

УДК 004.4:004.85

*Сосновська Є.Р., здобувач,  
Осипова О.І., к.е.н., доцент  
Київський національний економічний університет імені Вадима  
Гетьмана*

## **РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ АВТОМАТИЗОВАНОГО ЗБОРУ ТА АНАЛІЗУ ДАНИХ РИНКУ НЕРУХОМОСТІ**

Актуальність дослідження ринку нерухомості зумовлена постійною динамікою цін та величезним обсягом неструктурованих пропозицій на профільних веб-ресурсах. Для ефективного моніторингу та прийняття обґрунтованих рішень критично необхідною є автоматизація збору інформації та побудова надійних математичних моделей для оцінки вартості майна.

Метою даної роботи є розробка програмного забезпечення для автоматизованого збору даних з ринку нерухомості та створення на їх основі прогностичних моделей для оцінки вартості житла. Для досягнення цієї мети було реалізовано таку послідовність етапів дослідження: спочатку було розроблено програмний модуль для автоматизованого збору даних (парсер), далі сформовано та збережено набір даних з характеристиками об'єктів нерухомості, після цього виконано попередню обробку та аналіз даних, і на завершальному етапі побудовано та оцінено моделі машинного навчання для прогнозування вартості квартир.

Ядром розробленого програмного забезпечення є модуль автоматизованого збору даних з сайту lup.ua [1], реалізований мовою Python виключно за допомогою інструменту автоматизації браузерів Selenium [2]. Архітектура парсера побудована на механізмах явного очікування - скрипт синхронізується з роботою веб-ресурсу і ініціює вилучення даних лише після повної появи та рендерингу цільових об'єктів у DOM-структурі сторінки. Це гарантує стабільний збір інформації без необхідності застосування алгоритмів імітації поведінки користувача.

Парсер налаштований на вилучення строго визначеного набору характеристик кожного об'єкта:

- кількість кімнат;
- загальна площа, житлова площа та площа кухні;
- поверх та загальна поверховість будинку;
- рік будівництва;

- ціна за квадратний метр.

На етапі первинної обробки зібрана ціна за квадратний метр алгоритмічно перетворюється у підсумкову вартість квартири, яка виступає цільовою змінною для подальшого аналізу [3].

Перед навчанням моделей "сирі" дані пройшли етап глибокої обробки за допомогою бібліотеки Pandas [4]. Процес включав очищення датасету, трансформацію ознак та важливий крок - усунення мультиколінеарності шляхом аналізу та відкидання певних взаємозалежних змінних (ознак), що дозволило підвищити стабільність та інтерпретованість моделей.

Враховуючи суттєві економічні розбіжності у ціноутворенні, очищений датасет був концептуально розділений на дві незалежні групи: квартири в місті Києві та квартири в інших обласних центрах. На основі сформованих вибірок було побудовано окремі регресійні моделі з використанням алгоритмів машинного навчання бібліотеки Scikit-learn [5], зокрема Random Forest Regressor, k-Nearest Neighbors Regressor та Decision Tree Regressor. Розділення даних за географічною ознакою (Київ та інші міста) дозволило врахувати регіональні особливості формування цін на ринку нерухомості, що сприяє підвищенню точності прогнозування вартості квартир.

Розроблений програмний комплекс забезпечує повний цикл роботи з даними ринку нерухомості. Застосований підхід - від стабільного автоматизованого збору цільових метрик за допомогою Selenium до географічного кластеризації даних та очищення ознак від мультиколінеарності - доводить свою ефективність у створенні точних прогностичних інструментів для сфери PropTech.

#### **Список використаних джерел:**

1. ЛУН – пошук нерухомості в Україні. URL: <https://lun.ua/> (дата звернення: 06.03.2026).
2. Selenium Documentation. URL: <https://www.selenium.dev/documentation/> (дата звернення: 06.03.2026).
3. VanderPlas J. Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media, 2016. 546 p.
4. pandas - Python Data Analysis Library. URL: <https://pandas.pydata.org/docs/> (дата звернення: 06.03.2026).
5. Scikit-learn: Machine Learning in Python. URL: [https://scikit-learn.org/stable/user\\_guide.html](https://scikit-learn.org/stable/user_guide.html) (дата звернення: 06.03.2026).