі нерухомості. Можна констатувати відкладений попит на ринку, тобто, купують нерухомість відчутно менше, ніж цікавляться об'єктами. При цьому вартість квартир для купівлі значно збільшилася, у той час як ціна оренди, пройшовши через шлях суттєвого зменшення, помірно зросла.

### 1.3. Методи розрахунку вартості оренди нерухомого майна

Згідно із діючим законодавством України, орендна плата – це це фіксований платіж, тобто такий платіж, який орендар сплачує орендодавцю незалежно від наслідків своєї господарської діяльності (ч. 1 ст. 286 ГК) [11]. Обчислюється орендна плата за певний період часу: місяць, квартал, рік тощо. При укладенні договору оренди обидві сторони можуть керуватися нормами Закону № 2269 і Методикою № 786 (п. 4 ст. 1 Закону № 2269) [15]. За цими правовими актами можна визначити наступний порядок розрахунку орендної плати:

1. При оренді нежитлового приміщення, яке є частиною будівлі або споруди, орендна плата розраховується за наступною формулою:

$$P_i = \frac{P_{\#}}{S_{\#}} S_i,$$

де  $P_i$  – вартість орендованого приміщення, яке є частиною будівлі або споруди у грн;

$P_{\#}$  – вартість будівлі або споруди у цілому, визначена експертно;

$S_i$  – площа орендованого приміщення у квадратних метрах;

$S_{\#}$  – площа будівлі або споруди у цілому у квадратних метрах.

2. Розмір орендної плати за перший місяць оренди після укладення договору оренди визначається за формулою:

$$P_{\#} = \frac{P}{12} \frac{I_{\#}}{I_i},$$

де  $P_{\#}$  – орендна плата за перший місяць оренди у гривні;

$P_{\#}$  – орендна плата за рік оренди у гривні;

$I_{\#}$  – індекс інфляції за перший місяць оренди;

де  $P_{\#}$  – орендна плата за перший місяць оренди у гривні;

$P_{\#}$  – орендна плата за рік оренди у гривні;

$I_{\#}$  – індекс інфляції за перший місяць оренди;

$I_t$  – індекс інфляції за період з дати проведення експертної оцінки до дати укладення договору оренди або перегляду розміру орендної плати.

3. Розмір орендної плати за кожний наступний місяць оренди визначається коригуванням розміру орендної плати за попередній місяць на індекс інфляції за поточний місяць.

Якщо сторони не погоджуються розраховувати ціну за наведеною методикою, то розмір і склад орендної плати вони мають визначати за взаємною угодою (ч. 1 ст. 762 ЦК) [15]. Тобто, фактично, це означає, що ціна може бути розрахована і встановлена орендодавцем будь-яким методом, головне – щоб вона була погоджена із орендарем та зазначена у договорі. При розрахунку вартості оренди орендодавець зазвичай враховує наступні показники:

- витрати на обслуговування об'єкта оренди (наприклад, працівникам, що охороняють або прибирають орендоване приміщення);
- частка прибутку, яку орендодавець бажає отримати від здачі нерухомості в оренду;
  - сума, що може відшкодувати збитки, спричинені часом, на який нерухоме майно передається в оренду.

#### 1.4. Висновки до розділу 1

У цьому розділі досліджено поняття нерухомості, наведена кваліфікація об'єктів нерухомості за різними ознаками, поняття ринку нерухомості, проаналізовано взаємозв'язки, що його формують та взаємозв'язки, що є ключовими. Розглянуто динаміку ринку нерухомості саме на території України за минулі роки, сьогоdnішні стан ринку, попит та пропозицію та впоdовання споживачів. Також детальну увагу приділено методам розрахунку вартості оренди житла та чинникам, що мають прямий вплив на її формування в цілому.

## 2.1. Методи та моделі прогнозування

Прогнозування, як процес, що дозволяє отримати результати про стан певного об'єкта чи явища на основі системної оцінки його якісних та кількісних характеристик, широко використовується у багатьох галузях [6]. Наприклад, бізнес-планування, прогноз погоди, передбачення землетрусів та вулканічної активності, прогнозування розвитку економіки певної країни або передбачення можливих результатів політичних виборів тощо. Прогнозування являє собою спосіб передбачення, сформований на основі даних, що певним чином розподілені у часовому проміжку. Щоб побудувати гідну систему прогнозування важливо опанувати існуючі методи та моделі прогнозування, а також методи оцінки побудованого прогнозу та їх надійність [7].

Методи прогнозування поділяються на кількісні та якісні. Кількісні методи базуються на математичних моделях та історичних даних. У той час, як якісні спираються на професійну думку, досвід та інтуїцію. Розповсюдженими якісними методами побудови прогнозу є думка журі, модель очікування споживача та метод Делфі. Якісні методи поступаються операційними можливостями та високими точностями, що можуть забезпечити при побудові прогнозу кількісні методи, але при цьому, мають і свої переваги. Якісні методи допомагають включати у дослідження людський фактор та емпатію, а отже, знаходити та аналізувати достатньо специфічні взаємозв'язки між даними, що, скоріш за все, не будуть виявлені при застосуванні математичних моделей [3]. Спираючись на ті ж чинники, можна стверджувати, що деякі прогнози настільки сильно потребують людського зауваження та експертної думки, що не можуть бути піддані лише математиці. Наприклад, вищезгаданий метод Делфі, також відомий як метод експертних оцінок, дозволяє за допомогою проведення багатоетапної оцінки суджень експертами та, у деяких випадках, проведення брейнштормів між ними, дозволяє дійти консенсусу або виділити особливі



рішення певних проблем. Можна сказати, що велика кількість компаній, так чи інакше, користується його частинами при побудові бізнес-планів або проведенні ретроспектив [4].

У той же час, кількісні методи, що мають під собою сильне технічне підґрунтя та велику операційну могутність, передбачають побудову моделей, які можна поділити на дві групи: причинно-наслідкові (або ж моделі предметної області) та моделі часових рядів.

Моделі предметної області базуються на законах досліджуваної предметної області. Наприклад, закони загальної фізики та термодинаміки використовуються при побудові прогнозу погоди, а диференціальні рівняння при передбаченні кількісної зміни популяції видів. Таким чином, ці моделі описуються рівняннями і рішеннями поставлених проблем або шуканих прогнозів будуть розв'язки цих самих рівнянь.

Моделі часових рядів будуються за принципом знаходження залежності майбутніх значень від минулих відносно часових проміжків усередині самого процесу та побудови прогнозу на основі віднайденого взаємозв'язку. Перевагою даних моделей є універсальність їх застосування до різних предметних областей, так як їх загальний вигляд не потерпає змін від природи даних, а часовий ряд від них не залежить та їх не враховує. У залежності від виду взаємозв'язку між даними та прогнозованим значенням, серед моделей часових рядів виділяють структурні та статистичні. Статистичні моделі виражають залежність майбутнього значення від попередніх у часі за допомогою рівнянь. До них належать регресійні та авторегресійні моделі (ARIMAX, ARDLN), моделі експоненційного згладжування і т.д. Структурні моделі, наприклад, нейромережеві або побудовані на основі ланцюгів Маркова тощо, виявляють залежність прогнозованого значення від наданих даних у вигляді певної структури та чіткого алгоритму із правил переходу по ній.

Надалі, у рамках даної роботи, ми розглянемо існуючі кількісні та якісні методи прогнозування та виділимо найбільш ефективний метод кожного виду для нашого випадку. Опісля, побудуємо модель кількісного аналізу та проведемо

якісний аналіз даних, що з тих чи інших причин не можуть бути враховані у обраному кількісному методі. Це дозволить нам всебічно розглянути питання роботи, порівняти результати та ефективність застосування тих чи інших методів у нашому випадку, а також, спираючись на результати обох методів, зробити більш глибокий та обґрунтований висновок щодо вирішення поставленої задачі.

## 2.2. Методи кількісного аналізу. Модель лінійної регресії

У цьому розділі розглянемо методи кількісного аналізу. Як уже було зазначено вище, найбільш часто використовуваними кількісними методами є регресійні, авторегресійні моделі (ARIMAX, ARDLN), моделі експоненційного згладжування, нейромережеві моделі, структурні моделі побудовані на основі ланцюгів Маркова тощо. Для звуження вибору підходящої моделі було вирішено зупинитися на статистичних моделях, так як статистичні моделі доцільно використовувати у випадках, коли при врахуванні випадкових процесів, динаміка розглядаємої системи підкоряється не динамічним законам, а статистичним. Це дозволяє звужити і водночас розширити погляд на аналіз системи, так як статистичний підхід дозволяє враховувати та досліджувати питання вірогідності стосовно багатьох характеристик системи. При дослідженні наявних моделей, враховуючи фактори впливу та характеристики нашого питання, нами було обрано модель лінійної регресії.

Для подальшого розгляду саме цієї моделі уведемо поняття регресії. Регресія – це одnobічна залежність між випадковими змінними, тобто деяка функція, що виражає взаємозв'язок залежної шуканої змінної від незалежної або декількох незалежних змінних. Кажучи ближче до побудови, регресія – це спосіб

вибору такої функції, що мінімізує функцію втрат. Функція втрат (англ. «loss function») у той же час характеризує наскільки сильно побудована математична функція буде розходитися зі значеннями у заданих точках. Регресію можна розглядати як «інтерпольовану апроксимацію», так як метод полягає у проведенні кривої якомога ближче до заданих точок, але при цьому без виходу помилки за задані межі та збереженні простоти задля виявлення спільної тенденції. У цьому і полягає суть функції втрат – збереження рівномірного балансу між апроксимацією та інтерполяцією.

Метод лінійної регресії уперше було опубліковано Лежандром у 1805 році, хоча, стверджується, що Гаусс використовував його раніше, обчислюючи орбіту карликової планети Церери [1].

Модель лінійної регресії – різновид регресійної моделі, що відображає залежність вихідної змінної від вхідної або декількох вхідних змінних. Саме поняття лінійної регресії означає, що сімейство функцій, із якого буде здійснено вибір функції, являє собою лінійну комбінацію наперед заданих базисних функцій, але не пов'язане із самим базисними функціями, вони можуть бути як лінійними, так і ні.

Парна (або ж проста) лінійна регресія дозволяє розглядати зв'язок між однієї вхідною незалежною та однієї вхідною залежною змінними, застосовуючи функцію прямої:  $y = kx + b$ . Така пряма називається лінією регресії, її зображення графічне представлення наведено нижче. Параметрами моделі у цьому випадку будуть коефіцієнти  $k$  і  $b$ .

Більш розповсюдженим випадком є множинна лінійна регресія, що дозволяє встановити лінійну залежність між множиною вхідних незалежних змінних та однією вихідною залежною змінною. Побудувати множинну лінійну регресію можна наступним чином:

$$y = a_0x_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n + b,$$

де  $x_i$  – змінні,  $a_i$  – коефіцієнти, а  $b$  – зміщення.

Тепер, спираючись на усі наведені вище дані, задаймо модель лінійної регресії та заглибимося у розуміння її процесів із точки зору математичної статистики. Для цього відійдемо від теоретичних понять та задаймо модель наступним чином:

$$\vec{y} = X\vec{a} + \varepsilon,$$

де:

- $\vec{y} \in \mathbb{R}^l$  – цільова змінна;
- $\vec{a}$  – вектор параметрів моделі (при побудові моделей у машинному навчанні їх також часто називають вагами);
- $X$  – матриця спостережень;
- $\varepsilon$  – випадкова змінна, що відповідає за випадкову непрогнозовану помилку моделі.

Для того, щоб регресійна модель була саме лінійною, ми маємо накласти на неї ряд певних обмежень (спираючись при цьому на певну теорему, формулювання та необхідність застосування якої буде надано пізніше):

- математичне сподівання випадкових помилок має строго дорівнювати нулю:  $\forall i: \mathbb{E}(\varepsilon) = 0$ ;
- має виконуватися умова гомоскедастичності (незалежності дисперсії випадкових величин від номеру спостереження), інакше кажучи, дисперсія випадкових помилок має бути однаковою та скінченною:  $\forall i: \mathcal{D}(\varepsilon_i) = \sigma^2 < \infty$ ;
- випадкові помилки мають бути не корельованими:
- $\forall i \neq j: \text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ .

Також, при розгляді моделі саме лінійної регресії, розв'язком задачі пошуку оптимальних її параметрів має бути саме лінійна оцінка. Уведемо поняття лінійної оцінки, а разом із ним, незміщеної.

Оцінка  $\hat{a}_k$  параметрів  $a_k$  називається лінійною, якщо:

$$\hat{a}_k = a_{k0}y_{k0} + a_{k1}y_{k1} + \dots + a_{kL}y_{kL},$$

де  $\forall k$   $a_{kj}$  залежить тільки від даних спостережень і майже напевно не є лінійним.

Оцінка  $\hat{a}_k$  називається незміщеною, якщо математичне сподівання цієї оцінки дорівнює дійсному, але невідомому значенню оцінюваного параметру:  
 $E(\hat{a}_k) = a_k.$

Одним із методів обрахувати значення параметрів моделі є метод найменших квадратів (далі МНК). МНК дозволяє мінімізувати середньоквадратичну похибку між дійсним значенням залежної змінної та прогнозованим, яке є результатом роботи моделі. Описати його математично можна наступним чином:

$$\begin{aligned}\mathcal{L}(X, \vec{y}, \vec{a}) &= \frac{1}{2n} \sum_{i=1}^n (y_i - a^T x_i)^2 = \\ &= \frac{1}{2n} \|\vec{y} - X\vec{a}\|^2 = \frac{1}{2n} (\vec{y} - X\vec{a})^T (\vec{y} - X\vec{a}).\end{aligned}$$

Щоб розв'язати дану задачу, потрібно обрахувати похідні за параметрами моделі, прирівняти їх до нуля та розв'язати отримані рівняння відносно  $\vec{a}$ .

$$\frac{\partial \mathcal{L}}{\partial \vec{a}} = \frac{\partial}{\partial \vec{a}} \frac{1}{2n} (\vec{y}^T \vec{y} - 2\vec{y}^T X\vec{a} + \vec{a}^T X^T X \vec{a}) = \frac{1}{2n} (-2X^T \vec{y} + 2X^T X \vec{a});$$

$$\frac{\partial \mathcal{L}}{\partial \vec{a}} = 0 \Leftrightarrow \frac{1}{2n} (-2X^T \vec{y} + 2X^T X \vec{a}) = 0;$$

$$-X^T \vec{y} + X^T X \vec{a} = 0 \Leftrightarrow X^T X \vec{a} = X^T \vec{y};$$

$$\vec{a} = (X^T X)^{-1} X^T \vec{y}.$$

Згадаємо спрощене формулювання теореми Гаусса-Маркова, використання якої надалі буде доцільним. Дана теорема розглядає модель лінійної регресії із певними накладеними умовами (що були наведені вище) і стверджує, що за таких обмежень оцінка МНК є кращою оцінкою параметрів моделі серед усіх лінійних та незміщених оцінок, тобто такою, що має найменшу дисперсію. Також дана теорема постулює достатньо очевидний, але важливий

факт: якщо розташування випадкових точкових спостережень буде зосереджене вздовж експоненти, логарифму чи будь-якої іншої нелінійної функції, то будувати саме

лінійне рівняння регресії у цьому випадку є недоцільним. Інакше кажучи, наявність однакової функціональної форми означає наявність однакової функціональної залежності.

Серед інших методів оцінювання також варто виділити метод найменших модулів та метод максимальної вірогідності. Перший є значно менш чутливим до викидів значень, ніж МНК, що є його перевагою, але може мати більше, ніж один розв'язок та не має простої формули обчислення оцінки, що є серйозним недоліком. Метод максимальної вірогідності має місце коли відомі усі розподіли відхилень для усіх спостережень, а за умови нормальності таких відхилень, приводить до того ж результату що і метод найменших квадратів. Усе це, враховуючи теорему Гауса-Маркова, цілком обґрунтовує вибір методу найменших квадратів.

Успішність побудованого прогнозу можна оцінити, оцінивши величину похибки даного прогнозу. Знову наведемо певні викладки, що дозволять нам розкласти дану похибку на дисперсію та зсув (англ. «bias–variance decomposition», у перекладі – компроміс зсуву та дисперсії).

Нехай виконуються наступні твердження:

1. дійсне значення цільової змінної складається із деякої визначеної функції  $f(\vec{x})$  та випадкової помилки  $\varepsilon$ :  $y = f(\vec{x}) + \varepsilon$ ;
2. при цьому значення цільової змінної є нормально розподіленим:  $y \sim \mathcal{N}(f(\vec{x}), \sigma^2)$ ;
3. випадкова помилка також є нормально розподіленою із центром у точці нуль:  $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ ;
4. наближення невідомої, але визначеної функції  $f(\vec{x})$  відбувається лінійною функцією  $f^N(\vec{x})$ , котра є точковою оцінкою наближуваної функції у просторі сімейства функцій, тобто випадковою змінною, що має середнє значення та дисперсію.

Тобто, нам треба знайти таку функцію  $f^N(\vec{x})$ , яка якомога краще наближує функцію  $f(\vec{x})$ , при цьому вимірюючи середньоквадратичну похибку у відносно  $f^N(\vec{x})$  ми прагнемо, щоб вона була мінімальною, як для точок нашої вибірки, так і за її межами.

Тоді, враховуючи, усе вищезгадане, помилку  $\vec{x}$  можна розкласти наступним чином:

$$\begin{aligned} Err(\vec{x}) &= \mathbb{E}YZy - fV(\vec{x})^{\$} \setminus = \\ &= \mathbb{E}(y^{\$}) - 2\mathbb{E}Zy fV(\vec{x})^{\$} \setminus + \mathbb{E}YZ fV(\vec{x})^{\$} \setminus = \\ &= (y^{\$}) - 2\mathbb{E}GyfV_H + f^{\$}N_H. \end{aligned}$$

Спростимо, застосовуючи формулу визначення дисперсії через матсподівання:  $\mathcal{D}(x) = \mathbb{E}(x^{\$}) - \mathbb{E}(x)^{\$}$ . Розпишемо два члени рівності, використовуючи цю формулу та властивості матсподівання і дисперсії:

$$(y^{\$}) = \mathcal{D}(y) + \mathbb{E}(y^{\$}) = \sigma^{\$} + f^{\$};$$

$$\mathbb{E}Gf^{\$}N_H = \mathcal{D}(f^{\$}N_H) + f^{\$}N_H^{\$}.$$

Третій член рівності розпишемо, спираючись на той факт, що помилка та цільова змінна є незалежними одна від одної величинами:

$$\mathbb{E}Gyf^{\$}N_H = \mathbb{E}G(f + \varepsilon)f^{\$}N_H = \mathbb{E}Gf^{\$}N_H + \mathbb{E}G\varepsilon f^{\$}N_H = f^{\$}N_H + (\varepsilon) \mathbb{E}Gf^{\$}N_H = f^{\$}N_H.$$

Отже, маємо:

$$\begin{aligned} Err(\vec{x}) &= \mathbb{E}YZy - fV(\vec{x})^{\$} \setminus = \mathbb{E}(y^{\$}) - 2\mathbb{E}GyfV_H + f^{\$}N_H \\ &= \sigma^{\$} + f^{\$} - 2f\mathbb{E}GfV_H + f^{\$}N_H^{\$} = \end{aligned}$$



$$\begin{aligned}
 &= \sigma^2 + Z f - \mathbb{E} G f^{VH} + \mathcal{D} G f^{VH} = \\
 &= \sigma^2 + Bias G f^{VH} + \mathcal{D} G f^{VH}
 \end{aligned}$$

Із отриманого виразу можемо зробити висновок, що помилка прогнозу будь-якої моделі вигляду  $y = f(\vec{x}) + \varepsilon$  складається з наступних елементів:

- незнижуваної похибки:  $\sigma^2$  – оскільки усі три члени є невід’ємними,

формує обмеження знизу для математичного сподівання похибки;

- квадрату зсуву:  $Bias G f^{VH}$  – середня похибка за вибіркою (набором даних), спричинена спрощувальними припущеннями (наприклад, припущення лінійності);

- дисперсії:  $\mathcal{D} G f^{VH}$  – варіативність помилки, тобто, наскільки помилка буде відрізнятися при навчанні моделі на різних наборах даних.

Чим складнішою буде побудована модель (наприклад, при великій кількості вільних параметрів), тим більше точок вона охоплюватиме і тим меншим буде зсув. Проте, ця складність збільшуватиме дисперсію, тому на практиці, при побудові моделі, доводиться балансувати між цими двома значеннями.

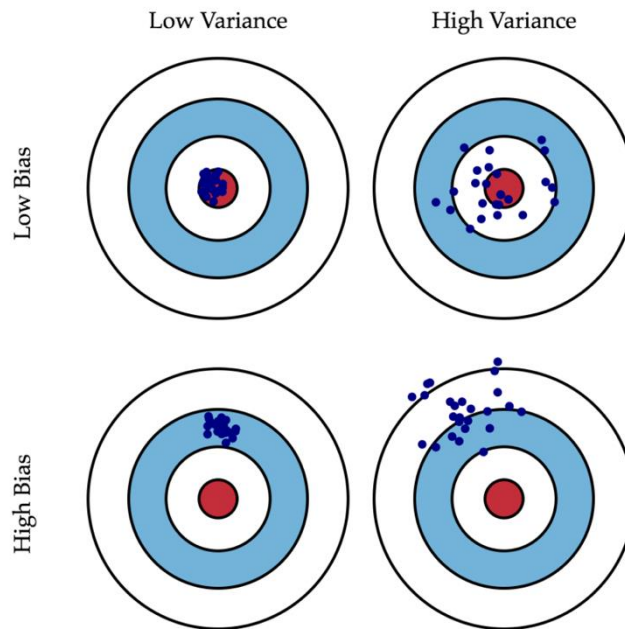


Рис. 2.1 – Візуальне зображення компромісу зсуву та дисперсії [7]

Через те, що тренувальний набір даних не узагальнюється, а повністю запам'ятовується моделлю, навіть невеликі зміни можуть призвести до неочікуваних результатів (перенавчання). Якщо ж модель буде мати недостатню складність, то вона не зможе виявити та запам'ятати закономірність, що призводить до навчання моделі іншому рішенню, зміщеному відносно правильного. Візуальне зображення залежності складності моделі від зсуву зображено за рис.2.2

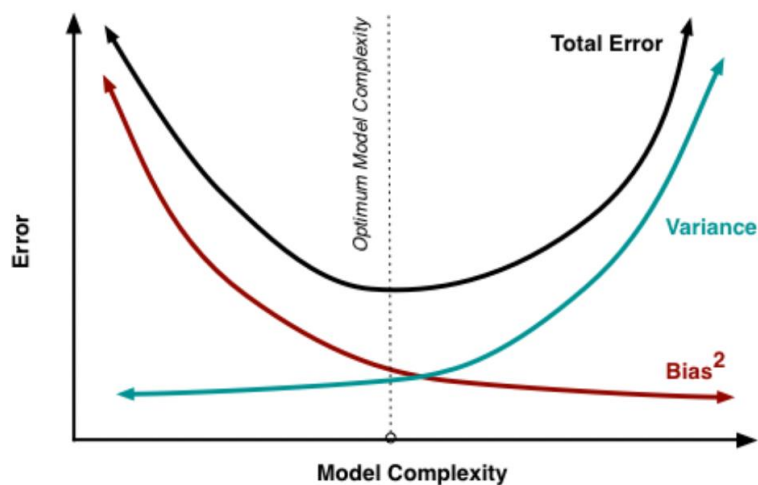


Рис. 2.2 – Візуальне зображення залежності складності моделі від зсуву [7]

Щоб оцінити наскільки підвищення складності моделі підвищує її точність, а також якісність побудованого моделлю прогнозу, можна використовувати ряд відомих метрик. Кожна з них має свої особливості застосування. Розглянемо їх детальніше [1].

Середньоквадратична помилка (англ. «Mean Squared Error» - MSE) застосовується у випадках, коли треба обійти саме великі помилки, що можуть бути допущені моделлю. За рахунок квадратичної залежності ці великі помилки стають набагато помітнішими. Середньоквадратична помилка розраховується за наступною формулою:

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

де  $n$  – кількість спостережень, за якими будується модель,  $y_i$  – значення залежної змінної для кожного такого спостереження,  $\hat{y}_i$  – значення прогнозованої змінної, передбачене моделлю.

Головний недолік MSE полягає у тому, що якщо на одному або декількох прикладах будуть фігурувати аномальні значення, то це призведе до великих помилок, на основі яких можуть бути зроблені невірні висновки про роботу моделі в цілому. З іншого боку, якщо модель не буде мати великих помилок взагалі, то це може призвести до зворотного результату – недооцінювання слабкості побудованої моделі.

Корінь із середньоквадратичної помилки (англ. «Root Mean Squared Error» - RMSE) розраховується як квадратний корінь із MSE:

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

Так як корінь є неспадною функцією, то MSE та RMSE можна мінімізувати, застосовуючи один і той самий функціонал. Різниця між ними полягає у тому, що вплив кожної помилки на RMSE є пропорційним величині квадрату цієї помилки. Саме тому, великі помилки викликать непропорційно великий вплив на RMSE, а значить, RMSE можна назвати чутливою до аномалій.

Середня абсолютна помилка (англ. «Mean Absolute Error» - MAE) розраховується як середнє значення абсолютних різниць між спостережуваними та прогнозованими даними за наступною формулою:

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

На відміну від MSE та RMSE, MAE є лінійною оцінкою, що означає приблизно однакову вагомість помилок. MAE широко застосовується у сферах, де помилка має інтерпретуватися відносно інших помилок саме лінійним чином, а не квадратично, як у попередніх, розглянутих метриках.

Середня абсолютна відсоткова помилка (англ. «Mean Absolute Percentage Error» - MAPE) є дуже простою в інтерпретації та розраховується наступним чином:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{|y_i|}$$

Перевагою цієї метрики є відсутність у неї розмірності, що дозволяє інтерпретувати її як у відсотках, так і у долях. Також, зручністю є те, що отриманий результат метрики одразу виражає відхилення від фактичного значення.

Підсумовуючи усю наведену інформацію, виділимо недоліки та переваги моделі лінійної регресії. Недоліками можуть бути:

- ☐ модель лінійної регресії передбачає виключно лінійну залежність між залежною та незалежною (або незалежними) змінними, що може бути досить нереалістичним у багатьох сценаріях, що звужує можливість використання моделі;
- ☐ модель може бути чутливою до «викидів», тобто, надто високих або низьких значень окремих точок, що можуть вплинути на параметри моделі, зсунувши отримуваний результат від правильного;
- ☐ також модель лінійної регресії може бути недостатньою для опису взаємозв'язків між змінними, у випадках, коли залежність між ними є складною або нелінійною.

У той же час варто зазначити переваги моделі лінійної регресії:

- ☐ модель лінійної регресії є достатньо простою у використанні та інтерпретації;
- ☐ модель може бути використана для виявлення статистичної залежності між змінними та оцінки впливу кожної змінної на залежну змінну;
- ☐ модель можна легко модифікувати для включення додаткових змінних, що можуть підвищити точність побудованих за допомогою моделі, прогнозів.

Усі перелічені вище метрики достатньо нелегко інтерпретувати, так як вони візуалізують саме значення помилки, що, без зазначення контексту, є складним для розуміння, наскільки ця помилка є прийнятною. Тому, досить доцільно оцінювати не значення помилок, а величину, що відображала б

наскільки побудована модель прогнозує дані краще, ніж модель, яка мала б тільки константні значення, або нульові коефіцієнти регресії.

Яскравим прикладом такої помилки є коефіцієнт детермінації (англ. «Coefficient of determination» або ж «R-квадрат»), який відображає частку дисперсії залежної змінної за наступною формулою:

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

Головною перевагою цієї метрики перед метриками, що оцінюють значення помилок є його інваріативність. Крім того, дана метрика лежить у діапазоні від « $-\infty$ » до «1». Значення, що будуть близькими до одиниці, кажуть про високу відповідність моделі до даних. Якщо ж значення метрики наближається до нуля, то це говорить про низьку відповідність моделі даним, а так як ми порівнюємо її роботу із роботою моделі, що основана на константах, то, на жаль, це означало б, що наша модель працює не краще, ніж працювала б із константами. За від'ємних значень метрики можемо зробити висновок про роботу моделі гіршу, ніж передбачення за середнім значенням.

Підсумовуючи, модель лінійної регресії є корисним інструментом для аналізу даних та передбачення результатів на основі вхідних даних, але перед використанням моделі важливо оцінити її обмеження та відповідність вимог. Беручи до уваги питання, що розглядається у цій роботі, дані, на яких будуватиметься модель, можна вважати лінійно залежними та представленими у числовому вигляді, а, отже, такими, на яких модель має добре відпрацювати показавши гідні результати.

### 2.3. Методи якісного аналізу. Метод морфологічного аналізу

Розглянемо якісні методи прогнозування. Серед них здебільшого виділяють такі методи як думка журі, модель очікування споживачів та метод експертних оцінок. Якщо при застосуванні методу думки журі судження експертів за допомогою обговорення зводяться до узагальненої думки (що має наразі широке використання у різних компаніях під назвою «мозковий штурм»), то при застосуванні моделі очікувань споживачів навпаки – проводиться опитування клієнтів, що можуть потенційно бути чи вже є зацікавленими у

питанні, що розглядається. Метод експертних оцінок полягає в анкетному опитуванні спеціально відібраних експертів. Основними опорними точками при складанні прогнозу за допомогою застосування якісних методів є:

- необхідність встановлення чіткої цілі прогнозування до початку побудови моделі;
- завчасне розуміння та виділення обмежень, що прямо вплинуть на точність побудованого прогнозу;
- виділення та розподілення за рівнями (частіше, складності прийняття) можливих рішень питання, що розглядається;
- створення можливості якісної оцінки прогнозу, враховуючи зміни та фактори впливу, що можуть прямо вплинути на отримані результати.

У рамках виконання нашої роботи у нас не було змоги здобути експертні оцінки вищепереліченими методами. Тому, для прогнозування вартості оренди житла якісними методами нами було обрано морфологічний аналіз, основного частиною процедури якого є якісний аналіз.

Морфологічний аналіз – яскравий приклад системного підходу у винахідництві. Метод було розроблено швейцарським вченим Ф. Цвіккі 1969 році. Завдяки цьому, Цвіккі зміг за досить короткий проміжок часу отримати патерни на інші винаходи в області будівництва ракет.

Метод морфологічного аналізу дозволяє як і знайти оригінальні ідеї, там де необхідна присутність людського фактору, так і проаналізувати цілий ряд рішень наявної проблеми, оцінити її фактори ризику, методи запобігання або уникнення, надавши у кожному випадку вірогідності усіх альтернатив [16]. Взагалі, головними цілями застосування методу можна вважати наступні:

- відбір параметрів, що критично та найбільше впливають на розв’язання поставленої задачі;
- системний аналіз усіх можливих рішень поставленої проблеми, що витікають як із технічних характеристик досліджуваної системи, побудованої на

базі відібраних параметрів, так і дослідження взаємозв'язків цих параметрів та суперечностей між ними;

– реалізація послідовного алгоритму, що складається із операцій, які дозволяють дійти до певних рішень.

Ідея морфологічного аналізу полягає у знаходженні найбільшої кількості, а гранично й усіх можливих способів розв'язання обраної задачі процедурою комбінування основних структурних елементів систем, що дозволяє обрати найдоцільніші з них. Головною проблемою та етапом у побудові системи, що буде піддана морфологічному аналізу є формування, так званої, морфологічної скрині. Аспект проблеми полягає у виявленні параметрів та альтернатив, що відповідають поставленій задачі та спрямовані на вирішення поставленого питання. Головною ж перевагою методу можна назвати можливість розглядати не тільки існуючі параметри, а й створювати нові, комбінуючи певним чином існуючі.

Напрямки застосування морфологічного аналізу можуть бути дуже різними у залежності від задачі, що розглядається. Класичним варіантом застосування можна справедливо назвати пошук нових або виявлення найефективніших та кращих організаційних, соціальних, економічних та інших складних систем. За такого способу проводиться функціонально-елементний аналіз системи, у результаті якого формується морфологічна скриня.

Узагальнено, процес побудови морфологічної скрині покроково можна описати наступним чином:

1. чітке та точне формулювання поставленої проблеми або задачі;
2. визначення параметрів, які прямо впливають на розв'язки задачі;
3. розподіл відібраних параметрів за альтернативами або значеннями (формування множини значень кожного параметру);
4. побудова морфологічної скрині, кожен елемент якої є варіантом системи, що аналізується;



5. оцінка усіх наявних варіантів розв'язку задачі (побудова матриці взаємо узгодженості параметрів);

6. вибір найкращого варіанту розв'язку поставленої задачі із морфологічної скрині.

Приклад морфологічної скрині зображено на малюнку нижче:

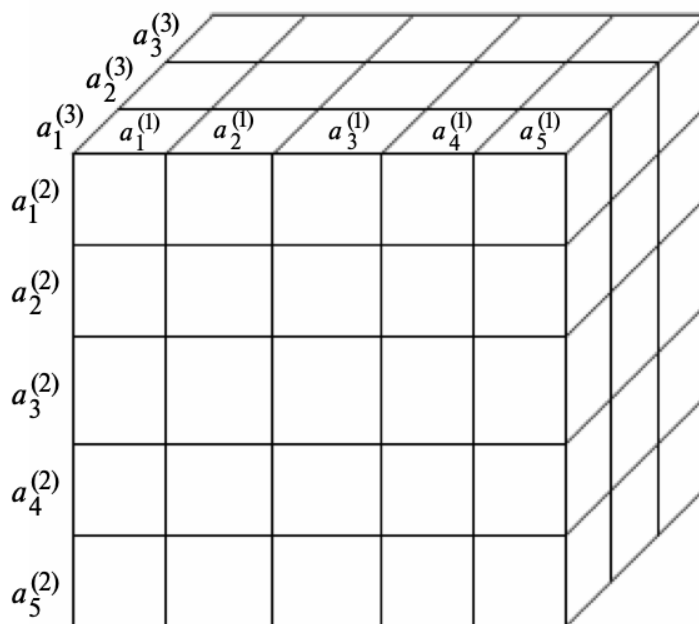


Рис. 2.3 – Зображення морфологічної скрині [16]

Вочевидь, її можна використовувати лише для рішень від одного до трьох вимірів, так як більша кількість вимірів є важкою для візуального зображення.

Приклад морфологічної таблиці, що відповідає морфологічній скрині:

| Параметри задачі |             |             |
|------------------|-------------|-------------|
| $F_1$            | $F_2$       | $F_3$       |
| $a_1^{(1)}$      | $a_1^{(2)}$ | $a_1^{(3)}$ |
| $a_2^{(1)}$      | $a_2^{(2)}$ | $a_2^{(3)}$ |
| $a_3^{(1)}$      | $a_3^{(2)}$ | $a_3^{(3)}$ |
| $a_4^{(1)}$      | $a_4^{(2)}$ | —           |
| $a_5^{(1)}$      | $a_5^{(2)}$ | —           |

Рис. 2.4 – Приклад морфологічної таблиці [16]

Якщо із кожного стовпчику даної таблиці обирати по одному значенню, то це буде конфігурацією системи, що розглядається. Розрахувати кількість конфігурацій можна за наступною формулою:

$$M = \prod_{i=1}^N m_i$$

де  $N$  – кількість параметрів, а  $m_i$  – кількість значень  $i$ -го параметра.

Для проведення аналізу доцільно використовувати експертне оцінювання, так як розглянути та оцінити власноруч усі конфігурації є близьким до неможливого, бо їх кількість може сягати тисяч. Це і є тим аспектом морфологічного аналізу, де варто застосовувати якісні методи оцінки.

Для забезпечення нормального функціонування методу морфологічного аналізу треба дотримуватися певних правил побудови морфологічних таблиць:

- релевантність параметрів – необхідно, щоб параметр залежав або впливав на хоча б один інший параметр. Для забезпечення цієї умови треба, щоб хоча б одна з альтернатив одного параметру була пов'язана матрицею взаємозв'язків із принаймі однією альтернативою іншого параметру. За невиконання цього правила, утвориться незалежний параметр, що ніяк не буде впливати на результат, але збільшує кількість конфігурацій, тим самим збільшуючи об'єми та складність обчислень;
- взаємовиключність альтернатив – необхідно, щоб альтернативи кожного параметра були взаємовиключними. За невиконання цього правила, необхідно переформулювати множину альтернатив або параметрів;
- повнота множини альтернатив – необхідно, щоб множина альтернатив для кожного параметру була повною, щоб сукупність альтернатив параметра являла собою повну групу подій. За невиконання цього правила, можна доповнити групу подій, додавши альтернативу «інші» або «немає» тощо.

Подальша суть проведення аналізу після формування морфологічної скрині та проведення експертної оцінки є виявлення та оцінка взаємозв'язків між параметрами, що є можливим, так як параметри не є незалежними. Дана процедура має назву оцінювання взаємної узгодженості. Її основна відмінність

від багатьох схожих оцінок полягає у її небінарності. У нашому випадку кожній парі взаємопов'язаних значень проставляється не бінарне значення «так» або «ні», а оцінка у межах від «-1» до «1», що напрямую характеризує ступінь узгодженості розглянутої пари параметрів за вибраного критерію. Серед очевидних значень крайніх меж можливої оцінки, варто зазначити, що оцінка «0» означає повну незалежність значень одне від одного, тобто вибір будь-якого з них ніяк не вплине на вибір іншого. Такі оцінки зводяться у таблицю, яка називається матрицею взаємної узгодженості параметрів. Приклад такої матриці наведемо нижче:

|       |             | $F_1$       |             |             |             |             | $F_2$       |             |             |             |             |
|-------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|       |             | $a_1^{(1)}$ | $a_2^{(1)}$ | $a_3^{(1)}$ | $a_4^{(1)}$ | $a_5^{(1)}$ | $a_1^{(2)}$ | $a_2^{(2)}$ | $a_3^{(2)}$ | $a_4^{(2)}$ | $a_5^{(2)}$ |
| $F_2$ | $a_1^{(2)}$ |             |             |             |             |             |             |             |             |             |             |
|       | $a_2^{(2)}$ |             | 0,2         |             |             |             |             |             |             |             |             |
|       | $a_3^{(2)}$ |             | 0,2         |             |             |             |             |             |             |             |             |
|       | $a_4^{(2)}$ |             |             |             | -0,8        |             |             |             |             |             |             |
|       | $a_5^{(2)}$ |             |             |             |             |             |             |             |             |             |             |
| $F_3$ | $a_1^{(3)}$ | -1          | -1          |             |             |             |             | -1          | -0,5        |             |             |
|       | $a_2^{(3)}$ |             |             | 0,5         | 0,2         |             |             |             |             |             |             |
|       | $a_3^{(3)}$ |             |             |             |             |             |             |             |             |             | 0,8         |

Рис. 2.5 – Приклад матриці взаємної узгодженості [16]

Можна розглядати лише половину таблиці, так як вона є симетричною, що логічно, враховуючи взаємність взаємозв'язків між параметрами.

У випадках, коли постає задача побудови прогнозування, як у нашій роботі, проблема власне синтезу складної системи виникає не одразу. Варто зазначити, що у цьому випадку дослідження починається із детального розгляду поставленої задачі, моделювання та аналізу можливих подій та аналіз поведінки або можливої зміни стану системи, побудованої на основі початкового аналізу поставленої задачі. Метод морфологічного аналізу є гнучким та універсальним інструментом, що може бути застосований у задачах передбачення, проте, як вже зазначалося вище, важливо чітко та точно розуміти та формулювати цілі дослідження, та не менш чітко виділяти які саме вхідні дані є необхідними для проведення такого дослідження [23]. Такий підхід обумовлює якісний та уважний відбір параметрів. Розглянемо можливі типи параметрів морфологічної таблиці.

1. Дихотомічні – параметри, альтернативи яких виражають наявність або відсутність певного елемента або характеристики в об’єкта дослідження. Також такий параметр може бути відповіддю на логічне питання щодо об’єкта.

2. Кількісні – параметри, що описують певною величиною деяку характеристику досліджуваного об’єкта. Такі параметри можуть бути обмеженими та необмеженими, а також ділитися на підтипи за видом представлення: числові, вербальні та порівняльні.

3. Якісні – параметри, альтернативи яких якісно відрізняються одна від одної.

У випадках сценарного аналізу часто буває доцільним використання методу двоетапного морфологічного аналізу. Його суть полягає у тому, що на першому етапі відбувається аналіз неконтрольованих зовнішніх факторів, а на другому – синтез рішень, які найбільш доцільно та ефективно враховувати при аналізі системи, складеної на першому етапі. При цьому будуються дві взаємопов’язані морфологічні таблиці. Морфологічна таблиця для першого етапу – морфологічна таблиця сценаріїв, таблиця для другого – морфологічна таблиця стратегій. При такому підході на другому етапі морфологічного аналізу вибір альтернатив параметрів залежатиме не від зовнішніх факторів впливу та чинників, а виключно від особи (людини-аналітика, керівника, експерта тощо), яка буде приймати це рішення, тож, таким чином, ми уникаємо оцінки ймовірності вибору альтернатив серед усіх альтернатив. Замість цього використовується величина очікуваної результативності, тобто вірогідності того, що вибір саме цієї альтернативи призведе до бажаних результатів.

Одним морфологічний аналіз є гнучким та варіативним інструментом, що дозволяє аналізувати побудовану систему рішень для певної задачі чи об’єкта навіть в умовах невизначеності цього об’єкта. Причин для такої невизначеності об’єкта, що розглядається, може бути багато, розглянемо основні з них:

1. Відсутність точної та детальної інформації про стан об’єкта та його «насиченість» параметрами у даний момент. Такий фактор може бути спричинений у випадку, якщо, наприклад, об’єкт розглядається у майбутньому,

через певний проміжок часу, або ж після настання події, що впливатиме на характеристичні параметри цього об'єкта та на нього самого у цілому. Також причина відсутності повної інформації може бути пояснена нестачею даних із технічних причин, наприклад, за дослідження об'єктів, чіткі параметри яких не можуть бути визначені із технічної точки зору.

2. Необхідність системного розгляду множини конфігурацій об'єкта спричиняє виникнення невизначеності. Як вже зазначалося вище, за допомогою морфологічної таблиці будується множина характеристик або параметрів, складові яких є відомими, але чітке передбачення яка саме конфігурація альтернатив параметрів з'явиться при наступній реалізації - неможливе.

3. Необхідність прийняття рішення, яку саме конфігурацію треба обрати. У цьому випадку невизначеність виникає не через вплив зовнішніх чинників, тому саме цей вид невизначеності ускладнює або унеможлиблює у деяких випадках автоматичну побудову морфологічних таблиць.

Існують типи задач, для яких застосування двоетапного морфологічного аналізу може виявитися недостатнім. У такому випадку здійснюється аналіз за допомогою сукупності методів якісного аналізу, однієї зі складових якої є двоетапний метод морфологічного аналізу. Наприклад, якщо у задачі необхідно розглянути не тільки саму проблему, а її можливі наслідки, оцінку ризиків, підготовленості, методів запобігання. У цьому випадку для наявності усебічного зв'язку використовується мережа морфологічних таблиць.

При побудові мережі морфологічних таблиць підготовчий етап буде відрізнятися та складатися із наступних кроків:

1. Формування концепції роботи із мережею морфологічних таблиць, тобто визначення тих частин або варіацій об'єкту дослідження із якими буде проводитися аналіз.

2. Далі, спираючись на цю концепцію, будуються морфологічні таблиці для кожної частини, створюючи мережу.

3. Оцінювання взаємозв'язків не тільки всередині кожної таблиці, а і між ними.

Власне, таблиці, із яких складається мережа можна умовно поділити на три рівні:

1. Перший рівень складається із таблиці, що вміщує спільні для усіх частин досліджуваного об'єкта параметрів та спеціальних таблиць із відмінними власним параметрами для кожного випадку (частини, що окремо розглядається) досліджуваного об'єкта або задачі.

2. На другому рівні може міститися таблиця, що оцінює можливі наслідки або ризики, що можуть бути спричинені у рамках поставленої задачі.

3. На третьому рівні знаходиться таблиця, що дозволяє оцінити рівень підготовленості, можливі методи запобігання у рамках задачі або ж пом'якшити наслідки, що були оцінені на попередньому рівні.

#### 2.4. Висновки до розділу 2

У цьому розділі розглянуто можливості та методи побудови точного та якісного прогнозу. Увага приділена різниці між кількісними та якісними методами прогнозування, детально розглянуто переваги та недоліки багатьох з них.

Обрано методи для прогнозування у рамках цієї роботи. Детально розібрано та обгрунтовано з математичної точки зору побудову лінійної регресії, розглянуто сфери її застосування, оцінено переваги та недоліки у

порівнянні з іншими кількісними методами. Обрано та детально розглянуто підхід морфологічного аналізу, його можливості використання. Приділено увагу спеціалізованій літературі та науковим працям відповідних напрямків задля поглибленого вивчення предмету дослідження.

### 3 ПОБУДОВА МОДЕЛІ ЛІНІЙНОЇ РЕГРЕСІЇ ТА ПРОВЕДЕННЯ МОРФОЛОГІЧНОГО АНАЛІЗУ

#### 3.1. Прогнозування за побудованою моделлю лінійної регресії

Для побудови якісної та точної моделі лінійної регресії, варто підійти до цього питання точно та алгоритмічно, тому, будуючи модель, будемо дотримуватися наступного покрокового алгоритму:

1. Першочергово зберемо та підготуємо дані, на яких власне і буде навчатися та прогнозувати результати наша модель. Визначимо залежну змінну та незалежні змінні, що будуть на неї впливати. Перевіримо дані на наявність прогалин чи аномалій.

2. Власне, побудуємо модель, застосовуючи метод найменших квадратів, який детально теоретично було описано у попередньому розділі.

3. Оцінимо побудовану модель за допомогою різних метрик, які також були розглянуті вище.

4. За необхідності, спираючись на отримані у попередньому пункті оцінки, покращуємо модель, наприклад, шляхом додавання нових незалежних змінних, або, навпаки, вилучення непотрібних.

5. Прогнозуємо шукані значення вже оціненою моделлю.

Починаючи з першого кроку, ми задалися питанням пошуку підходящих даних. Для побудови моделі лінійної регресії прогнозування вартості оренди житла у Києві, нам необхідні реальні дані, що якісно та чітко виражали б характеристики квартири саме у Києві. Для цього нами було обрано та виконано парсинг сайт [rieltor.ua](https://www.reltor.ua) (скрипт цього коду буде наведено у додатках нижче).

Для нашого датасету було обрано наступні характеристики:

- ціна (щомісячна оплата оренди квартири);
- район (за адміністративним поділом Києва на райони);



- станція метро (найближча до розглядаємої квартири, що вказана в оголошенні);
- кімнатність квартири;
- поверховість (на якому поверху розташована квартира);
- загальна площа;
- жила площа;
- комфортність проживання (визначення за результатами опитування, що відображає комфортність життя у Києві за районами та проставлена у датасеті у відповідності із районами).

Усі дані було оброблено, очищено та приведено до формального, прийняттого нами вигляду за використання мови програмування python (лістинг коду буде приведено у додатках).

Також, для зручності, усі дані було переведено у числовий формат. Наприклад, районам суто формально було проставлено у відповідність числа від одного до десяти.

Станції метро було «оцифровано», спираючись, на їх нумерацію усередині київського метрополітену (наприклад, на картосхемі київського метро станція «Політехнічний інститут» позначена номером «116», саме так і відображено у нашому датасеті.

Ціну, де та не була вказана у гривні, було переведено у гривню за сучасним курсом ( $1 \text{ USD} = 37.8 \text{ грн.}$  на момент проведення парсингу).

У рамках розгляду поверховості було прийняте рішення не розглядати поверховість будинку, а саме квартири, а також не розглядати нежилу площу, а тільки загальну та, власне, жилу.

Врешті-решт, маємо датасет у такому вигляді (наведемо тільки перші декілька рядків):

|    | Ціна    | Район | Станція метро | Кімнатність | Поверховість | Комфортність | Загальна площа | Жила площа |
|----|---------|-------|---------------|-------------|--------------|--------------|----------------|------------|
| 0  | 22000.0 | 1     | 317.0         | 2           | 4            | 3.04         | 50             | 36         |
| 1  | 22000.0 | 1     | 317.0         | 2           | 3            | 3.04         | 45             | 12         |
| 2  | 22000.0 | 1     | 317.0         | 2           | 4            | 3.04         | 45             | 14         |
| 3  | 20000.0 | 1     | 316.0         | 2           | 10           | 3.04         | 65             | 45         |
| 4  | 20000.0 | 1     | 316.0         | 3           | 10           | 3.04         | 65             | 46         |
| 5  | 25000.0 | 3     | 226.0         | 2           | 18           | 3.06         | 77             | 32         |
| 6  | 55245.0 | 1     | 221.0         | 3           | 24           | 3.04         | 86             | 60         |
| 7  | 31305.5 | 2     | 217.0         | 2           | 3            | 3.01         | 50             | 20         |
| 8  | 16000.0 | 2     | 113.0         | 2           | 6            | 3.01         | 76             | 44         |
| 9  | 31305.5 | 2     | 117.0         | 1           | 3            | 3.01         | 50             | 20         |
| 10 | 15000.0 | 3     | 224.0         | 1           | 15           | 3.06         | 37             | 15         |
| 11 | 15000.0 | 3     | 224.0         | 1           | 15           | 3.06         | 37             | 21         |
| 12 | 16000.0 | 2     | 113.0         | 2           | 6            | 3.01         | 77             | 40         |

Табл. 3.1 – Вигляд зібраного датасету

Також перед побудовою лінійної регресії важливо перевірити корельованість даних, на яких навчатиметься модель, так як модель лінійної регресії ґрунтується на припущенні про наявність лінійної залежності між залежною та незалежними змінними. Якщо залежні та незалежні змінні некорельовані, то модель лінійної регресії не буде вірно описувати залежність між цими змінними. Але якщо змінні мають високу кореляцію, то модель лінійної регресії може якісно спрогнозувати значення залежної змінної на основі значень незалежних змінних. Зобразимо графічно кореляцію між зібраними нами даними:

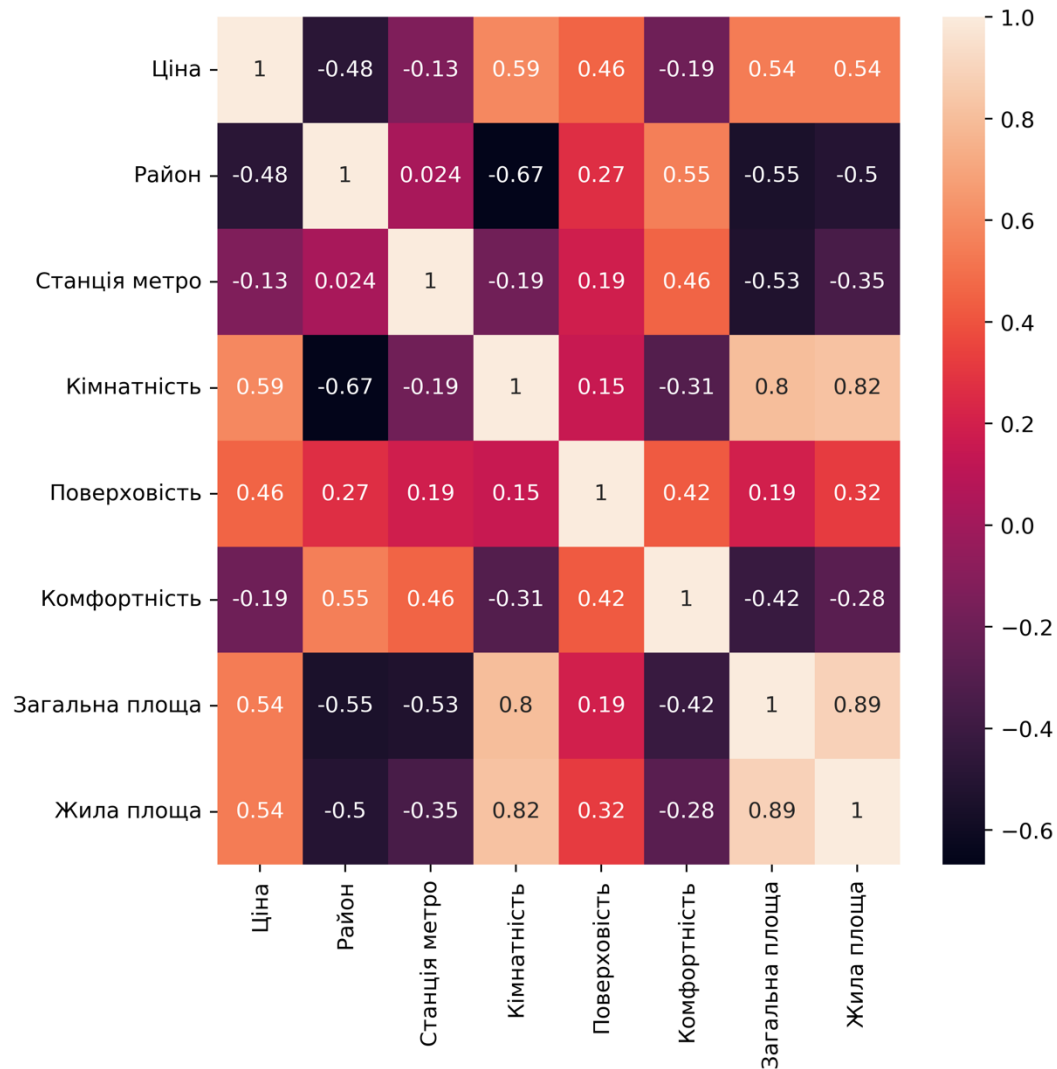


Рис. 3.1 – Візуалізація корельованості даних

Далі, використовуючи, метод найменших квадратів, власне, будемо модель. Щоб не обтяжувати роботу переліченням коду та тавтологічним повторенням теоретичних даних, помістимо лістинг програми у додатках та уникнемо зайвого нагадування вже відомого теоретичного підходу. Як і зазначалося у теорії, для оцінки якості побудованого прогнозу треба оцінити його за допомогою певних метрик.

Нами було обрано наступні метрики: коефіцієнт детермінації, середньоквадратична помилка, корінь із середньоквадратичної помилки та

середня абсолютна помилка. За цими метриками ми отримуємо наступні результати (округлюючи до десятитисячних):

- $R^2 = 0.72394$ ;
- $MAE = 3.12658$ ;
- $MSE = 22.47621$ ;

Маючи такі результати, можемо справедливо назвати модель достатньо точною, якісною та гарно натренованою (і, що важливо, такою, що не зазнала ні «перенавчання» ні є недостатньо складною). Розподіливши попередньо дані на тестову та тренувальну вибірки, застосуємо побудовану моделі на тренувальній вибірці.

Отримуємо наступні результати (знову таки, наведемо тільки перші декілька рядків, із повним датасетом та результатами можна буде ознайомитись у додатках):

|    | Район | Станція метро | Поверховість | Комфортність | Загальна площа | predicted value | true value |
|----|-------|---------------|--------------|--------------|----------------|-----------------|------------|
| 0  | 1     |               | 317          | 3            | 3,04           | 45              | 19         |
| 1  |       | 436,94        | 22 000,00    |              |                |                 |            |
| 2  | 3     |               | 226          | 18           | 3,06           | 77              | 20         |
| 3  |       | 759,29        | 25 000,00    |              |                |                 |            |
| 4  | 2     |               | 117          | 3            | 3,01           | 50              | 22         |
| 5  |       | 160,93        | 31 305,50    |              |                |                 |            |
| 6  | 2     |               | 113          | 6            | 3,01           | 76              | 21         |
| 7  |       | 942,22        | 16 000,00    |              |                |                 |            |
| 8  | 2     |               | 116          | 6            | 3,01           | 73              | 22         |
| 9  |       | 353,35        | 14 000,00    |              |                |                 |            |
| 10 | 1     |               | 316          | 10           | 3,04           | 65              | 26         |
| 11 |       | 138,65        | 20 000,00    |              |                |                 |            |
| 12 | 2     |               | 116          | 6            | 3,01           | 73              | 22         |
| 13 |       | 353,35        | 14 000,00    |              |                |                 |            |
| 14 | 1     |               | 317          | 3            | 3,04           | 45              | 19         |
| 15 |       | 436,94        | 22 000,00    |              |                |                 |            |
|    | 1     |               | 120          | 8            | 3,04           | 75              | 31         |
|    |       | 223,23        | 44 196,00    |              |                |                 |            |
|    | 1     |               | 316          | 10           | 3,04           | 65              | 26         |
|    |       | 138,65        | 20 000,00    |              |                |                 |            |
|    | 3     |               | 224          | 15           | 3,06           | 37              | 23         |
|    | а     | 910,37        | 15 000,00    |              |                |                 |            |
|    | 2     |               | 312          | 6            | 3,01           | 42              | 18         |
|    | б     | 238,09        | 17 000,00    |              |                |                 |            |
|    | 3     |               | 224          | 15           | 3,06           | 37              | 23         |
|    | Л     | 910,37        | 15 000,00    |              |                |                 |            |
|    | 1     |               | 122          | 5            | 3,04           | 65              | 28         |
|    | •     | 539,60        | 16 000,00    |              |                |                 |            |
|    | 5     |               | 321          | 22           | 3,08           | 26              | 11         |
|    |       | 883,57        | 15 000,00    |              |                |                 |            |
|    | 3     | 1             | 317          | 3            | 3,04           | 45              | 19         |
|    |       | 436,94        | 22 000,00    |              |                |                 |            |
|    | •     |               |              |              |                |                 |            |

Метрики на тестовій вибірці будуть наступними:

- $R^2 = 0.75741$ ;
- $MAE = 3.64977$ ;
- $MSE = 34.05682$ ;
- $RMSE = 4.82055$ .

R-квадрат у нашому випадку більше 75 відсотків, це справедливо можна вважати добре натренованою моделлю, помилки ж MAE та MSE допускають не надто сильні відхилення. Цілком, можна з упевненістю назвати модель правильно побудованою, якісно натренованою та такою, що прогнозує вірні результати, які можуть бути використані надалі.

### 3.2. Прогнозування застосуванням методу морфологічного аналізу

У цьому розділі ми будемо проводити двоетапний морфологічний аналіз, спираючись на теоретичні відомості, наведені раніше. Враховуючи, спрогнозовані за допомогою моделі лінійної регресії логічні та лінійно послідовні результати, задаймося питанням, як змінився би прогноз за включення даних та параметрів, що не є лінійними. Проведення морфологічного аналізу дозволить нам за допомогою включення людського фактору та експертних оцінок оцінити вплив на отримані результати таких параметрів, що ми не змогли б оцінити за допомогою моделі лінійної регресії та багатьох інших, побудованих штучно.

Ідея полягає у дослідженні за допомогою морфологічного аналізу питання: чи зміниться ціна на оренду квартири у Києві наступної зими за наявності впливу нелінійних факторів?

Поняття нелінійних факторів уведене, відсилаючись на побудовану раніше модель лінійної регресії. У нашому конкретному випадку ми розглянемо такий узагальнений жакхливий фактор, як війна. Уточнено, з'ясуємо вплив на економіку ринку оренди житла міста Києва у майбутньому (наступної зими, підібравши параметри, спираючись на минулу) під впливом розгорнутої повномасштабної війни на території нашої держави.

Задачею першого етапу морфологічного дослідження є побудова морфологічної таблиці, що розкриває взаємозв'язок обраних нами параметрів та надає можливість комплексного погляду на ситуацію та сценаріїв, що в ній утворюються. При виборі параметрів, як вже зазначалося, спиратися будемо на умови життя у Києві минулої зими. Дослідження проводимо незалежно, ні зі сторони ріелтора, ні того, хто шукає житло, а виключно, у рамках відкритого питання побудови якомога більш точного прогнозу, що враховував би різносторонні фактори впливу та складався не тільки із очевидних лінійно залежних параметрів. Також варто зазначити, що дослідження проводиться неупереджено, із метою оцінити реальний вплив обраних параметрів (але, звісно, із надією на покращення).

Задля проведення дослідження було обрано наступні параметри:

1. Доступність квартири для оренди у Києві:

- низька;
- середня;
- висока.

Інакше кажучи, наскільки перенасичений ринок та чи легко винайняти квартиру.

2. Наявність дефіциту потужності енергосистеми:
  - відсутній;
  - незначний;
  - помірний;
  - сильний;
  - блекаут.
3. Відсутність опалення (відносно частки годин у добі):
  - відсутнє;
  - переважно відсутнє;
  - наявне;
  - повністю наявне.
4. Відносна кількість людей, що виїхали з міста:
  - висока;
  - середня;
  - низька.
5. Відносна кількість людей, що знаходяться у пошуку житла:
  - висока;
  - середня;
  - низька.
6. Наявність бойових дій на території міста Києва:
  - бойові дії є;
  - бойових дій немає.
7. Наявність бойових дій на території України:
  - збільшилися територіально;
  - не змінилися;
  - зменшилися.

Параметр впливає на кількість людей, що вимушені переїхати у пошуках нового житла.

8. Наявність та частота обстрілів у Києві:
  - часто;
  - рідко;
  - обстріли відсутні.



Морфологічна таблиця першого етапу з визначеними характеристичними параметрами наведена в таблиці:

| Характеристичні параметри |                        |                         |                           |
|---------------------------|------------------------|-------------------------|---------------------------|
| Доступність житла         | Дефіцит електроенергії | Відсутність опалення    | Кіл-сть людей, що виїхали |
| 1                         | 2                      | 3                       | 4                         |
| 1.1. висока               | 2.1. відсутній         | 3.1. відсутнє           | 4.1. висока               |
| 1.2. середня              | 2.2. незначний         | 3.2. переважно відсутнє | 4.2. середня              |
| 1.3. низька               | 2.3. помірний          | 3.3. переважно наявне   | 4.3. низька               |
|                           | 2.4. сильний           | 3.4. повністю наявне    |                           |
|                           | 2.5. блекаут           |                         |                           |
|                           |                        |                         |                           |

| Характеристичні параметри     |                               |                                 |                        |
|-------------------------------|-------------------------------|---------------------------------|------------------------|
| Кіл-сть людей у пошуках житла | Наявність бойових дій у Києві | Наявність бойових дій в Україні | Обстріли у Києві       |
| 5                             | 6                             | 7                               | 8                      |
| 5.1. висока                   | 6.1. є                        | 7.1. збільшилися                | 8.1. часто             |
| 5.2. середня                  | 6.2. немає                    | 7.2. не змінилися               | 8.2. рідко             |
| 5.3. низька                   |                               | 7.3. зменшилися                 | 8.3. обстріли відсутні |

Т

аб. 3.1 – Морфологічна таблиця першого етапу

Ймовірності альтернатив визначимо за допомогою експертного оцінювання, оцінюючи кожен у проміжку від «-1» до «1», де «-1» означає, що альтернативи є повністю несумісними і конфігурація з цією парою альтернатив неможлива; «1» - альтернативи є повністю пов'язаними, вибір однієї з них тягне за собою вибір іншої; «0» - повна незалежність альтернатив, вибір однієї з них, не впливає на вибір іншої. Далі наведемо таблицю оцінених сценаріїв морфологічної таблиці:

| Доступність житла |     | Дефіцит електроенергії |      | Відсутність опалення |     | Кіл-сть людей, що виїхали |     |
|-------------------|-----|------------------------|------|----------------------|-----|---------------------------|-----|
| висока            | 0.3 | відсутній              | 0.15 | відсутнє             | 0.2 | висока                    | 0.4 |
| середня           | 0.5 | незначний              | 0.3  | переважно відсутнє   | 0.3 | середня                   | 0.4 |
| низька            | 0.3 | помірний               | 0.4  | переважно наявне     | 0.3 | низька                    | 0.2 |
|                   |     | сильний                | 0.15 | повністю наявне      | 0.2 |                           |     |

| Кіл-сть людей у пошуках житла |     | Наявність бойових дій у Києві |     | Наявність бойових дій в Україні |     | Обстріли у Києві  |      |
|-------------------------------|-----|-------------------------------|-----|---------------------------------|-----|-------------------|------|
| висока                        | 0.5 | є                             | 0.3 | збільшилися                     | 0.2 | часто             | 0.2  |
| середня                       | 0.3 | немає                         | 0.7 | не змінилися                    | 0.5 | рідко             | 0.65 |
| низька                        | 0.2 |                               |     | зменшилися                      | 0.3 | обстріли відсутні | 0.15 |

Таб. 3.2 – Оцінена таблиця альтернатив параметрів першого етапу

Надалі, за алгоритмом методу, маємо оцінити взаємозв'язки між обраними нами параметрами. Отримуємо наступну таблицю взаємозв'язків:

|                               |                   | Доступність житла у Києві |         |        | Дефіцит електроенергії |           |          |         | Відсутність опалення |            |        |        |
|-------------------------------|-------------------|---------------------------|---------|--------|------------------------|-----------|----------|---------|----------------------|------------|--------|--------|
|                               |                   | висока                    | середня | низька | відсутній              | незначний | помірний | сильний | відсутнє             | пер. немає | пер. є | наявне |
| Дефіцит електроенергії        | відсутній         | 0.3                       | 0.2     | 0.3    |                        |           |          |         |                      |            |        |        |
|                               | незначний         | -0.2                      | 0.2     | 0.15   |                        |           |          |         |                      |            |        |        |
|                               | помірний          | 0.1                       | 0.3     | 0.15   |                        |           |          |         |                      |            |        |        |
|                               | сильний           | 0.4                       | 0.15    | 0.3    |                        |           |          |         |                      |            |        |        |
| Відсутність опалення          | відсутнє          | 0.4                       | 0.3     | -0.3   | -0.3                   | 0.3       | 0.8      | 0.9     |                      |            |        |        |
|                               | пер. немає        | 0.4                       | 0.1     | 0.2    | -0.1                   | 0.1       | 0.6      | 0.85    |                      |            |        |        |
|                               | пер. є            | 0.2                       | 0.15    | 0.1    | 0.5                    | -0.2      | 0.1      | 0.3     |                      |            |        |        |
|                               | є                 | 0.5                       | 0.35    | 0.1    | 0.7                    | 0.3       | 0.1      | -0.4    |                      |            |        |        |
| Кіл-сть людей, що виїхали     | висока            | -0.2                      | 0.2     | 0.4    | 0.0                    | 0.0       | 0.0      | 0.0     | 0.0                  | 0.0        | 0.0    | 0.0    |
|                               | середня           | -0.3                      | 0.2     | 0.2    | 0.0                    | 0.0       | 0.0      | 0.0     | 0.0                  | 0.0        | 0.0    | 0.0    |
|                               | низька            | -0.2                      | 0.2     | 0.1    | 0.0                    | 0.0       | 0.0      | 0.0     | 0.0                  | 0.0        | 0.0    | 0.0    |
| Кіл-сть людей у пошуках житла | висока            | -0.8                      | 0.2     | 0.9    | 0.0                    | 0.0       | 0.0      | 0.0     | 0.0                  | 0.0        | 0.0    | 0.0    |
|                               | середня           | 0.2                       | 0.6     | 0.4    | 0.0                    | 0.0       | 0.0      | 0.0     | 0.0                  | 0.0        | 0.0    | 0.0    |
|                               | низька            | 0.85                      | 0.3     | -0.8   | 0.0                    | 0.0       | 0.0      | 0.0     | 0.0                  | 0.0        | 0.0    | 0.0    |
| Бойові дії у Києві            | є                 | -0.2                      | 0.1     | 0.4    | -0.4                   | 0.1       | 0.2      | 0.6     | -0.5                 | 0.1        | 0.25   | 0.6    |
|                               | немає             | 0.3                       | 0.15    | -0.4   | -0.6                   | -0.4      | -0.1     | 0.2     | -0.65                | -0.35      | -0.1   | 0.25   |
| Бойові дії в Україні          | збільшилися       | 0.0                       | 0.0     | 0.0    | -0.3                   | 0.3       | 0.6      | 0.7     | -0.75                | -0.25      | 0.75   | 0.85   |
|                               | не змінилися      | 0.0                       | 0.0     | 0.0    | 0.4                    | 0.2       | 0.1      | 0.2     | 0.15                 | 0.15       | 0.3    | 0.1    |
|                               | зменшилися        | 0.0                       | 0.0     | 0.0    | 0.5                    | 0.1       | -0.4     | -0.2    | 0.7                  | 0.35       | -0.5   | -0.3   |
| Обстріли у Києві              | часто             | 0.3                       | 0.1     | 0.4    | -0.7                   | -0.2      | 0.7      | 0.9     | -0.65                | -0.2       | 0.8    | 0.9    |
|                               | рідко             | 0.05                      | 0.05    | 0.25   | 0.1                    | 0.2       | 0.1      | 0.1     | 0.1                  | 0.3        | 0.15   | 0.1    |
|                               | обстріли відсутні | -0.2                      | 0.15    | 0.1    | 0.7                    | 0.4       | -0.5     | -0.3    | 0.65                 | 0.35       | -0.5   | -0.4   |

| Кіл-сть людей, що виїхали |         |        | Кіл-сть людей у пошуках житла |         |        | Наявність бойових дій у Києві |       | Наявність бойових дій в Україні територіально |              |            |
|---------------------------|---------|--------|-------------------------------|---------|--------|-------------------------------|-------|-----------------------------------------------|--------------|------------|
| висока                    | середня | низька | висока                        | середня | низька | є                             | немає | збільшилися                                   | не змінилися | зменшилися |
|                           |         |        |                               |         |        |                               |       |                                               |              |            |
|                           |         |        |                               |         |        |                               |       |                                               |              |            |
|                           |         |        |                               |         |        |                               |       |                                               |              |            |
| -0.4                      | 0.1     | 0.6    |                               |         |        |                               |       |                                               |              |            |
| 0.2                       | 0.05    | 0.15   |                               |         |        |                               |       |                                               |              |            |
| 0.5                       | 0.1     | -0.1   |                               |         |        |                               |       |                                               |              |            |
| 0.8                       | 0.5     | -0.6   | -0.5                          | 0.2     | 0.5    |                               |       |                                               |              |            |
| -0.4                      | 0.5     | 0.55   | 0.3                           | 0.2     | -0.3   |                               |       |                                               |              |            |
| 0.7                       | 0.4     | -0.3   | 0.6                           | 0.2     | -0.4   | 0.7                           | -0.4  |                                               |              |            |
| 0.3                       | 0.6     | 0.3    | 0.2                           | 0.15    | -0.1   | -0.2                          | 0.4   |                                               |              |            |
| -0.6                      | 0.4     | 0.7    | 0.1                           | 0.2     | 0.3    | -0.85                         | 0.9   |                                               |              |            |
| 0.9                       | 0.65    | -0.6   | -0.6                          | 0.25    | 0.5    | 1                             | 0.15  | 0.6                                           | 0.1          | 0.05       |
| 0.25                      | 0.1     | 0.6    | 0.4                           | 0.1     | -0.3   | -0.95                         | 0.5   | -0.4                                          | 0.6          | 0.5        |
| -0.8                      | -0.1    | 0.75   | 0.7                           | 0.15    | -0.5   | -1                            | 0.8   | -0.7                                          | 0.5          | 0.85       |

Табл. 3.3 – Оцінена таблиця взаємозв'язків між альтернативами параметрів

Задачею другого етапу морфологічного дослідження є відповідь на питання про те, як зміниться вартість оренди квартири у Києві під впливом факторів, розглянутих на першому етапі.

Задля оцінки такого впливу було обрано очевидну групу параметрів:

- сильно впаде;
- помірно впаде;
- не зміниться;
- помірно зросте;
- сильно зросте.

Морфологічна таблиця другого етапу:

| Ціна на оренду житла |
|----------------------|
| 9                    |
| 9.1. сильно впаде    |
| 9.2. помірно впаде   |
| 9.3. не зміниться    |
| 9.4. помірно зросте  |
| 9.5. сильно зросте   |

Табл. 3.4 – Морфологічна таблиця другого етапу

Проміжні результати представимо у вигляді таблиці конфігурацій, у якій будуть наведені усі можливі конфігурації. Вона будується наступним чином:

1. На першому кроці розраховується добуток ймовірностей незалежних ймовірностей відповідних альтернатив кожної конфігурації.
2. На другому – добуток зміщених на одиницю значень матриці взаємозв'язків, які відповідають усім можливим парам альтернатив цієї конфігурації.
3. Далі розраховуємо ненормоване значення ймовірності конфігурації, як добуток значень перших двох кроків.
4. Остаточню, знаходимо суму таких ненормованих значень і розраховуємо нормовані значення ймовірностей конфігурацій шляхом ділення кожного ненормованого значення на їх суму.

У рамках нашої роботи ми маємо достатньо велику кількість альтернатив та параметрів, тож числові розрахунки проведемо автоматично, за допомогою скрипту на мові програмування Python. Лістинг якого буде наведено у додатках.

Отримуємо наступну таблицю конфігурацій із розрахованими оцінками (наведемо лише декілька її рядків):

|       | F1  | F2  | F3  | F4  | F5  | F6  | F7  | F8  | p        | C        | p*C          | P(s)         |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|----------|----------|--------------|--------------|
| 0     | a1  | a1  | a1  | a1  | a1  | a1  | a1  | a1  | 0.000022 | 0.007097 | 1.532852e-07 | 5.513324e-09 |
| 1     | a1  | a1  | a1  | a1  | a1  | a1  | a1  | a2  | 0.000070 | 0.001426 | 1.000971e-07 | 3.600266e-09 |
| 3     | a1  | a1  | a1  | a1  | a1  | a1  | a2  | a1  | 0.000054 | 0.012114 | 6.541804e-07 | 2.352940e-08 |
| 4     | a1  | a1  | a1  | a1  | a1  | a1  | a2  | a2  | 0.000176 | 0.009441 | 1.656970e-06 | 5.959749e-08 |
| 6     | a1  | a1  | a1  | a1  | a1  | a1  | a3  | a1  | 0.000032 | 0.000969 | 3.138261e-08 | 1.128762e-09 |
| ...   | ... | ... | ... | ... | ... | ... | ... | ... | ...      | ...      | ...          | ...          |
| 7772  | a3  | a4  | a4  | a3  | a3  | a2  | a2  | a3  | 0.000019 | 0.391495 | 7.399258e-06 | 2.661347e-07 |
| 7773  | a3  | a4  | a4  | a3  | a3  | a2  | a3  | a1  | 0.000015 | 1.428342 | 2.159653e-05 | 7.767787e-07 |
| 7774  | a3  | a4  | a4  | a3  | a3  | a2  | a3  | a2  | 0.000049 | 1.486807 | 7.306169e-05 | 2.627865e-06 |
| 7775  | a3  | a4  | a4  | a3  | a3  | a2  | a3  | a3  | 0.000011 | 0.525113 | 5.954783e-06 | 2.141801e-07 |
| Total | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN      | NaN      | 2.780268e+01 | 1.000000e+00 |

Рис. 3.2 – Таблиця конфігурацій першого етапу

Як бачимо, маємо 7775 конфігурацій о автоматично розраховані оцінки для кожної з них.

Далі, враховуючи зв'язки та знайдені нами оцінки конфігурацій, знаходимо оцінки окремих альтернатив параметрів як суми ймовірностей всіх конфігурацій, що містять цю альтернативу:

|    | F1    | F2    | F3    | F4    | F5    | F6    | F7    | F8    |
|----|-------|-------|-------|-------|-------|-------|-------|-------|
| a1 | 0.137 | 0.075 | 0.085 | 0.288 | 0.422 | 0.312 | 0.269 | 0.413 |
| a2 | 0.590 | 0.194 | 0.271 | 0.455 | 0.443 | 0.688 | 0.541 | 0.505 |
| a3 | 0.273 | 0.465 | 0.335 | 0.257 | 0.135 | NaN   | 0.189 | 0.082 |
| a4 | NaN   | 0.266 | 0.309 | NaN   | NaN   | NaN   | NaN   | NaN   |
| a5 | NaN   | NaN   | NaN   | NaN   | NaN   | NaN   | NaN   | NaN   |
| a6 | NaN   | NaN   | NaN   | NaN   | NaN   | NaN   | NaN   | NaN   |

Рис. 3.3 – Переоцінена таблиця альтернатив параметрів

Переходячи до другого етапу, зазначимо таблицю оцінок матриці взаємозв'язків першого та другого етапів, яку використовуватимемо надалі:

|                               |                   | Ціна на оренду житла у Києві |               |              |                |               |
|-------------------------------|-------------------|------------------------------|---------------|--------------|----------------|---------------|
|                               |                   | сильно впаде                 | помірно впаде | не зміниться | помірно зросте | сильно зросте |
| Доступність житла у Києві     | висока            | 0.4                          | 0.2           | -0.1         | -0.4           | -0.8          |
|                               | середня           | 0.1                          | 0.3           | 0.9          | 0.3            | 0.1           |
|                               | низька            | -0.8                         | -0.3          | 0.2          | 0.45           | 0.85          |
| Дефіцит електроенергії        | відсутній         | -0.85                        | -0.5          | -0.2         | 0.5            | 0.7           |
|                               | незначний         | -0.7                         | -0.3          | 0.2          | 0.25           | 0.4           |
|                               | помірний          | 0.75                         | 0.6           | -0.3         | -0.3           | 0.4           |
|                               | сильний           | 0.9                          | 0.75          | -0.2         | -0.45          | -0.65         |
| Відсутність опалення          | відсутнє          | 0.9                          | 0.75          | -0.2         | -0.45          | -0.65         |
|                               | переважно немає   | 0.75                         | 0.6           | -0.3         | -0.3           | 0.4           |
|                               | переважно є       | -0.7                         | -0.3          | 0.2          | 0.25           | 0.4           |
|                               | є                 | -0.85                        | -0.5          | -0.2         | 0.5            | 0.7           |
| Кіл-сть людей, що виїхали     | висока            | 0.0                          | 0.0           | 0.0          | 0.0            | 0.0           |
|                               | середня           | 0.0                          | 0.0           | 0.0          | 0.0            | 0.0           |
|                               | низька            | 0.0                          | 0.0           | 0.0          | 0.0            | 0.0           |
| Кіл-сть людей у пошуках житла | висока            | -0.7                         | -0.55         | -0.4         | 0.75           | 0.9           |
|                               | середня           | 0.15                         | 0.3           | 0.6          | 0.3            | 0.15          |
|                               | низька            | 0.8                          | 0.6           | -0.1         | -0.4           | -0.6          |
| Бойові дії у Києві            | є                 | 0.9                          | 0.7           | -0.25        | -0.65          | -0.9          |
|                               | немає             | -0.9                         | -0.65         | -0.25        | 0.7            | 0.9           |
| Бойові дії в Україні          | збільшилися       | 0.0                          | 0.0           | 0.0          | 0.0            | 0.0           |
|                               | не змінилися      | 0.0                          | 0.0           | 0.0          | 0.0            | 0.0           |
|                               | зменшилися        | 0.0                          | 0.0           | 0.0          | 0.0            | 0.0           |
| Обстріли у Києві              | часто             | 0.8                          | 0.6           | -0.2         | -0.6           | -0.85         |
|                               | рідко             | -0.2                         | 0.15          | 0.3          | 0.15           | -0.2          |
|                               | обстріли відсутні | -0.4                         | -0.2          | -0.4         | 0.7            | 0.75          |

Табл. 3.5 – Таблиця оцінки взаємозв'язків першого та другого етапів

Проміжні значення другого етапу також шукатимемо за допомогою таблиці конфігурацій, що будується за тим самим принципом, що і на першому

|       | F1  | F2  | F3  | F4  | F5  | F6  | F7  | F8  | P(s)         | r1_8     | ... | R1_8     | R2_8     | R3_8     | R4_8     | R5_8     | R1_8*P(s)    | R2_8*P(s)    | R3_8*P(s)    | R4_8*P(s)    | R5_8*P(s)    |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|--------------|----------|-----|----------|----------|----------|----------|----------|--------------|--------------|--------------|--------------|--------------|
| 0     | a1  | a1  | a1  | a1  | a1  | a1  | a1  | a1  | 5.513324e-09 | 0.409374 | ... | 0.202000 | 0.634165 | 0.102319 | 0.059842 | 0.001673 | 1.113693e-09 | 3.496360e-09 | 5.641185e-10 | 3.299261e-10 | 9.226505e-12 |
| 1     | a1  | a1  | a1  | a1  | a1  | a1  | a1  | a2  | 3.600266e-09 | 0.181944 | ... | 0.100555 | 0.510523 | 0.186228 | 0.192697 | 0.009997 | 3.620253e-10 | 1.838018e-09 | 6.704704e-10 | 6.937618e-10 | 3.599082e-11 |
| 3     | a1  | a1  | a1  | a1  | a1  | a1  | a1  | a2  | 2.352940e-08 | 0.409374 | ... | 0.202000 | 0.634165 | 0.102319 | 0.059842 | 0.001673 | 4.752947e-09 | 1.492153e-08 | 2.407508e-09 | 1.408037e-09 | 3.937626e-11 |
| 4     | a1  | a1  | a1  | a1  | a1  | a1  | a2  | a2  | 5.959749e-08 | 0.181944 | ... | 0.100555 | 0.510523 | 0.186228 | 0.192697 | 0.009997 | 5.992835e-09 | 3.042588e-08 | 1.109872e-08 | 1.148428e-08 | 5.957789e-10 |
| 6     | a1  | a1  | a1  | a1  | a1  | a1  | a3  | a1  | 1.128762e-09 | 0.409374 | ... | 0.202000 | 0.634165 | 0.102319 | 0.059842 | 0.001673 | 2.280103e-10 | 7.158218e-10 | 1.154939e-10 | 6.754691e-11 | 1.888974e-12 |
| ...   | ... | ... | ... | ... | ... | ... | ... | ... | ...          | ...      | ... | ...      | ...      | ...      | ...      | ...      | ...          | ...          | ...          | ...          | ...          |
| 7772  | a3  | a4  | a4  | a3  | a3  | a2  | a2  | a3  | 2.661347e-07 | 0.006156 | ... | 0.001491 | 0.066442 | 0.075314 | 0.502264 | 0.354488 | 3.966993e-10 | 1.768263e-08 | 2.004375e-08 | 1.336700e-07 | 9.434159e-08 |
| 7773  | a3  | a4  | a4  | a3  | a3  | a2  | a3  | a1  | 7.767787e-07 | 0.018468 | ... | 0.011575 | 0.343958 | 0.259924 | 0.305896 | 0.078647 | 8.991011e-09 | 2.671793e-07 | 2.019034e-07 | 2.376133e-07 | 6.109170e-08 |
| 7774  | a3  | a4  | a4  | a3  | a3  | a2  | a3  | a2  | 2.627865e-06 | 0.008208 | ... | 0.002607 | 0.125261 | 0.214008 | 0.445597 | 0.212527 | 6.849562e-09 | 3.291679e-07 | 5.623851e-07 | 1.170969e-06 | 5.584930e-07 |
| 7775  | a3  | a4  | a4  | a3  | a3  | a2  | a3  | a3  | 2.141801e-07 | 0.006156 | ... | 0.001491 | 0.066442 | 0.075314 | 0.502264 | 0.354488 | 3.192561e-10 | 1.423065e-08 | 1.613084e-08 | 1.075751e-07 | 7.592433e-08 |
| Total | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN          | NaN      | ... | NaN      | NaN      | NaN      | NaN      | NaN      | 7.971076e-02 | 2.540131e-01 | 1.501874e-01 | 2.485415e-01 | 2.675472e-01 |

етапі.

Рис. 3.4 – Таблиця конфігурацій другого етапу

Результат, поданий у вигляді наступної таблиці, буде означати очікувану результативність кожної із альтернатив.

| Ціна на оренду житла у Києві |       |
|------------------------------|-------|
| сильно впаде                 | 0.080 |
| помірно впаде                | 0.254 |
| не зміниться                 | 0.150 |
| помірно зросте               | 0.249 |
| сильно зросте                | 0.268 |

Табл. 3.6 – Таблиця оцінки результатів

### 3.3. Висновки на основі отриманих результатів

У цій частині роботи проведемо аналіз отриманих результатів за проведеним у розділі прогнозування двома методами – кількісним та якісним.

За побудованим за допомогою кількісного методу прогнозу можна побачити певні результати, які, на нашу думку, краще відобразити графічно. Наприклад. Наведемо графік розподіл цін на оренду квартири у Києві у Печерському районі, де червона лінія відображає розподіл реальних цін на ринку оренди нерухомого майна у Києві, а синя – спрогнозованих.

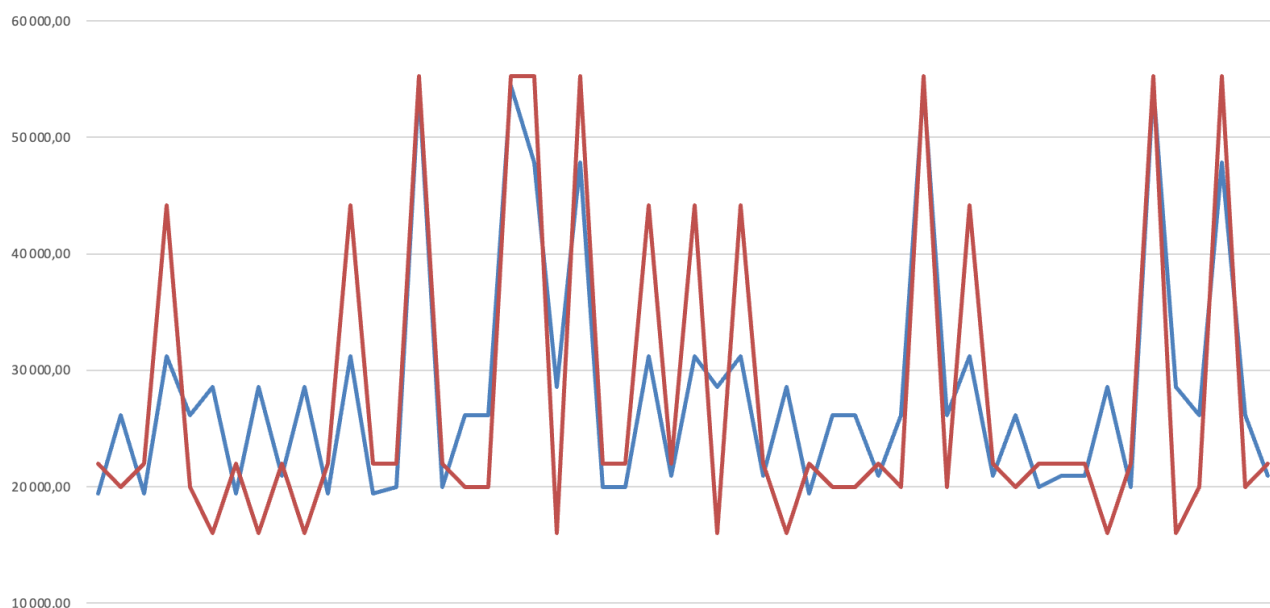


Рис. 3.5 – Графік реальної та прогнозованої цін у Печерському районі

Розглянемо також для порівняння ситуацію у більш віддаленому від центру району та менш облаштованому інфраструктурно, наприклад, у Дніпровському.



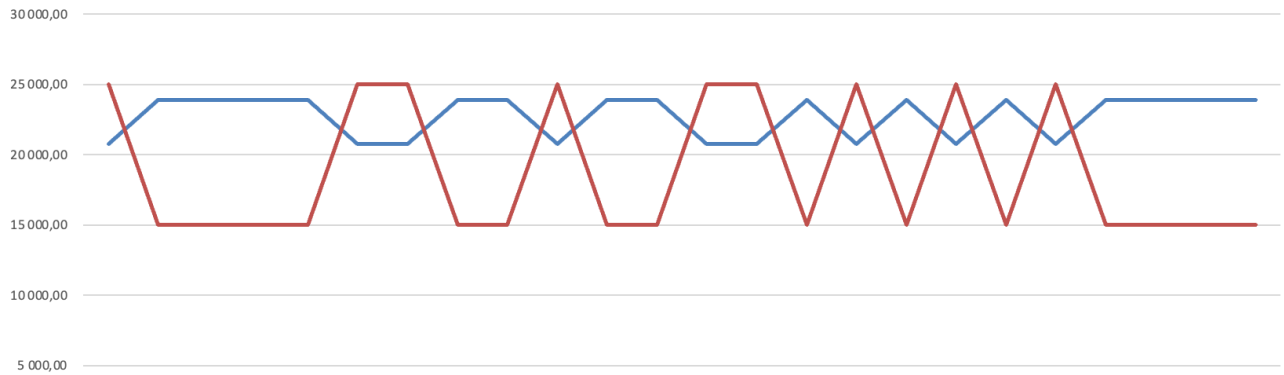


Рис. 3.6 – Графік реальної та прогнозованої цін у Дніпровському районі

Як бачимо, модель прогнозує як можливий приріст, так і спадання ціни, що логічно, так як ми відсортували дані лише по району, не враховуючи інших параметрів. Якщо зв'язок між спрогнозованими даними і заданими параметрами зрозумілий, то задаймося питанням, відповідь на яке надають нам результати проведеного морфологічного аналізу: як війна вплине на змінення ціни на оренду житла у Києві?

Неабиякою перевагою морфологічного аналізу є відмінність його параметрів від параметрів моделі лінійної регресії, що дозволяє досягнути оцінки за іншими аспектами, не пов'язаними із лінійною регресією, а значить такими, що не можуть повипливати одне на одного. Результати ММА було наведено у роботі вище.

Для відображення спільної картини та отримання загального суцільного прогнозу, сумістимо отримані результати. Для цього скоригуємо отримані побудованою моделлю лінійної регресії результати за допомогою оцінок, що надає нам проведений морфологічний аналіз.

| Ціна на оренду житла у Києві |       |      |
|------------------------------|-------|------|
| сильно впаде                 | 0,08  | 10%  |
| помірно впаде                | 0,254 | 125% |
| не зміниться                 | 0,15  | 55%  |
| помірно зросте               | 0,249 | 130% |
| сильно зросте                | 0,268 | 90%  |

Табл. 3.6 – Таблиця оцінки результатів

Початково, оцінимо у відсотковому співвідношенні вірогідний вплив кожної альтернативи результуючого параметру морфологічного аналізу на отримані результати. Отримаємо наступну таблицю:

Надалі, переоцінимо отримані за допомогою моделі лінійної регресії результати методом зважування, проводячи його наступним чином:

$$P_i = P_{\#} \prod_{\#} a_{\#}^{b_{\#}},$$

де:

- $P_i$  – шукана результуюча ціна;
- $P_{\#}$  – ціна, отримана як результат прогнозу методу лінійної регресії;
- $a_{\#}$  – отримана у результаті морфологічного аналізу оцінка кожної альтернативи результуючого параметру;
- $b_{\#}$  – оцінений нами вище відсотковий вплив кожної альтернативи.

За допомогою Excel проведемо ці нескладні розрахунки. Для наочності, наведемо лише перші декілька рядків отриманого результату, із повним обсягом можна буде ознайомитися у додатках.

|    | Район | Станція метро | Поверховість | Комфортність | Загальна площ | predicted value | true valu | result    |
|----|-------|---------------|--------------|--------------|---------------|-----------------|-----------|-----------|
| 0  | 1     | 317           | 3            | 3,04         | 45            | 19 436,94       | 22 000,00 | 18 910,20 |
| 1  | 3     | 226           | 18           | 3,06         | 77            | 20 759,29       | 25 000,00 | 20 196,71 |
| 2  | 2     | 117           | 3            | 3,01         | 50            | 22 160,93       | 31 305,50 | 21 560,36 |
| 3  | 2     | 113           | 6            | 3,01         | 76            | 21 942,22       | 16 000,00 | 21 347,58 |
| 4  | 2     | 116           | 6            | 3,01         | 73            | 22 353,35       | 14 000,00 | 21 747,57 |
| 5  | 1     | 316           | 10           | 3,04         | 65            | 26 138,65       | 20 000,00 | 25 430,29 |
| 6  | 2     | 116           | 6            | 3,01         | 73            | 22 353,35       | 14 000,00 | 21 747,57 |
| 7  | 1     | 317           | 3            | 3,04         | 45            | 19 436,94       | 22 000,00 | 18 910,20 |
| 8  | 1     | 120           | 8            | 3,04         | 75            | 31 223,23       | 44 196,00 | 30 377,08 |
| 9  | 1     | 316           | 10           | 3,04         | 65            | 26 138,65       | 20 000,00 | 25 430,29 |
| 10 | 3     | 224           | 15           | 3,06         | 37            | 23 910,37       | 15 000,00 | 23 262,40 |
| 11 | 2     | 312           | 6            | 3,01         | 42            | 18 238,09       | 17 000,00 | 17 743,84 |
| 12 | 3     | 224           | 15           | 3,06         | 37            | 23 910,37       | 15 000,00 | 23 262,40 |
| 13 | 1     | 122           | 5            | 3,04         | 65            | 28 539,60       | 16 000,00 | 27 766,18 |
| 14 | 5     | 321           | 22           | 3,08         | 26            | 11 883,57       | 15 000,00 | 11 561,53 |
| 15 | 1     | 317           | 3            | 3,04         | 45            | 19 436,94       | 22 000,00 | 18 910,20 |
| 16 | 3     | 224           | 15           | 3,06         | 37            | 23 910,37       | 15 000,00 | 23 262,40 |
| 17 | 2     | 117           | 3            | 3,01         | 50            | 22 160,93       | 31 305,50 | 21 560,36 |
| 18 | 2     | 312           | 6            | 3,01         | 42            | 18 238,09       | 17 000,00 | 17 743,84 |

Рис. 3.7 – Отриманий датасет із скоригованою оцінкою

Графічне представлення отриманого скоригованого морфологічним аналізом прогнозу та дійсних значень можна побачити на наступному графіку, де зеленою ламаною позначено спрогнозовані дані, а синьою – дійсні.

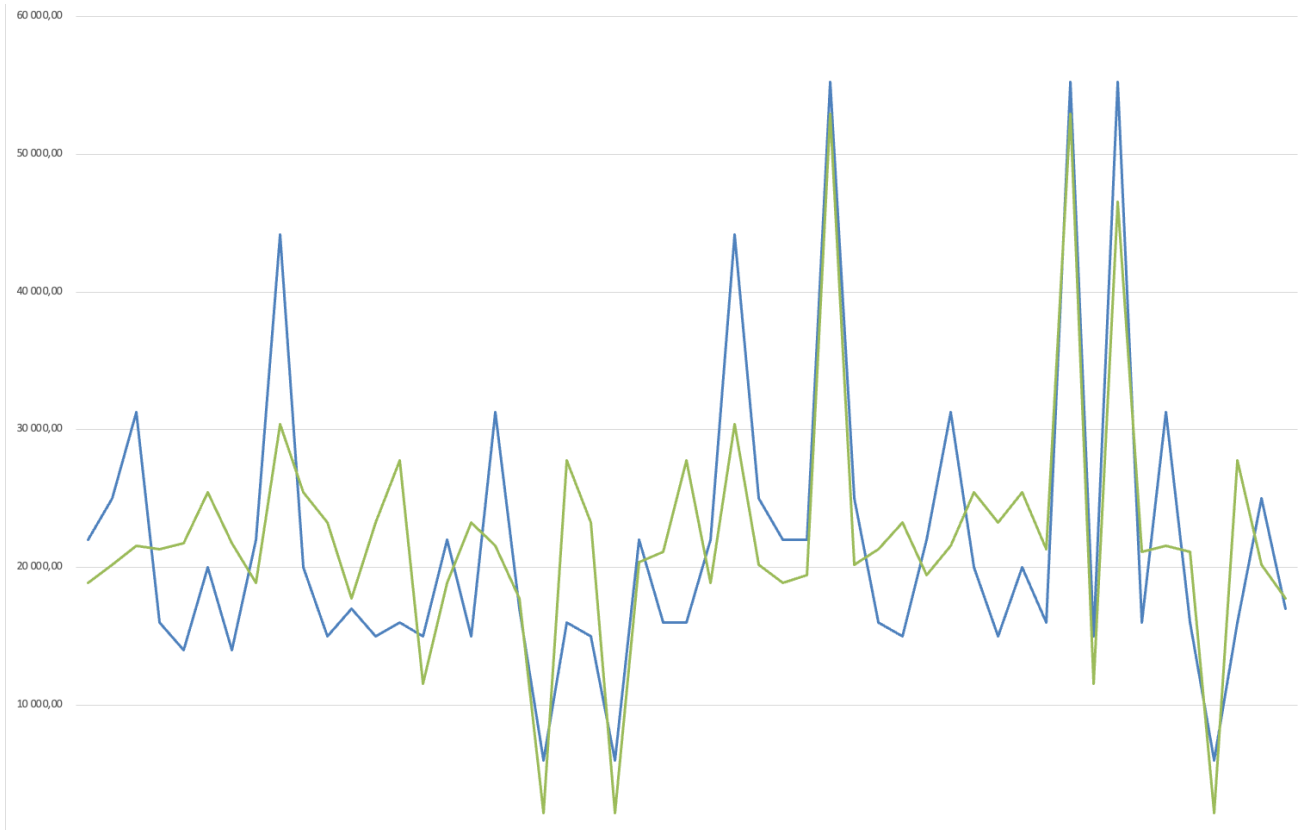


Рис. 3.8 – Графік реальної та прогнозованої цін після коригування

Із проведеного аналізу та побудованого прогнозу можна зробити висновок про рівнозначне як помірне спадання, та і помірне зростання ціни. Можна встановити, що ціна малоімовірно надто сильно зросте чи надто сильно сягатиме низьких порогів. У той же час, не можна сказати, що ціна на оренду житла у Києві не зазнає значущих змін. Збільшення же чи зменшення ціни можна пояснити як економічними факторами (наприклад, такими як розташування квартири у «дорогому» районі Києві), що впливають на побудовану модель лінійної регресії, так і параметрами, вплив яких не є очевидним, та які було розглянути при побудові прогнозу якісним методом – методом морфологічного аналізу.

Таким чином, у залежності від заданих характеристик конкретної квартири (таких як, район розташування, площа тощо) можна рівносильно передбачити помірне зростання або помірне спадання ціни, враховуючи фактори впливу, що завдає війна на ринок оренди нерухомості у Києві.

### 3.4. Висновки до розділу 3

У цьому розділі застосовано методи прогнозування задля досягнення результатів поставленої задачі. Побудовано модель лінійної регресії, прогнозування якої визначене як точне та якісне відповідними метриками. Проведено підготовчий процес відбору параметрів та проведено двоетапний морфологічний аналіз задля отримання якісної характеристики у рамках поставленої задачі.

Отримані двома методами результати проаналізовано, наведено графічну візуалізацію та скомпановано у єдине ціле задля досягнення мети дослідження -

Побудови якомога більше точного та якісного прогнозу вартості оренди житла на основі реальних даних, чітких характеристик та з урахуванням якісних параметрів та факторів впливу й чинників, таких, як війна.

## 4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ

У даному розділі буде проведено оцінювання основних характеристик для майбутнього програмного продукту, що спеціалізується на дослідженні демографічного стану.

Дана реалізація буде сприяти проведенню усіх необхідних досліджень, що дасть змогу якісно дослідити питання не лише в Україні, проте у всьому світі.

Також в даному дослідженні показано різні варіанти реалізації для забезпечення найбільш коректної та оптимальної стратегії вибору, що має вплив на економічні фактори та сумісність з майбутнім програмним продуктом. Для цього застосовувався апарат функціонально-вартісного аналізу.

Функціонально-вартісний аналіз (ФВА) передбачає собою технологію, що 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.

## 4.1 Постановка задачі проектування

У роботі застосовується метод ФВА для проведення техніко-економічного аналізу розробки системи прогнозу стійкості фінансових показників. 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 компанії.

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  $F$  – розробка можливого програмного продукту, який дозволяє спрогнозувати вартість оренди житла у Києві кількісними та якісними методами, базуючись на наборі даних, що складено самостійно. Беручи за основу цю функцію, можна виділити наступні:

- $F_{\#}$  – вибір кількісного методу аналізу та прогнозування;
- $F_{\$}$  – вибір якісного методу аналізу та прогнозування;
- $F_{\%}$  – вибір мови програмування та середовища розробки.

Кожна з цих функцій має декілька варіантів реалізації (наведемо лиш декілька для кожної).

Функція  $F_{\#}$ :

1. Лінійна регресія.

2. Нейромережева модель.

Функція  $F_{\$}$ :

1. Метод експертних оцінок.
2. Метод морфологічного аналізу.

Функція  $F_{\%}$ :

1. Python, PyCharm.
2. Python, Jupyter Notebook.
3. R, Visual Studio Code.
4. R, Google Colaboratory.

Варіанти реалізації основних функцій наведені у морфологічній карті системи на наступному рисунку:

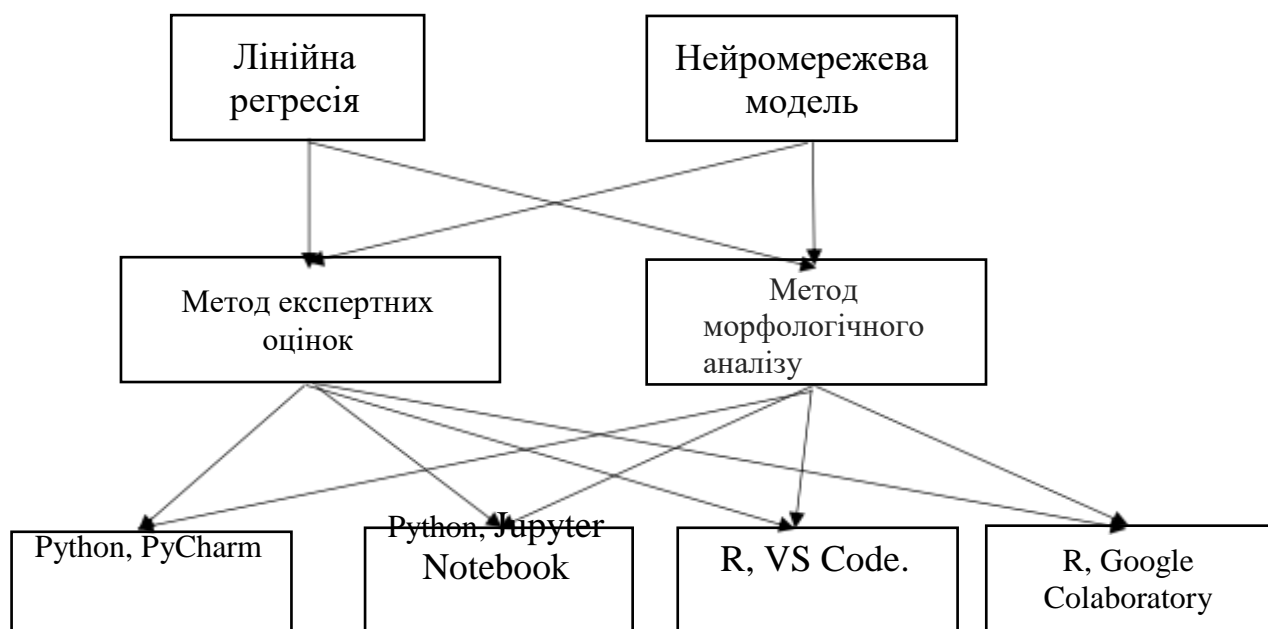


Рисунок 4.1 – Морфологічна карта



Морфологічна карта відображає множину всіх можливих варіантів основних функцій. Позитивно-негативна матриця наведена у таблиці 4.1.

| Функції  | Варіанти реалізації | Переваги                                                                                              | Недоліки                                                                                |
|----------|---------------------|-------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|
| $F_{\#}$ | 1                   | Проста реалізація, швидкість виконання, потужність                                                    | Можливі викиди при неоднорідності обраного набору даних                                 |
|          | 2                   | Більш велика точність, можливість додання більшої складності                                          | Більш складна у реалізації, потребує більших умінь та навичок                           |
| $F_{\$}$ | 1                   | Відображає реальну картину та настрої суспільства                                                     | Залежить від великої кількості зовнішніх факторів впливу                                |
|          | 2                   | Дозволяє провести більш ґрунтовний аналіз із математичним підходом                                    | Для здобуття необхідних оцінок все                                                      |
|          |                     |                                                                                                       | одно потребує проведення експертного оцінювання                                         |
| $F_{\%}$ | 1                   | Зручна мова програмування із багатьма вбудованими бібліотеками та середовище розробки, зручне для неї | Python може потребувати певних навичок у роботі, а PyCharm використовує багато пам'яті, |
|          | 2                   | Зручний у використанні, особливо при написанні коду на Python                                         | Більші потужності не є безкоштовними                                                    |

|  |   |                                                                                                                                                 |                                                                                                      |
|--|---|-------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|
|  |   | Зручна мова                                                                                                                                     |                                                                                                      |
|  | 3 | програмування для роботи зі статистичними даними, можна суміщати із іншими мовами, тим паче VS Code також підтримує багато мов                  | R має достатньо посередню швидкість, є складним, а із VS Code можуть виникати проблеми із дебагінгом |
|  | 4 | Ті самі переваги мови R, при цьому Google Colaboratory може забезпечувати високу потужність та є дуже зручним і при цьому повністю безкоштовним | Google Colaboratory важко та довго працює із великими наборами даних                                 |
|  |   |                                                                                                                                                 |                                                                                                      |
|  |   |                                                                                                                                                 |                                                                                                      |

Таблиця 4.1 – Позитивно-негативна матриця

На основі аналізу позитивно-негативної матриці 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  $F_{\#}$ :

Надаємо перевагу потужності та швидкості без задіяння великих не необхідних складностей, тому обираємо перший варіант.

Функція  $F_{\$}$ :

Надаємо перевагу стабільності, науковості та математичному підходу, тому варіант 1 має бути відкинтий.

Функція  $F_{\%}$ :

Реалізація програми на мові Python є найбільш сприйнятливою, має кращий розподіл переваг та недоліків серед інших. Тому відкинемо варіанти 3 і 4.

Таким чином, будемо розглядати такий варіанти реалізації ПП:

$$F_1 1 - F_2 2 - F_3 1$$

$$F_1 1 - F_2 2 - F_3 2$$

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

#### 4.3 Обґрунтування системи параметрів програмного продукту

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:

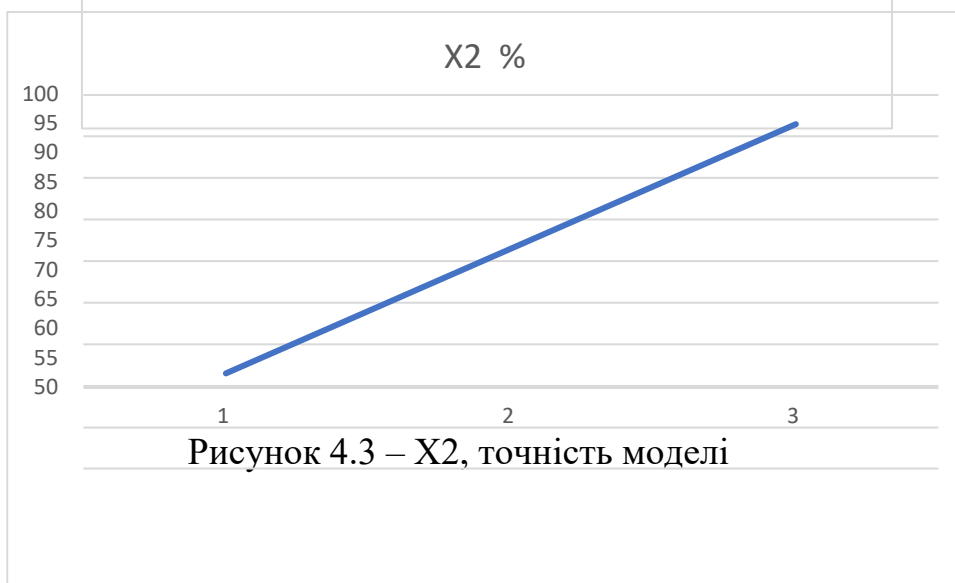
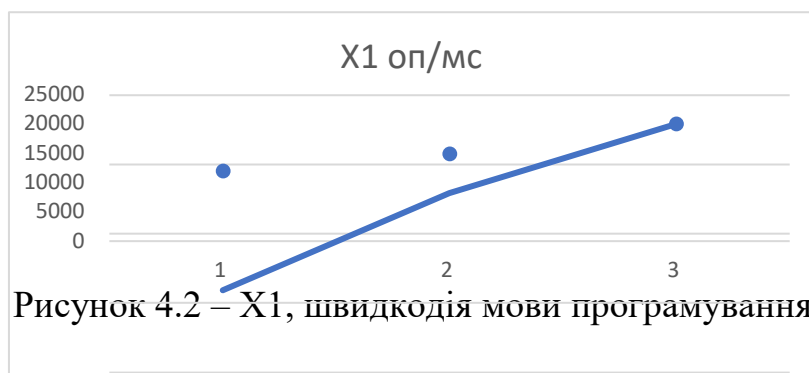
- $X1$  – швидкодія мови програмування;
- $X2$  – об'єм пам'яті для обчислень та збереження даних;
- $X3$  – час навчання даних;
- $X4$  – час навчання моделі.

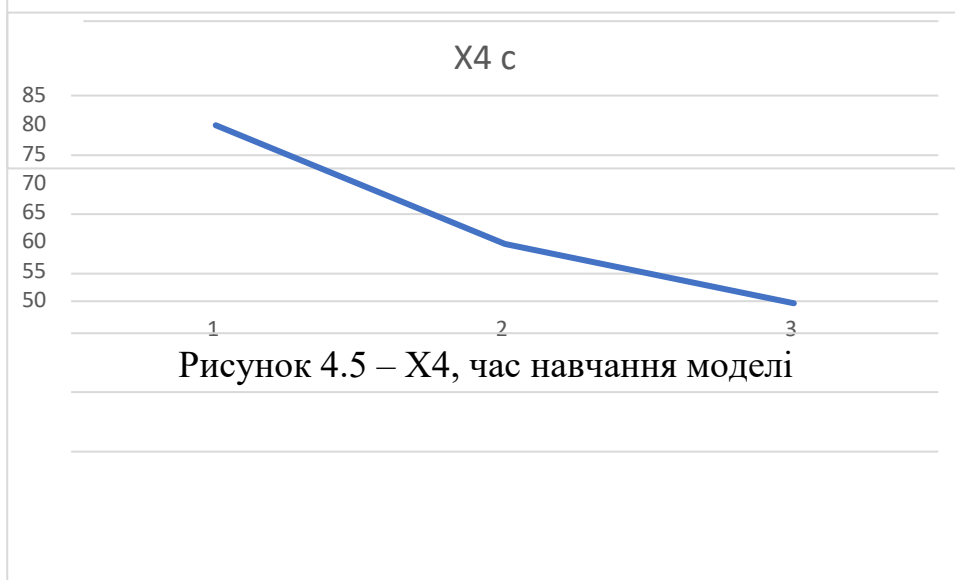
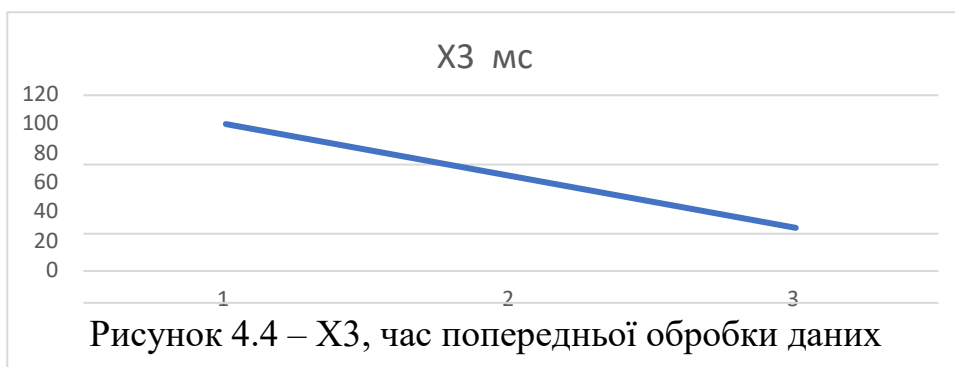
Гірші, середні і кращі значення параметрів вибираються на основі вимог замовника й умов, що характеризують експлуатацію програмного продукту, як показано у таблиці 4.2.

| Назва<br>Параметра               | Умовні<br>позначення | Одиниці<br>виміру | Значення параметра |         |       |
|----------------------------------|----------------------|-------------------|--------------------|---------|-------|
|                                  |                      |                   | гірші              | середні | кращі |
| Швидкодія мови<br>програмування  | X1                   | оп/мс             | 8000               | 15000   | 20000 |
| Точність моделі                  | X2                   | %                 | 65                 | 80      | 95    |
| Час попередньої обробки<br>даних | X3                   | мс                | 100                | 85      | 70    |
| Час навчання моделі              | X4                   | с                 | 80                 | 70      | 65    |

Таблиця 4.2 – Основні параметри програмного продукту

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





#### 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.

| 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                                  | 3 | 4 | 3 | 4 | 4 | 4 | 26                | 8,5                   | 72,25        |
| X2                   | Точність моделі              | %              | 2                                  | 1 | 2 | 2 | 1 | 1 | 2 | 11                | -6,5                  | 42,25        |

|    |                               |    |    |    |    |    |    |    |    |    |      |       |
|----|-------------------------------|----|----|----|----|----|----|----|----|----|------|-------|
| X3 | Час попередньої обробки даних | мс | 3  | 4  | 3  | 4  | 3  | 3  | 3  | 23 | 5,5  | 30,25 |
| X4 | Час навчання моделі           | с  | 1  | 2  | 1  | 1  | 2  | 2  | 1  | 10 | -7,5 | 56,25 |
|    | Разом                         |    | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 70 | 0    | 201   |

Таблиця 4.3 – Результати ранжування параметрів

Для перевірки степені достовірності експертних оцінок, визначимо наступні параметри:

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

$$R_{\&} = \sum_{i=1}^N r_{\& i} = \frac{N(n+1)}{2} = 70, \quad (4.1)$$

де  $N$  – число експертів та  $n$  – кількість параметрів;

б) середня сума рангів:

$$T = \frac{1}{n} R_{\&} = 17,5. \quad (4.2)$$

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

$$\Delta_{\&} = R_{\&} - T. \quad (4.3)$$

Сума відхилень по всіх параметрах повинна дорівнювати 0;



г) загальна сума квадратів відхилення:

$$S = N \Delta_{\&}^2 = 201. \quad (4.4)$$

Порахуємо коефіцієнт узгодженості:

$$W = \frac{12S}{N(n_{\&} - n)} = \frac{12 \cdot 201}{7(4_{\&} - 4)} = 0,820 > W_{\&} = 0,67. \quad (4.5)$$

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.

| Параметри | Експерти |   |   |   |   |   |   | Кінцева оцінка | Числове значення |
|-----------|----------|---|---|---|---|---|---|----------------|------------------|
|           | 1        | 2 | 3 | 4 | 5 | 6 | 7 |                |                  |
| X1 і X2   | ≥        | ≥ | ≥ | ≥ | ≥ | ≥ | ≥ | ≥              | 1,5              |
| X1 і X3   | ≥        | ≤ | ≥ | ≤ | ≥ | ≥ | ≥ | ≥              | 1,5              |
| X1 і X4   | ≥        | ≥ | ≥ | ≥ | ≥ | ≥ | ≥ | ≥              | 1,5              |
| X2 і X3   | ≤        | ≤ | ≤ | ≤ | ≤ | ≤ | ≤ | ≤              | 0,5              |
| X2 і X4   | ≥        | ≤ | ≥ | ≥ | ≤ | ≤ | ≥ | ≥              | 1,5              |
| X3 і X4   | ≥        | ≥ | ≥ | ≥ | ≥ | ≥ | ≥ | ≥              | 1,5              |

Таблиця 4.4 – Попарне порівняння параметрів.

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

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

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

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

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

$$b_i = N a_i \quad (4.8)$$

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

$$K_{bi} = \frac{b_i^1}{\sum_{i=1}^n b_i^1}, \quad (4.9)$$

$$b_i^1 = N a_i b_i \quad (4.10)$$

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

| Параметри $x_i$ | Параметри $x_i$ |     |     |     | 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               | 1,5 | 1,5 | 1,5 | 5,5         | 0,344    | 21,25       | 0,360      | 77,875      | 0,361      |
| X2              | 0,5             | 1   | 0,5 | 1,5 | 3,5         | 0,219    | 12,25       | 0,208      | 44,875      | 0,208      |
| X3              | 0,5             | 1,5 | 1   | 1,5 | 4,5         | 0,281    | 16,25       | 0,275      | 59,125      | 0,274      |
| X4              | 0,5             | 0,5 | 0,5 | 1   | 2,5         | 0,156    | 9,25        | 0,157      | 34,125      | 0,158      |
| Всього:         | -               |     |     |     | 16          | 1        | 59          | 1          | 216         | 1          |

Таблиця 4.5 – Розрахунок вагомості параметрів

#### 4.5 Аналіз рівня якості варіантів реалізації функцій

Визначаємо рівень якості кожного варіанту виконання основних функцій окремо.

Абсолютні значення параметрів X2 (точність моделі), X3 (час попередньої обробки даних) та X4 (час навчання моделі) 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):

$$K_2(j) = \frac{1}{n} \sum_{i=1}^n K_{\&i} B_i, \quad (4.11)$$

де:

- $n$  – кількість параметрів;
- $K_{\&i}$  – коефіцієнт вагомості  $i$ -го параметра;
- $B_i$  – оцінка  $i$ -го параметра в балах.

| Основні функції | Варіант реалізації функції | Параметри | Абсолютне значення параметра | Бальна оцінка параметра | Коефіцієнт вагомості параметра | Коефіцієнт рівня якості |
|-----------------|----------------------------|-----------|------------------------------|-------------------------|--------------------------------|-------------------------|
| F1              | 1                          | X1        | 18000                        | 5                       | 0,361                          | 1,805                   |
| F2              | 2                          | X2        | 88                           | 4                       | 0,208                          | 0,832                   |
| F3              | 1                          | X3        | 87                           | 3                       | 0,274                          | 0,822                   |
|                 | 1                          | X4        | 73                           | 4                       | 0,158                          | 0,632                   |
|                 | 2                          | X3        | 85                           | 3                       | 0,274                          | 0,822                   |
|                 | 2                          | X4        | 67                           | 5                       | 0,158                          | 0,79                    |

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

За даними з таблиці 4.6 за формулою:

$$K_2 = K_{\text{ту}}[F_{\#}] + K_{\text{ту}}[F_{\$}] + \dots + K_{\text{ту}}[F_{\text{о}}], \quad (4.12)$$

визначаємо рівень якості кожного з варіантів:

$$K_{K1} = 1,805 + 0,832 + 0,822 + 0,632 = 4,091;$$

$$K_{K2} = 1,805 + 0,832 + 0,822 + 0,79 = 4,249 .$$

Як видно з розрахунків, кращим є 2 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:

$$T_0 = T_p \cdot K_p \cdot K_{\text{СК}} \cdot K_M \cdot K_{\text{СТ}} \cdot K_{\text{СТ.М}}, \quad (4.13)$$

де:

- $T_P$  – трудомісткість розробки ПП;
- $K_P$  – 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 степеню 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:  $K_P = 1,8$ . 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  $K_{СТ} = 0,7$ . Тоді загальна трудомісткість програмування першого завдання дорівнює:

$$T_1 = 90 \cdot 1,8 \cdot 0,7 = 113,4 \text{ людино-днів.}$$

Проведемо аналогічні розрахунки для подальших завдань.

Для другого завдання (використовується алгоритм третьої групи складності, стіпень новизни Б), тобто  $T_P = 35$  людино-днів,  $K_P = 0,9$ ,  $K_{СК} = 1$ ,  $K_{СТ} = 0,7$ :

$$T_2 = 35 \cdot 0,9 \cdot 0,7 = 22,05 \text{ людино-днів.}$$

Складаємо трудомісткість відповідних завдань для кожного з обраних варіантів реалізації програми, щоб отримати їх трудомісткість: 157,5

$$T_I = (113,4 + 22,05 + 4,8 + 22,05) \cdot 8 = 1298,4 \text{ людино-годин.}$$

$$T_{II} = (113,4 + 22,05 + 6,91 + 22,05) \cdot 8 = 1315,3 \text{ людино-годин.}$$

Найбільш високу трудомісткість має варіант II.

В розробці беруть участь два аналітики з окладом 22000 грн. та один програміст із окладом 20000. Визначимо середню зарплату за годину за формулою:

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

де:

- $M$  – місячний оклад працівників;
- $T_{\text{н}}$  – кількість робочих днів тиждень;
- $t$  – кількість робочих годин в день.

$$C_{\text{ч}} = \frac{22000 + 22000 + 20000}{3 \cdot 21 \cdot 8} = 126,98 \text{ грн.} \quad (4.15)$$

Тоді, розрахуємо заробітну плату за формулою:

$$C_{\text{ЗП}} = C_{\text{ч}} \cdot T_{\&} \cdot K_{\text{д}}, \quad (4.16)$$

де:

- $C_{\text{ч}}$  – величина погодинної оплати праці програміста;
- $T_{\&}$  – трудомісткість відповідного завдання;
- $K_{\text{д}}$  – норматив, який враховує додаткову заробітну плату.

Зарплата розробників за варіантами становить:

$$\text{I. } C_{\text{ЗП}} = 126,98 \cdot 1298,4 \cdot 1,2 = 197844,99 \text{ грн.}$$

$$\text{II. } C_{\text{ЗП}} = 126,98 \cdot 1315,3 \cdot 1,2 = 200420,15 \text{ грн.}$$

Відрахування на єдиний соціальний внесок становить 22%:

$$\text{I. } C_{\text{ВІД}} = C_{\text{ЗП}} \cdot 0,22 = 197844,99 \cdot 0,22 = 43525,90 \text{ грн.}$$

$$\text{II. } C_{\text{ВІД}} = C_{\text{ЗП}} \cdot 0,22 = 200420,15 \cdot 0,22 = 44092,43 \text{ грн.}$$

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

Так як одна ЕОМ обслуговує одного аналітика з окладом 22000 грн., з коефіцієнтом зайнятості 0,2 то для однієї машини отримаємо:

$$C_G = 12 \cdot M \cdot K_3 = 12 \cdot 22000 \cdot 0,2 = 52800 \text{ грн.}$$

З урахуванням додаткової заробітної плати:

$$C_{\text{ЗП}} = C_G \cdot (1 + K_3) = 52800 \cdot (1 + 0,2) = 63360 \text{ грн.}$$

Відрахування на соціальний внесок:

$$C_{\text{ВІД}} = C_{\text{ЗП}} \cdot 0,22 = 63360 \cdot 0,22 = 13939,2 \text{ грн.}$$

Амортизаційні відрахування розраховуємо при амортизації 30% та вартості ЕОМ – 30000 грн.

$$C_A = K_{\text{ТМ}} \cdot K_A \cdot Ц_{\text{ПР}} = 1,3 \cdot 0,3 \cdot 30000 = 11700 \text{ грн., де:}$$

- $K_{TM}$  – 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:

$$C_P = K_{TM} \cdot C_{ПР} \cdot K_P = 1,3 \cdot 30000 \cdot 0,04 = 1560 \text{ грн.},$$

де  $K_P$  – відсоток витрат на поточні ремонти.

Ефективний годинний фонд часу ПК за рік розраховуємо за формулою:

$$\begin{aligned} T_{\text{ЕФ}} &= (D_K - D_B - D_C - D_P) \cdot t_z \cdot K_B = (365 - 104 - 12 - 16) \cdot 8 \cdot 0,5 = \\ &= 932 \text{ години, де:} \end{aligned}$$

- $D_K$  – 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:

$$C_{\text{ЕЛ}} = T_{\text{ЕФ}} \cdot N_C \cdot K_3 \cdot C_{\text{ЕН}} = 932 \cdot 0,2 \cdot 0,5 \cdot 4,86 = 452,95 \text{ грн.}, \text{ де:}$$

- $N_C$  – середньо-споживча потужність приладу;
- $K_3$  – коефіцієнтом зайнятості приладу;
- $C_{\text{ЕН}}$  – тариф за 1 КВт-годин електроенергії.



Накладні витрати розраховуємо за формулою:

$$C_H = C_{\text{ПР}} \cdot 0,67 = 30000 \cdot 0,67 = 20100 \text{ грн.}$$

Тоді, річні експлуатаційні витрати будуть:

$$C_{\text{ЕКС}} = C_{\text{ЗП}} + C_{\text{ВІД}} + C_A + C_P + C_{\text{ЕЛ}} + C_H, (4.17)$$

$$C_{\text{ЕКС}} = 63360 + 13939,2 + 11700 + 1560 + 452,95 + 20100 = 111112,15 \text{ грн.}$$

Собівартість однієї машино-години ЕОМ дорівнюватиме:

$$C_{\text{М-Г}} = C_{\text{ЕКС}} / T_{\text{ЕФ}} = 111112,15 / 932 = 119,22 \text{ грн/год.}$$

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:

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

$$\text{I. } C_M = 119,22 \cdot 1298,4 = 154795,25 \text{ грн.}$$

$$\text{II. } C_M = 119,22 \cdot 1315,3 = 156810,07 \text{ грн.}$$

Накладні витрати складають 67% від заробітної плати:

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

$$\text{I. } C_H = 197844,99 \cdot 0,67 = 132556,14 \text{ грн.}$$

$$\text{II. } C_H = 200420,15 \cdot 0,67 = 134281,50 \text{ грн.}$$

Отже, вартість розробки ПП за варіантами становить:

$$C_{\text{пп}} = C_{\text{зп}} + C_{\text{вд}} + C_{\text{м}} + C_{\text{н}}, \quad (4.20)$$

$$\text{I. } C_{\text{пп}} = 197844,99 + 43525,90 + 154795,25 + 132556,14 = 528722,53 \text{ грн.}$$

$$\text{II. } C_{\text{пп}} = 200420,15 + 44092,43 + 156810,07 + 134281,50 = 535604,15 \text{ грн.}$$

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

Розрахуємо коефіцієнт техніко-економічного рівня за формулою:

$$K_{\text{ТЕР}'} = K_{\text{кн}} / C_{\text{фн}}, \quad (4.21)$$

$$K_{\text{ТЕР}1} = 4,091 / 528722,53 = 0,774 \cdot 10^{-5},$$

$$K_{\text{ТЕР}2} = 4,249 / 535604,15 = 0,793 \cdot 10^{-5}.$$

Як бачимо, найбільш ефективним є перший варіант реалізації програми з коефіцієнтом техніко-економічного рівня  $K_{\text{ТЕР}1} = 0,793 \cdot 10^{-5}$ .

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  $K_{\text{ТЕР}} = 6,851 \cdot 10^{-5}$ .

Цей варіант реалізації програмного продукту має такі параметри:

- Вибір кількісного методу прогнозування: лінійна регресія;
- Вибір якісного методу прогнозування: метод морфологічного аналізу;
- Вибір мови програмування та середовища розробки: Python, Jupyter Notebook.

Даний варіант виконання програмного комплексу дає користувачу зручний інтерфейс, швидку реалізацію програми та доступний функціонал для роботи.

#### 4.8 Висновки до четвертого розділу

У цьому розділі було проведено повний функціонально-вартісний аналіз програмного продукту. Також було знайдено оцінку основних функцій програмного продукту.

В результаті виконання функціонально-вартісного аналізу програмного комплексу що розроблюється, було визначено та проведено оцінку основних функцій програмного продукту, а також знайдено параметри, які його характеризують. На основі проведеного аналізу вибрано варіант реалізації програмного продукту.

## ВИСНОВКИ

У даній дипломній роботі було розроблено та застосовано алгоритм прогнозування вартості оренди житла у Києві в умовах якісними та кількісними методами на основі реальних даних, зібраних самостійно упродовж певного періоду часу. Результатом виконаної роботи є побудований за допомогою методу лінійної регресії прогноз, ефективність якого визначена задовільною обраними метриками. Побудований прогноз скориговано з урахуванням обраного якісного методу аналізу – методу морфологічного аналізу, що дозволяє оцінити вплив факторів війни у рамках поставленої задачі.

У першому розділі даної роботи було розглянуто та проаналізовано теоретичні дані про ринок нерухомості, його складові, чинники та фактори, що впливають на вартість оренди житла та методи її розрахунку. У другому розділі наведено теоретичні відомості щодо використаних кількісного та якісного методів прогнозування, особлива увага приділена математичному обґрунтуванню кожного методу, їх недолікам та перевагам. У третьому розділі виконано реалізацію згаданих методів, оцінено їх ефективність, порівняно результати та скомпоновано до спільного результату. Четвертий розділ присвячено функціонально-вартісному аналізу розробленого продукту.

Поставлену задачу можна вважати успішно розв'язаною, у випадку необхідності прогнозування цін на реальних даних для кожного розглядаємого об'єкта, тобто, при необхідності визначення чітких математичних відповідей та можливих розподілів. Побудований алгоритм не може надати відповідь на узагальнені запитання про динаміку ціни, бо, як зазначено вище, кожен об'єкт має бути розглянутий окремо із урахуванням його характеристик.

Недоліком розробленого алгоритму можна вважати його пристосованість до даних певного вигляду та суб'єктивності при визначенні параметрів морфологічного аналізу та проведення експертних оцінок. Також, враховуючи збір даних у реальному часі, як недолік можна зазначити необхідність постійного оновлення вихідних даних, що відбувається вручну. Покращити наведений алгоритм можна за допомогою автоматизованого збору даних експертизи та збору вихідних даних і приведення їх до потрібного вигляду у реальному часі. Також, задля досягнення більшої точності результатів можна побудувати більш складну модель, використовуючи інші методи кількісного аналізу.

Розроблений алгоритм для прогнозування може бути використаний у науково-дослідницьких цілях та потенційно бути частиною більших проєктів та програмних продуктів. Найбільш доцільним способом його використання, на думку авторів, було б застосування його у межах сайтів для пошуку житла для оренди, що дозволяло б різним сторонам ринку оренди нерухомості оцінювати наперед ситуацію, що може скластися на цьому ринку.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Cox, D. R. «Principles of Statistical Inference» - Cambridge University Press, 2006.
2. Friedman, J., Ordway, Nick (1995): Analysis and evaluation of incomegenerating real estate. TRANS. from English., – M.: McGraw-hill
3. Graham J. Hooley, Michael K. Hussey «Quantitative Methods in Marketing, Second Edition» - London: International Thompson Business Press, 1999.
4. Harold A. Linstone, Murray Turoff. «The Delphi Method: Techniques and Applications», 1975.
5. Hyndman, R.J., & Athanasopoulos, G. «Forecasting: principles and practice», 2nd edition – OTexts: Melbourne, Australia, 2018.
6. J. Scott Armstrong «Forecasting for Marketing» - The Wharton School, University of Pennsylvania, 1999.
7. R.U. Ayres, Technological Forecasting and Long-Time Planning. McGraw-Hill, 1969.
8. Sir James Steuart An Inquiry Into the Principles of Political Oeconomy  
Університет Лозанни  
[https://books.google.com.ua/books?id=uN8TAAAAQAAJ&redir\\_esc=y](https://books.google.com.ua/books?id=uN8TAAAAQAAJ&redir_esc=y)
9. Zwicky F., Wilson A.G. - «New Methods of Thought and Procedure // Contributions to the Symposium on Methodologies», 1967.
10. Горошко М. Ф., Кулішов В. В. Мікроекономіка: Навчальний посібник — К.: Ельга, 2003.
11. Господарський кодекс України від 16.01.2003 // Відомості Верховної Ради, 2003. – № 18-22
12. 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. Кузнєцової. – К.: «Юрінком Інтер», 2004. – Кн. 1. – С. 147.

13. Закон України «Про іпотеку» від 05.06.2003 р. № 898-ІУ // Відомості Верховної Ради України. – 2003. – № 38. – С. 313.
14. Закон України "Про оподаткування прибутку підприємств" від 28.12.94 // ВВР України. – 1994. – № 334/94.
15. Закон України «Про оренду державного та комунального майна» від 16.10.04.1992р. // ВВР України. – 1992. – № 30. – 416 с.
16. Згуровский М.З., Панкратова Н.Д. «Технологическое предвидение» — К.: ІВЦ «Видавництво «Політехні- ка»», 2005.
17. Одрин В.М. «Морфологический синтез систем: морфологические методы поиска» — К.: Ин-т кибернетики им. В.М. Глушкова АН УССР, 1986.
18. Онлайн-джерело: Bovas Abraham, Johannes Ledolter «Statistical Methods for Forecasting». Посилання - <https://onlinelibrary.wiley.com>.
19. Онлайн-джерело: <https://forbes.ua/money/budivnitstva-ne-zapuskayutsya-popitu-nemae-ale-tsini-ne-padayut-doslidzhennya-rinku-nerukhomosti-pid-chas-viyni-vid-ernst-amp-young-13032023-12337>»
20. Онлайн-джерело: «<https://misto.lun.ua/2022>»
21. Пилипенко Д. Є. «Оцінка достовірності результатів аналізу перехресного впливу при розв’язанні задач технологічного передбачення». — 2008.
22. Ринок нерухомості: навч. посіб. [для студ. вищ. навч. закл.] / В. В. Кузьмін, О. В. Лепьохіна. — Х. : Вишемирський В. С., 2008. — 219 с. : іл., табл. ; 20 см. — Бібліогр.: с. 215—219 (113 назв). — 2 000 пр.
23. Савченко І. О. «Методологічне і математичне забезпечення розв’язання задач передбачення на основі модифікованого методу морфологічного аналізу» — 2011.

## ДОДАТОК А ЛІСТИНГ ПРОГРАМИ

```

#збір даних (пасринг сайту)
from bs4 import BeautifulSoup
import requests
import csv

def get_html(url):
    r = requests.get(url)
    if r.ok:
        return r.text
    print(r.status_code)

def write_csv(data):
    with open('parsing_new.csv', 'a') as f:
        writer = csv.writer(f)
        writer.writerow((data['price'],
                           data['district'],
                           data['chips'],
                           data['rooms+square+level']))

def get_total_page(html):
    soup = BeautifulSoup(html, "html.parser")
    pages = soup.find('div', class_="jss225 jss14").find_all('a',
class_="jss61 jss35 jss46 jss49 jss58 jss226 jss228")[-1].get('href')
    total_pages = pages.split('=')[2]
    return int(total_pages)

def get_page_data(html):
    soup = BeautifulSoup(html, "html.parser")
    announcement = soup.find_all('div', class_='catalog-card-content')
    for i in announcement:
        try:
            price = i.find('div', class_='catalog-card-price').text
        except:
            price = ''

        try:
            district = i.find('div', class_='catalog-card-region').text
        except:
            district = ''

        try:
            district = i.find('div', class_='catalog-card-region').text
        except:
            district = ''

        try:
            chips = i.find('div', class_='catalog-card-chips').text
            #chips.replace('грн', 'грн. дата: ')
        except:
            chips = ''

```



```

try:
    rsl= i.find('div', class_='catalog-card-details').text
    rsl.replace ('поверх', '')
except:
    rsl = ''

    data = {'price': price,
            'district': district,
            'chips': chips,
            'rooms+square+level': rsl}
    write_csv(data)

def main():
    url = 'https://rieltor.ua/kiev/flats-rent/?sort=-
default#10.19/50.4381/30.5257'
    page_count = '&page='
    #total_pages = get_total_page(get_html(url))

    for i in range(0, 2):
        url_gen = url + page_count + str(i)
        html = get_html(url_gen)
        get_page_data(html)

import pandas as pd
csv_file_path = "/Users/anastasialickovaha/Documents/Диплом/dataset.csv"
df = pd.read_csv(csv_file_path, sep = ',')
df.dropna(subset=['Ціна'], inplace=True)
df.reset_index(drop=True, inplace=True)
df['Район'] = df['Район'].str.strip()
district_map_districts = {
    'Печерський р-н': 1,
    'Шевченківський р-н': 2,
    'Голосіївський р-н': 3,
    'Дніпровський р-н': 4,
    'Дарницький р-н': 5,
    'Оболонський': 6,
    'Подільський': 7,
    'Святошинський': 8,
    "Солом'янський": 9,
    'Деснянський': 10
}
df['Район'] = df['Район'].replace(district_map_districts)
district_map_metro = {
    'Берестейська': 113,
    'Васильківська': 224,
    'Вокзальна': 117,
    'Дніпро': 122,
    'Іподром': 226,
    'Кловська': 316,
    'Либідська': 221,
    "Лук'янівська": 312,
    'Майдан Незалежності': 217,
    'Печерська': 317,
    'Політехнічний інститут': 116,
    'Славутич': 321,

```

```

'Хрещатик': 120
}
df['Станція метро'] = df['Станція метро'].replace(district_map_metro)
district_map_comfortable = {
    1: 3.04, 2: 3.01, 3: 3.06, 4: 3.1, 5: 3.08, 6: 3.06, 7: 3.07, 8:
    3.02, 9: 3.06, 10: 3.03
}
df['Комфортність'] = df['Район'].map(district_map_comfortable)
df.rename(columns={'Кімнатність ': 'Кімнатність'}, inplace=True)
df['Кімнатність'] = df['Кімнатність'].str.replace('кімната', '')
df['Кімнатність'] = df['Кімнатність'].str.replace('кімнати', '')
df['Поверховість'] = df['Поверховість'].str.extract(r'(\d+)')
df['Загальна площа'] = df['Площа'].str.split(' / ').str[0]
df['Жила площа'] = df['Площа'].str.split(' / ').str[1]
df = df.drop(['Площа'], axis=1)

import tensorflow.compat.v1 as tf
tf.disable_v2_behavior()

import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split

X_data = np.linspace(-1, 1, 100)
num_coef=8
coef =[1, 20, 3, 4, 6, 7, 300, 2]
y_data=0
for i in range (num_coef) :
    y_data += coef[i] * np.power(X_data, i)

y_data_without_noise = np.copy(y_data)

y_data += np.random.randn(*X_data.shape) * 20.5

from sklearn.preprocessing import StandardScaler

X_data = X_data.reshape(-1, 1)
y_data = y_data.reshape(-1, 1)

X_data = StandardScaler().fit_transform(X_data.reshape(-1, 1))

plt.figure(figsize=(8, 6))
plt.title("custom dataset")
plt.scatter(X_data, y_data, c='red', label='y_data')
plt.plot(X_data, y_data_without_noise, label='y_data without noise')
plt.legend()
plt.show()

X_train, X_test, Y_train, Y_test = train_test_split(X_data, y_data,
test_size=0.2)

degree = 3
learning_rate = 0.01
n_epochs = 1000
reg = 0.0001

```

```

X = tf.placeholder(tf.float32, shape=(None, 1), name='X')
y = tf.placeholder(tf.float32, shape=(None, 1), name='Y')

w = tf.Variable(tf.random.uniform([degree + 1, 1], -1.0, 1.0), name='w')

y_pred = tf.reduce_sum([tf.multiply(tf.pow(X, i), w[i]) for i in
range(degree + 1)], 0)
cost = tf.reduce_mean(tf.square(y_pred - y)) + reg *
tf.reduce_sum(tf.square(w))
optimizer =
tf.train.GradientDescentOptimizer(learning_rate=learning_rate).minimize(c
ost)

save_values = tf.train.Saver()

with tf.Session() as sess:
    sess.run(tf.global_variables_initializer())

    for i in range(n_epochs + 1):
        if (10 <= i <= 90 and i % 10 == 0) or i % 100 == 0:
            if i == 0:
                print(i, ' ', '%.3f' %
cost.eval(feed_dict={X:X_train, y:Y_train}))
            else:
                print(i, ' ', '%.3f' % cost.eval(feed_dict={X:X_train,
y:Y_train}))
                save_values_after = save_values.save(sess, '')
                sess.run(optimizer, feed_dict={X: X_train, y:Y_train})
                save_values_after = save_values.save(sess, '')
                w = sess.run([w[i] for i in range(degree + 1)])

predicted_test = 0
predicted = 0
for i in range(len(w)):
    predicted += w[i] * np.power(X_data, i)
    predicted_test += w[i] * np.power(X_test, i)

from sklearn.metrics import mean_squared_error, r2_score

print('RMSE: ', np.power(mean_squared_error(Y_test, predicted_test),
1/2))
print('R^2', r2_score(Y_test, predicted_test))

plt.figure(figsize=(8, 6))
plt.scatter(X_data, y_data, c='blue', label='полином с шумами')
plt.plot(X_data, y_data_without_noise, color = 'green', label='полином
без шумов')
plt.plot(X_data, predicted, color = 'red', label='регрессионная модель')
plt.legend(loc='best')
plt.show()

import pandas as pd
df = pd.DataFrame(data={'Регуляризация': [0.0001, 0.001, 0.01, 0.1, 1],
                        'RMSE': [51.190, 51.198, 55.652, 68.030, 89.282],

```

```

'R^2': [0.755, 0.755, 0.738, 0.658, 0.453]],
index=[1, 2, 3, 4, 5])

tf.reset_default_graph()
from sklearn.datasets import make_friedman3

X, Y = make_friedman3(n_samples=1000)

fig, ax = plt.subplots(2, 2)
fig.set_figwidth(8)
fig.set_figheight(6)
plt.suptitle('make_friedman2')

colors = ['blue', 'red', 'black', 'green']
labels = ["X1", "X2", "X3", "X4"]

plot = 1
for i, j in zip([X[:, 0], X[:, 1], X[:, 2], X[:, 3]], [colors[0],
colors[1], colors[2], colors[3]]):
    plt.subplot(2, 2, plot)
    plt.gca().set_title(labels[plot - 1])
    plt.scatter(i, Y, color=j)
    plot += 1

fig.tight_layout()
plt.show()

from sklearn.preprocessing import StandardScaler

X = StandardScaler().fit_transform(X)
Y = Y.reshape(-1, 1)
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2)

degree = 3
learning_rate = 0.01
n_epochs = 1000
reg = 0.0001

X = tf.placeholder(tf.float32, shape=(None, 4), name='X')
Y = tf.placeholder(tf.float32, shape=(None, 1), name='Y')
w = tf.Variable(tf.random.uniform([4, 1, degree], -1.0, 1.0), name='w')
bias = tf.Variable(tf.random.uniform([1, 1], -1.0, 1.0), name='bias')

y_pred = 0
for i in range(1, degree + 1):
    y_pred += tf.matmul(tf.pow(X, i), w[:, :, i - 1])
y_pred += bias
cost = tf.reduce_mean(tf.square(y_pred - Y)) + reg *
(tf.reduce_sum(tf.abs(w)) + tf.abs(bias))
optimizer =
tf.train.GradientDescentOptimizer(learning_rate).minimize(cost)

save_values_second_dataset = tf.train.Saver()

```

```

with tf.Session() as sess:
    sess.run(tf.global_variables_initializer())

    print('\033[1m'+ 'Epoch      MSE'+ '\033[0;0m')
    for i in range(n_epochs):
        if (10 <= i <= 90 and i % 10 == 0) or i % 100 == 0:
            if i == 0:
                print(i, '      ', '%.3f' % cost.eval(feed_dict={X:X_train,
Y:Y_train}))
            else:
                print(i, '      ', '%.3f' % cost.eval(feed_dict={X:X_train,
Y:Y_train}))
            save_values_after = save_values_second_dataset.save(sess, '')
            sess.run(optimizer, feed_dict={X:X_train, Y:Y_train})

            save_values_after = save_values_second_dataset.save(sess, "")
            w = sess.run([w[:, :, i] for i in range(degree)] + [bias])

    test_predict = 0
    for i in range(degree):
        test_predict += tf.matmul(np.power(X_test, i + 1), w[i])
    test_predict += w[degree]

    def rmse(y_true, y_pred):
        return tf.keras.backend.sqrt(tf.keras.losses.MSE(y_true, y_pred))

    print(rmse(Y_test, test_predict))

    print('RMSE: ', np.power(mean_squared_error(Y_test, test_predict), 1/2))
    print('R^2', r2_score(Y_test, test_predict))

    import pandas as pd
    import numpy as np
    import itertools

    from google.colab import drive
    drive.mount('/content/drive/')
    path = '/content/drive/My Drive/Colab Notebooks/MSS/MMA.xlsx'

    evaluated_1 = pd.read_excel(path, sheet_name='1', index_col=0)

    connections_1 = pd.read_excel(path, sheet_name='2', header=[0, 1],
index_col=[0, 1])

    lists, data = [], []
    nums_1 = [3, 4, 4, 3, 3, 2, 3, 3]
    for i in nums_1:
        lists.append(['a' + str(x+1) for x in range(i)])
    for i in itertools.product(*lists):
        data.append(list(i))
    config_1 = pd.DataFrame(data, columns = ['F1', 'F2', 'F3', 'F4', 'F5',
'F6', 'F7', 'F8'])

    p_arr, c_arr = [], []
    for i, s in config_1.iterrows():

```

```

p, c = 1, 1
for j in range(1, 9):
    p = p * evaluated_1.loc[s['F'+str(j)], 'F'+str(j)]
p_arr.append(p)
for y in range(1, 8):
    for x in range(y+1, 9):
        c = c * (1 + connections_1.loc['F'+str(x),
'F'+str(y)].loc[s['F'+str(x)], s['F'+str(y)]])
    c_arr.append(c)

config_1['p'] = p_arr
config_1['C'] = c_arr
config_1['p*C'] = config_1['p'] * config_1['C']
config_1['P(s)'] = config_1['p*C'] / config_1['p*C'].sum()
config_1.loc['Total', :] = config_1[['p*C', 'P(s)']].sum(axis=0)

config_1.loc[config_1['P(s)'] > 0, :]

results_1 = pd.DataFrame(columns=['F1', 'F2', 'F3', 'F4', 'F5', 'F6',
'F7', 'F8'], index=['a1', 'a2', 'a3', 'a4', 'a5', 'a6'])
for i, j in zip(range(1, 9), nums_1):
    for k in range(1, j+1):
        results_1.at['a'+str(k), 'F'+str(i)] =
config_1.loc[config_1['F'+str(i)] == 'a'+str(k), 'P(s)'].sum()
        results_1['F'+str(i)] = results_1['F'+str(i)].apply(lambda x: round(x,
3) if str(x) != np.nan else x)

connections_2 = pd.read_excel(path, sheet_name='3', header=[0, 1],
index_col=[0, 1])

config_2 = pd.DataFrame(data, columns = ['F1', 'F2', 'F3', 'F4', 'F5',
'F6', 'F7', 'F8'])
config_2['P(s)'] = config_1['P(s)']

nums_2 = [5, 0]
for i, s in config_2.iterrows():
    r_8 = [1 for i in range(5)]
    for j in range(1, 9):
        x = 9
        for y in range(5):
            r_8[y] = r_8[y] * (1 + connections_2.loc['F'+str(j),
'F'+str(x)].loc[s['F'+str(j)], 'a'+str(y+1)])
        config_2.loc[i, ['r1_8', 'r2_8', 'r3_8', 'r4_8', 'r5_8']] = r_8

sum_8 = config_2[['r1_8', 'r2_8', 'r3_8', 'r4_8', 'r5_8']].sum(axis=1)
config_2['R1_8'], config_2['R2_8'], config_2['R3_8'], config_2['R4_8'],
config_2['R5_8'] = config_2['r1_8'] / sum_8, config_2['r2_8'] / sum_8,
config_2['r3_8'] / sum_8, config_2['r4_8'] / sum_8, config_2['r5_8'] /
sum_8

config_2['R1_8*P(s)', config_2['R2_8*P(s)', config_2['R3_8*P(s)',
config_2['R4_8*P(s)', config_2['R5_8*P(s)'] = config_2['R1_8'] *
config_2['P(s)', config_2['R2_8'] * config_2['P(s)', config_2['R3_8'] *
config_2['P(s)', config_2['R4_8'] * config_2['P(s)', config_2['R5_8'] *
config_2['P(s)']

```

```

config_2.loc['Total', :] = config_2[['R1_8*P(s)', 'R2_8*P(s)',
'R3_8*P(s)', 'R4_8*P(s)', 'R5_8*P(s)']].sum(axis=0)

results_2 = pd.DataFrame(columns=['F9'], index=['a1', 'a2', 'a3', 'a4',
'a5'])

for k in range(1, 6):
    results_2.at['a'+str(k), 'F9'] = config_2.loc['Total',
'R'+str(k)+'_8*P(s)']
results_2['F9'] = results_2['F9'].apply(lambda x: round(x, 3) if str(x)
!= np.nan else x)

```