DOI: [10.32702/2307-2105-2022.1.81](https://doi.org/10.32702/2307-2105-2022.1.81)

УДК 330.4, 519.2

*I. В. Мірошниченко,
к .е. н., доцент кафедри математичного моделювання та статистики,
Київський національний економічний університет імені Вадима Гетьмана
ORCID ID: 0000-0002-1307-7889
О. В. Крашеніннікова,
магістрант спеціальності «Економічна кібернетика та Дата Сайанс»,
Київський національний економічний університет імені Вадима Гетьмана
ORCID ID: 0000-0003-1303-814X*

ПРОГНОЗУВАННЯ ЦІНИ НА НЕРУХОМІСТЬ З ВИКОРИСТАННЯМ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ

*I. Miroshnichenko,
PhD in Economics, Associate Professor of the Department of Mathematical Modeling and
Statistics, Kyiv National Economic University named after Vadym Hetman
O. Krashenninnikova,
Master's student of "Economic cybernetics and Data science",
Kyiv National Economic University named after Vadym Hetman*

REAL ESTATE PRICE FORECASTING USING MACHINE LEARNING ALGORITHMS

Прогнозування вартості нерухомості є важливим економічним показником. Точний прогноз ціни на житло важливий для майбутніх власників будинків, забудовників, інвесторів, оцінювачів, податкових департаментів та інших учасників ринку нерухомості, таких як іпотечні кредитори та страховики. Тому наявність моделі прогнозування ціни на житло допомагає заповнити важливу інформаційну прогалину та підвищити ефективність ринку нерухомості.

Стаття присвячена застосуванню методів машинного навчання для побудови моделей прогнозування ціни на заміську нерухомість. Виконано критичний аналіз ринку житлової нерухомості України та визначено фактори, які впливають на формування ціни заміської нерухомості. Встановлено доцільність використання алгоритмів машинного навчання для прогнозування ціни на заміську нерухомість. Побудовано чотири моделі прогнозування ціни та визначено, що найбільшу ефективність має алгоритм XGBoost. Проведено порівняння результатів прогнозування алгоритму XGBoost та множинної лінійної регресії на основі випадково обраного будинку з вибірки. Визначено, що модель XGBoost краще описує дані і ймовірність спрогнозувати справжню ціну є вищою.

The subject of structural analysis and active development of the real estate market is of great importance for the country's economy. For this reason, modeling and prediction of real estate prices are important and relevant areas of research. A practical and accurate model for predicting house prices would become a valuable tool for homeowners, developers, property investors, appraisers, mortgage lenders, insurers, tax agencies, and other players in the real estate market. Lately, machine learning techniques have become the primary source for advanced analysis,

forecasting, and visualization of real estate prices. Therefore, having in place an accurate price prediction model helps fill in an important information gap and improve the efficiency of the real estate market.

The article focuses on the appropriate use of machine learning methods to build models for predicting suburban real estate prices. For this purpose, a critical analysis of Ukraine's residential real estate market was carried out. The study found that Ukraine's suburban real estate market revived against the backdrop of the COVID-19 pandemic and the lockdown in the spring of 2020, leading to an increase in property prices and the number of houses built. In addition, several factors affecting price formation for suburban real estate at the macro and micro levels were identified. It was found that the use of machine learning algorithms for predicting prices of suburban real estate is advisable. The following four price prediction models were built: multiple linear regression, regularization, random forest, and XGBoost. After checking the efficiency of the models on the test dataset and comparing the results with each other, it was determined that in terms of RMSE and MAPE, the XGBoost algorithm provides the most accurate and efficient method. A comparison was made between the predicted results of the XGBoost algorithm and the multiple linear regression algorithm based on a randomly selected house from the sample. It was found that the XGBoost model better describes the data and is more likely to accurately predict real-life prices.

The article summarizes the findings of the data analysis and testing and finds that the XGBoost model can effectively predict and analyze housing prices to a certain degree. At the same time, the accuracy and efficiency of the RE price modeling can be further improved by using more advanced machine learning techniques.

Ключові слова: ринок житлової нерухомості; ціна на заміську нерухомість; лінійна множинна модель; регуляризація; випадковий ліс; XGBoost; прогнозування.

Key words: residential real estate market; suburban real estate price; multiple Linear Regression; regularization; random forest; XGBoost; forecasting.

Постановка проблеми. Ринок житла відіграє важливу роль у структурі економіки. Зокрема, він встановлює зв'язок між споживачами житла та будівельною галуззю, визначає взаємозв'язок між попитом та пропозицією на нерухомість, встановлює ціни на житло, визначає найбільш ефективні форми житла, є механізмом перерозподілу житла серед громадян, тощо.

Група людей, зацікавлених у розумінні його сучасного стану та прогнозуванні тенденцій, не обмежується покупцями та продавцями, орендарями та орендодавцями, брокерами або забудовниками. Приватні інвестори, фінансові групи, інвестиційні та страхові компанії, банки та держава також зацікавлені у вивченні того, як працює вітчизняний ринок нерухомості.

Предмет структурного аналізу та активного розвитку ринку нерухомості та суміжних ринків має велике значення для економіки нашої країни. З цієї причини моделювання та прогнозування цін на нерухомість є важливим та актуальним напрямком досліджень. Методи машинного навчання стали основним джерелом вдосконалених можливостей для аналізу, прогнозування та візуалізації цін на нерухомість.

Аналіз останніх досліджень і публікацій. Аналіз ринку нерухомості є предметом досліджень вітчизняних та зарубіжних вчених. Аналіз динаміки та ціноутворення на ринку нерухомості розглядаються в роботах А.М. Асаула [1], Г.М. Стерника [2], Ю.М. Манцевича [3]; застосування нейронних мереж для прогнозування цін на нерухомість – в роботах В.С. Григорківа [4], Н.О. Філіпчука [4], О.М. Ярошенко [4], В.А. Вороніна [5]; використання методів нечіткого моделювання для вибору інструментів прогнозування на ринку нерухомості – в роботі Н. К. Максишко [6].

Однак поки що немає загальної концепції прогнозування цін на житло на основі математичного апарату. Вибір методів моделювання повністю залежить від мети дослідження та структури даних.

Формулювання цілей статті (постановка завдання). Метою дослідження є аналіз ринку житлової нерухомості України та розробка математичних моделей прогнозування ціни на заміські будинки.

Виклад основного матеріалу дослідження.

Нерухомість – це нерухоме майно, яке складається із землі та благоустрою, що включає земельну ділянку, а також фізичні, матеріальні та непридатні для пересування об'єкти (будівлі, споруди, тощо) [7].

Ринок житла представляє собою житлову нерухомість, яка продається та купується безпосередньо покупцями або через посередників з нерухомості [8]. Він є невід'ємною частиною економіки країни. Цей ринок є важливим не лише на економічному рівні, а й на соціальному, оскільки задовольняє одну з основних потреб людини – потребу в житлі [9, с. 19].

Ринок житла об'єднує різних зацікавлених сторін, таких як власників будинків, які продають свою нерухомість, людей, які прагнуть придбати нерухомість, орендарів, інвесторів, які купують та продають нерухомість виключно для інвестиційних цілей, підрядників, людей, які займаються ремонтами та агентів з нерухомості, які виступають посередниками в процесі купівлі або продажу нерухомості [10].

Житло може бути міським і заміським та відноситися до різних класів: економ, комфорт, бізнес і преміум. Ринок житлової нерухомості ділиться на дві категорії: первинний та вторинний види власності.

Змінам на ринку житла завжди надається таке значення через взаємозв'язок цін на житло та споживчі витрати. Зміни цін на будинки та кількість продажів впливають на те, скільки грошей люди повинні витратити. Спостереження за ринком житла допомагає оцінити загальний попит на товари та послуги.

Окрім впливу на загальні витрати домогосподарств, купівля та продаж будинків також безпосередньо впливає на економіку. Інвестиції в житло – це невелика, але непередбачувана частина загального обсягу виробництва в економіці.

Нерухомість, як і інші активи, також підлягає попиту та пропозиції.

До факторів, що впливають на попит на житло, належать:

1. Реальні доходи: якщо реальні доходи збільшуються, попит на житло зростає через підвищення рівня життя;
 2. Наявність кредиту: чим більше банки-позичальники та будівельні товариства готові надати кредити, тим більше людей позичатимуть і витратять на житло, а отже, і ціни на житло будуть вищими;
 3. Економічне зростання: коли економіка перебуває на етапі відновлення та підвищення ділового циклу, заробітна плата зростає. Це збільшить попит на будинки;
 4. Населення: коли населення зростає або збільшується кількість домогосподарств, попит на житло зростає;
 5. Зайнятість / безробіття: чим вищий рівень безробіття в економіці, тим менше людей зможе дозволити собі купити нове житло;
 6. Впевненість: якщо споживачі відчувають оптимізм щодо майбутнього стану економіки, вони, швидше за все, підуть на придбання будинку, тим самим збільшуючи попит.
- У міру того, як будинки дешевшають, все більше людей вирішують, що можуть собі дозволити собі його купити, тому кількість попиту збільшується.

Фактори, що впливають на пропозицію житла, включають:

1. Витрати на виробництво: чим вища собівартість, тим менше будинків буде зведено, зменшуючи пропозицію будинків, що надходять на ринок.
2. Урядова політика: якщо уряд збільшить оподаткування та/або зменшить субсидії на нові будинки, нових будинків буде менше;
3. Кількість будівельних компаній: залежно від їх цілей, чим більше будівельних компаній, тим більша ймовірність збільшення пропозиції житла;
4. Технології та інновації: завдяки вдосконаленню технологій та інновацій у будівельній галузі будинки стають дешевшими та простішими в будівництві, збільшуючи таким чином пропозицію;
5. Державні витрати на будівництво нового соціального житла: уряд може впливати на пропозицію житла, збільшуючи витрати на нове соціальне житло.

Крива пропозиції на житло має позитивний нахил. Це означає, що із зростанням ціни на будинки, зростає і кількість, що постачаються на ринок.

Точкою, де зустрічаються попит і пропозиція, є рівновага на ринку. Тобто існує ідеальна відповідність між кількістю, яку хочуть придбати покупці, та кількістю, яку хочуть продати продавці [11].

Деякі фактори підвищують ціни на нерухомість, а інші знижують їх. Тож, безумовно, варто уважно спостерігати за причинами, які можуть вплинути на ціни на нерухомість [12].

Умовно, фактори, що впливають на ціну нерухомості можна поділити на макро та мікрофактори.

До макрофакторів, які впливають на вартість нерухомості, можна віднести наступні:

1. Економічні фактори. Загальні показники економіки впливають на ринок нерухомості [13]. Якщо економіка переживає зростання, гарний рівень зайнятості та умови праці, більше людей можуть дозволити собі придбати нерухомість, що призводить до зростання вартості власності [14]. У випадку, якщо зростання зайнятості чи заробітної плати сповільнюється, то менша кількість людей може собі дозволити житло або ж може бути менше можливостей переїхати [15].

Окрім цього, до економічних факторів можна віднести: рівень економічного розвитку країни, темпи зростання ВВП, рівень інфляції, грошово-кредитна політика, коливання економічної діяльності, тощо [16, с. 96].

2. Інвестиційний потенціал. На вартість майна також впливає потенціал, який він представляє для інвесторів. Такі фактори, як дохід від оренди, який інвестор може очікувати від власності, та приріст капіталу, який він матиме, коли згодом буде продавати нерухомість, відіграють свою роль [13].

2. Ринкові фактори. Попит та пропозиція визначають ціни на будь-якому ринку. Якщо високий попит на будинки поблизу, очікуйте, що значення збільшиться. Якщо в певному районі є надмірна кількість нерухомості, але недостатньо покупців, то ціни, як правило, знижуються [17].

4. Населення та демографія. До соціальних факторів належать: житлові умови населення, необхідність придбання землі, нерухомості. Демографічні фактори: тенденції чисельності населення, розміру сім'ї, рівнів смертності та народжуваності, тощо. Демографічні та соціальні фактори дають інформацію про попит на нерухомість та її розмір [16, с. 97].

5. Законодавчі фактори. Податкові пільги, відрاهування та субсидії – це деякі з способів уряду тимчасово збільшити попит на нерухомість, доки вони існують [18]. Окрім цього, фактори державного регулювання включають: обмеження продажу нерухомості, контроль за використанням землі, нормативне встановлення ставок орендної плати, будівельні норми, зонування тощо [16, с. 97].

6. Політичні фактори та стабільність. Відсутність єдності влади, наявність революцій, воєн або суперечок в країні не найкращим чином відображається на ринку нерухомості [19].

Окрім цього, політико-психологічними факторами, які впливають на вартість нерухомості є:

- рівень довіри населення до банківської системи, ринку нерухомості та перспектив розвитку міст та сіл;

- рівень оптимізму населення, який залежить від великої кількості факторів, починаючи від перспективи розвитку країни, та закінчуючи станом житлового фонду;

- стан політичної стабільності, а отже, позитивні чи негативні очікування від ринку нерухомості [20, с. 252].

Оцінка всіх макрофакторів, що впливають на вартість будинку, є цінною, проте, варто уважніше розглянути мікрофактори, які впливають на ціни на житло на місцевому рівні.

1. Місцезнаходження. Чим краще розташування, тим вища вартість нерухомості. Нерухомість у центрі великого, густонаселеного міста, безсумнівно, буде вище, ніж подібні об'єкти нерухомості, але, які знаходяться на околиці [21, с. 36].

Близькість до інфраструктури, багатих районів або визначних пам'яток, таких як торгові центри або пляж, – все це фактори, які можуть збільшити вартість об'єкта. На відміну від цього, близькість до аеропорту, або проживання в низько соціально-економічному районі призведе до зниження ціни на нерухомість [17].

2. Розмір будинку та планування. Характеристики та загальний розмір майна також впливатимуть на його вартість. Вартість будинку приблизно оцінюється за ціною за квадратний фут. Жилий простір – є найважливішим для покупців та оцінювачів. Будинок із чотирма спальнями, швидше за все, буде коштувати більше, ніж будинок із двома спальнями в тій же зоні. Так само такі функції, як додаткові кімнати, гаражі, басейни, подвір'я, і це все відіграє важливу роль у вартості нерухомості [13].

2. Вік та стан майна. Як правило, будинки, які є новішими, оцінюються за вищою вартістю. Той факт, що такі важливі елементи будинку, як сантехніка, електрика, дах та побутова техніка є більш новими, і, отже, менш схильні до поломок, може призвести до економії для покупця [14]. Відкладене технічне обслуговування є одним з основних факторів, що може знизити вартість будинку [22].

4. Можливість потенційного ремонту. Якщо у покупця є можливість вдосконалити та персоналізувати майно шляхом ремонту будинку, вартість перепродажу цього майна зростає [16].

Нова ринкова тенденція, що розпочалася кілька років тому, призводить до розвитку сектору заміської житлової нерухомості, представленого переважно котеджними будинками, які можуть конкурувати зі столичним ринком. Хоча ринок житлової нерухомості все ще знаходиться на початку свого розвитку в багатьох регіонах України, висока динаміка зростання цього ринку є характерною для більшості регіонів, які мають потенціал для такого розвитку.

У Києві, Дніпропетровську та Харкові котеджні міста, зазвичай, розвиваються місцевими жителями, а туристичні регіони, такі як Одеса та Карпати, приваблює для іноземців та українців, які приїжджають на відпочинок. Загалом попит майже однаковий у всіх обласних центрах. За різними оцінками, 40-60% українських покупців воліють купувати будинок за містом, а не квартиру [23].

Загалом на рику заміської нерухомості в Україні пропонується декілька типів будинків:

1. невеликі будиночки 40-100 кв. метрів, які коштують 30-50 тисяч доларів. Сюди відносять дачні будинки, котеджі, таунхауси.

2. сучасні будинки площею 100-200 кв. метрів без внутрішнього оздоблення, за ціною 50-100 000 доларів.

3. будинки з ремонтом та якісною обробкою з площею від 150 кв. метрів, які коштують понад 150 000 доларів, та які можна побудувати за власним проектом, переважно виготовлені з цегли.

4. будинки преміум-класу площею від 500 кв. м. Їх вартість можуть сягати декількох мільйонів доларів залежно від місця розташування [24].

Зараз, враховуючи значне підвищення цін, вибір покупців стає більш раціональним. Найбільший попит спостерігається вже не на великі за розміром заміські будинки, а на котеджі площею 100-250 м² [25].

На тлі пандемії та жорсткого локдауну навесні 2020 року, у багатьох українців був страх та обмежений вибір місць для відпочинку. Така ситуація змусила людей звернутися до ринку заміської нерухомості. Через це попит на будинки за містом дуже сильно зріс.

За даними OLX, рейтинг продажу будинків на околиці столиці зріс на 68% у другому кварталі 2020 року, у порівнянні з першим кварталом. За словами експертів DOM.RIA, з квітня по червень 2020 року попит на сегмент будинків площею до 120 м² потроївся. Аналіз Lun.ua показав, що з початком карантину кількість запитів про купівлю будинків досягла рекордного рівня за останні роки.

На рис. 1 можна побачити кількість запитів «купити будинок» в Google за останні 4 роки. Як видно, запит сягнув свого піку протягом весни-літа 2020.

Рисунок 1. Кількість запитів «купити будинок» в Google [26]

Забудовники відреагували на зростаючий попит покупців. На початок 2021 року кількість нових проєктів зросла у 2,5 рази порівняно з початком 2020 року.

Статистика OLX показує, що ціни на нерухомість за містом також зросли в 2020 році. За даними ЛУН Місто, середні ціни на будинки, котеджі та таунхауси за 11 місяців зросли на 15% у Києві та на 35% у Київській області.

Експерти зазначають, що попит на заміську нерухомість буде продовжувати зростати, а це в свою чергу призведе до збільшення кількості та якості запропонованих проєктів [26].

Тепер можна перейти до процесу підготовки даних та побудови математичних моделей. У цьому дослідженні використовувалася набір даних округу Кінг, США, для побудови моделей, які допоможуть передбачити ціни на житло в цьому регіоні з урахуванням набору ознак. Набір даних для аналізу складається з ціни на будинки від продажів у період з травня 2018 року по травень 2019 року. Разом із ціною будинку він складається з інформації про характеристики будинку, які можна побачити у табл.1.

Інформаційною базою дослідження послужили дані з інтернет-ресурсу <https://www.kaggle.com/harlfoxem/housesalesprediction>.

Таблиця 1.
Значення змінних у наборі даних [27]

Показник	Опис показника
Id	Унікальний ідентифікатор кожного проданого будинку
Date	Дата продажу будинку
Price	Ціна кожного проданого будинку
Bedrooms	Кількість спальень
Bathrooms	Кількість ванних кімнат
Sqft_living	Квадратний фут житлової площі будинків
Sqft_lot	Квадратний фут земельної площі будинків (площі ділянки)
Floors	Кількість поверхів
Waterfront	Фіктивна змінна: будинок виходить на набережну чи ні
View	Індекс від 0 до 4 того, наскільки хорошим є вид з будинку
Condition	Індекс від 1 до 5 за станом будинку
Grade	Індекс від 1 до 13 за відповідністю будівельному та дизайнерському рівням
Sqft_above	Квадратний фут внутрішнього простору житла, що знаходиться над рівнем землі
Sqft_basement	Квадратний фут внутрішнього приміщення, яке знаходиться нижче рівня землі
Yr_built	Рік побудови будинку
Yr_renovated	Рік останнього ремонту будинку
Zipcode	В якому районі поштового індексу знаходиться будинок
Lat	Широта
Long	Довгота

Оскільки у наборі даних є змінні зі значеннями індексів, то необхідно окремо зупинитися на інтерпретації кожного з них.

У табл. 2 наведено значення кожного індексу показника view (наскільки хорошим є вид з будинку).

Таблиця 2.
Значення індексів показника view (наскільки хорошим є вид з будинку) [28]

View	Значення
0	Невідомо
1	Задовільний
2	Середній
3	Гарний
4	Чудовий

У табл. 3 наведено значення кожного індексу показника condition (стан будинку).

Таблиця 3.
Значення індексів показника condition (стан будинку) [28]

Condition	Значення	Опис
1	Погане	Потрібно багато ремонтів, що свідчить про серйозне погіршення стану
2	Задовільне	Необхідно багато відкладеного обслуговування.
3	Середнє	Залежно від часу ремонту, нормальний стан для будинку такого віку.
4	Гарне	Стан вище норми для віку будинку
5	Дуже гарне	Відмінне обслуговування та оновлення будинку.

У табл. 4 наведено значення кожного індексу показника grade (відповідність будівельному та дизайнерському рівням).

Таблиця 4.
Значення індексів показника grade (відповідність будівельному та дизайнерському рівням) [29]

Grade	Опис
1–3	Не відповідає мінімальним будівельним стандартам; зазвичай дуже примітивна будівля без зручностей.
4	В основному старе низькоякісне будівництво. Будинок не відповідає нормам.
5	Більш низькі витрати на будівництво і якість обробки. Будинок невеликий, простий дизайн.
6	Використовувалися неякісні матеріали і прості конструкції.
7	Середня оцінка конструкції і дизайну.
8	У будинках такої якості зазвичай використовувалися більш якісні матеріали як для зовнішньої, так і для внутрішньої обробки.
9	Кращий архітектурний дизайн з додатковим дизайном екстер'єру та інтер'єру і гарною якістю.
10	Оздоблення стало краще, більше якісного дизайну видно в планах поверхів і збільшеної площі.
11	Індивідуальний дизайн і більш якісна обробка, з додатковими зручностями, відмінною сантехнікою та іншими більш розкішними опціями.
12	Індивідуальний дизайн і відмінні будівельники. Всі матеріали найвищої якості, є всі зручності.
13	Зазвичай спроектований і побудований за індивідуальним замовленням, наближається до рівня особняка.

Для початку, необхідно зробити аналіз вхідних даних, що є першочерговим етапом для будь-яких досліджень. З цією метою були використано такий програмний засіб як мова програмування R. На рис. 2 можна побачити перших 10 спостережень за всіма відображеними стовпцями.

id	price	date	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	grade	sqft_above	sqft_basement	yr_built	yr_renovated	zipcode	lat	long
1	221900	2018-10-13	3	1.00	1180	5650	1.0	0	0	3	7	1180	0	1955	0	98178	47.5112	-122.257
2	538000	2018-12-09	3	2.25	2570	7242	2.0	0	0	3	7	2170	400	1951	1991	98125	47.7210	-122.319
3	180000	2019-02-25	2	1.00	770	10000	1.0	0	0	3	6	770	0	1933	0	98028	47.7379	-122.233
4	604000	2018-12-09	4	3.00	1960	5000	1.0	0	0	5	7	1050	910	1965	0	98136	47.5208	-122.393
5	510000	2019-02-18	3	2.00	1680	8080	1.0	0	0	3	8	1680	0	1987	0	98074	47.6168	-122.045
6	1225000	2018-05-12	4	4.50	5420	101930	1.0	0	0	3	11	3890	1530	2001	0	98053	47.6561	-122.005
7	257500	2018-06-27	3	2.25	1715	6819	2.0	0	0	3	7	1715	0	1995	0	98003	47.3097	-122.327
8	291650	2019-01-15	3	1.50	1060	9711	1.0	0	0	3	7	1060	0	1963	0	98198	47.4095	-122.315
9	229500	2019-04-15	3	1.00	1780	7470	1.0	0	0	3	7	1050	730	1960	0	98146	47.5123	-122.337
10	323000	2019-03-12	3	2.50	1890	6560	2.0	0	0	3	7	1890	0	2003	0	98038	47.3684	-122.031

Рисунок 2. Перші десять спостережень
Джерело: розроблено автором на основі даних [30]

Оскільки серед змінних є рік будівництва та рік останнього ремонту, то було вирішено власноруч додати дві додаткові змінні: вік будинку та вік останнього ремонту. Ці показники мають допомогти краще зрозуміти дані та виконати моделювання.

Важливим етапом є здійснення кореляційного аналізу, щоб дослідити залежність між величинами. Для цього візуалізуємо матрицю кореляцій у вигляді теплової карти, яку можна побачити на рис. 6.

З огляду на цей рисунок можна зробити висновок, що між деякими незалежними змінними існує сильний зв'язок, що може свідчити про наявність мультиколінеарності. Мультиколінеарність є проблемою, оскільки вона підриває статистичну значимість незалежної змінної.

За даними можна стверджувати, що найвищі позитивні значення кореляції ціни (більше 0.5) є з такими показниками: sqft_living, grade, sqft_above, bathrooms. Це означає, що зі збільшенням цих значень – ціна також збільшується.

У ту саму чергу, найнижче значення кореляції ціни є з показником date, яке дорівнює нулю.

Рисунок 6. Кореляційна матриця

Джерело: розроблено автором на основі даних [30]

Як було видно в кореляційній матриці, кореляція між yr_built та age становить 1. Це може призвести до обманливих результатів моделі, тому краще видалити один з цих показників. Ці показники несуть в собі однакову інформацію, тому неважливо який з них буде видалений. Ми вирішили прибрати такі показники як yr_built, оскільки, як було показано у першому розділі цієї роботи, вік будинку грає важливу роль. Окрім цього, кореляція між yr_renovated та repair_age становить 0.92, це значення є досить високим. Оскільки кореляційний зв'язок між ціною та yr_renovated є кращим (0.13), то видаляємо repair_age.

Перед тим як перейти безпосередньо до моделювання, необхідно провести ряд підготовчих заходів. Важливим кроком є виявлення викидів у ціні. Спостереження, які розглядаються як потенційні викиди, відображаються як пункти на графіку. Виходячи з цього, на рис. 8, існує декілька потенційних викидів.

Рисунок 8. Бокс-плот викидів ціни
Джерело: розроблено автором на основі даних [30]

Для визначення потенційних викидів існує багато методів, але у нашій роботі ми вирішили скористатися методом інтерквартильного діапазону (IQR).

Після того як інтерквартильний метод виявлення викидів був застосований до даних, залишилося 20467 спостережень, тобто 1146 значень були визначені як викиди.

Перед тим як переходити до моделювання, першочергово необхідно нормалізувати дані. У своїй роботі я нормалізувала лише незалежні змінні, оскільки залежна змінна (ціна) буде використовуватися для прогнозування, а тому її значення має бути в початкових одиницях виміру.

Також, поділ даних на навчальні та тестові набори є важливою частиною оцінки моделей аналізу даних. Свою вибірку даних ми поділили таким чином, що 70% значень стали навчальним набором (14326 спостережень), а 30% (6139) залишилися для тестування. Тепер можна переходити до моделювання. Для початку було побудовано множинну лінійну модель. Після запуску необхідного коду в R було отримано результат, який представлений на рис. 9.

```
call:
lm(formula = price ~ ., data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-660734  -49264   -2606   42898  568210

Coefficients:
(Intercept)      301109.7      36708.5      8.203 2.55e-16 ***
date           9723.5       721.3     13.480 < 2e-16 ***
bedrooms1     -51037.2     30133.7    -1.694 0.090346 .
bedrooms2     -58320.0     29286.1    -1.991 0.046457 *
bedrooms3     -61337.1     29257.3    -2.096 0.036058 *
bedrooms4     -61168.8     29301.7    -2.088 0.036856 *
bedrooms5     -76415.1     29443.0    -2.595 0.009459 **
bedrooms6     -94205.0     30204.4    -3.119 0.001819 ***
bedrooms7    -163566.8     34705.1    -4.713 2.46e-06 ***
bedrooms8    -127903.9     42341.4    -3.021 0.002526 **
bedrooms9     -67938.2     67575.6    -1.005 0.314738
bedrooms10   -174140.1     67416.8    -2.583 0.009803 **
bathrooms      8226.1      1269.4     6.481 9.44e-11 ***
sqft_living    55052.4      2794.4    19.701 < 2e-16 ***
sqft_lot      11632.9       756.9     15.370 < 2e-16 ***
floors1.5       2944.2      2881.1     1.022 0.306838
floors2       -16220.0     2490.4    -6.513 7.61e-11 ***
floors2.5     -15675.5     10053.7    -1.559 0.118977
floors3       -57422.3     5408.0   -10.618 < 2e-16 ***
floors3.5     -12639.5     43676.2    -0.289 0.772287
waterfront1   149340.3     15494.9     9.638 < 2e-16 ***
view1         63978.8     6343.1    10.086 < 2e-16 ***
view2         66070.4     3812.5    17.330 < 2e-16 ***
view3         93559.6     5729.4    16.330 < 2e-16 ***
view4        166212.7     9653.3    17.218 < 2e-16 ***
condition2     81815.1     22253.0     3.677 0.000237 ***
condition3     99291.4     20990.9     4.730 2.26e-06 ***
condition4    118395.7     20991.6     5.640 1.73e-08 ***
condition5    151883.0     21101.3     7.198 6.43e-13 ***
grade         49811.9     1253.2    39.748 < 2e-16 ***
sqft_above    41422.0     2925.8    14.158 < 2e-16 ***
sqft_basement1 6990.0      2932.4     2.384 0.017150 *
yr_renovated1  35618.0     3999.5     8.906 < 2e-16 ***
age          11970.7     1357.2     8.820 < 2e-16 ***
zipcode98002   1267.1      9461.1     0.134 0.893458
zipcode98003   -9910.1     8447.2    -1.173 0.240740
zipcode98004   466799.8     16101.6    28.991 < 2e-16 ***
zipcode98005   289926.5     16442.8    17.632 < 2e-16 ***
zipcode98006   244172.9     13651.1    17.887 < 2e-16 ***
zipcode98007   210697.5     17047.7    12.359 < 2e-16 ***
zipcode98008   191527.7     16220.4    11.808 < 2e-16 ***
zipcode98010   92210.6     13946.1     6.612 3.93e-11 ***
zipcode98011   77572.7     20865.1     3.718 0.000202 ***
zipcode98014   62562.0     23410.3     2.672 0.007539 **
zipcode98019   37915.6     22732.6     1.668 0.095358 .
zipcode98022   20213.5     12733.5     1.587 0.112438
zipcode98023  -26477.8     7658.2    -3.457 0.000547 ***

zipcode98038    44064.0     10636.3     4.143 3.45e-05 ***
zipcode98039    573387.4     45145.8    12.701 < 2e-16 ***
zipcode98040    371421.2     14475.4    25.659 < 2e-16 ***
zipcode98042     8372.4      8894.8     0.941 0.346583
zipcode98045    102870.0     19944.4     5.158 2.53e-07 ***
zipcode98052    201596.6     17797.0    11.328 < 2e-16 ***
zipcode98053    187672.0     19281.8     9.733 < 2e-16 ***
zipcode98055    16374.1     10543.2     1.553 0.120433
zipcode98056    69118.3     11634.1     5.941 2.90e-09 ***
zipcode98058    18547.2     10171.8     1.823 0.068266 .
zipcode98059    78959.2     11426.6     6.910 5.05e-12 ***
zipcode98065    114804.2     18276.7     6.281 3.45e-10 ***
zipcode98070    48640.7     13694.8     3.552 0.000384 ***
zipcode98072    103647.8     20816.9     4.979 6.47e-07 ***
zipcode98074    179675.4     16996.8    10.571 < 2e-16 ***
zipcode98075    207007.1     16457.1    12.579 < 2e-16 ***
zipcode98077    116900.1     21778.2     5.368 8.10e-08 ***
zipcode98092     -8143.6      8315.5    -0.979 0.327437
zipcode98102    345649.6     18387.8    18.798 < 2e-16 ***
zipcode98103    247431.1     16731.4    14.788 < 2e-16 ***
zipcode98105    312715.0     17559.4    17.809 < 2e-16 ***
zipcode98106    65117.0     12397.6     5.252 1.52e-07 ***
zipcode98107    244179.5     17232.2    14.170 < 2e-16 ***
zipcode98108    66666.5     13642.3     4.887 1.04e-06 ***
zipcode98109    344057.6     18065.4    19.045 < 2e-16 ***
zipcode98112    366535.4     16256.0    22.548 < 2e-16 ***
zipcode98115    240467.8     17007.6    14.139 < 2e-16 ***
zipcode98116    231279.9     13894.9    16.645 < 2e-16 ***
zipcode98117    231593.7     17167.7    13.490 < 2e-16 ***
zipcode98118    117071.1     12052.2     9.714 < 2e-16 ***
zipcode98119    331540.8     17140.3    19.343 < 2e-16 ***
zipcode98122    246547.1     14999.5    16.437 < 2e-16 ***
zipcode98125    119165.0     18359.2     6.491 8.82e-11 ***
zipcode98126    133973.9     12634.7    10.604 < 2e-16 ***
zipcode98133    66959.3     18891.6     3.544 0.000395 ***
zipcode98136    203410.5     12938.5    15.721 < 2e-16 ***
zipcode98144    186513.6     13899.2    13.419 < 2e-16 ***
zipcode98146    58420.0     11573.2     5.048 4.52e-07 ***
zipcode98148    24343.0     16190.3     1.504 0.132719
zipcode98155    53567.2     19738.8     2.714 0.006660 ***
zipcode98166    60419.5     10677.8     5.658 1.56e-08 ***
zipcode98168    12807.9     11175.8     1.146 0.251798
zipcode98177    132414.7     19759.0     6.701 2.14e-11 ***
zipcode98178    16568.5     11542.2     1.435 0.151176
zipcode98188    5447.1      11484.7     0.474 0.635301
zipcode98198    -2425.0      8974.0    -0.270 0.786993
zipcode98199    287579.7     16523.5    17.404 < 2e-16 ***
lat            25402.6      5823.2     4.362 1.30e-05 ***
long           -7232.9      4456.7    -1.623 0.104631

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 85490 on 14221 degrees of freedom
Multiple R-squared:  0.8328,    Adjusted R-squared:  0.8316 
F-statistic: 681.3 on 104 and 14221 DF,  p-value: < 2.2e-16
```

Рисунок 9. Результат множинної лінійної регресії
Джерело: розроблено автором на основі даних [30]

За даним рисунком можна зробити певні висновки.

1. *p-значення* – дуже близькі до нуля, що вказує на те, що ми можемо відкинути нульову гіпотезу, що дозволяє зробити висновок про наявність залежності між ціною та незалежними змінними.

2. *Множинний R-квадрат, Відрегульований R-квадрат* – отриманий R2 дорівнює 0,8328, або приблизно 83% дисперсії, знайденої у залежній змінній, можна пояснити змінною незалежних змінних. Скоригований R2 дорівнює 0,8316, або приблизно 83% дисперсії, знайденої у залежній змінній, можна пояснити змінною незалежних змінних.

3. *F-статистика* – становить 681,3, що відносно більше, ніж 1, враховуючи розмір наших даних. Оскільки $p\text{-value} < 2.2e-16$, то можна вважати, що ми можемо відкинути нульову гіпотезу, що дозволяє зробити висновок про наявність залежності між ціною та незалежними змінними.

Наступним важливим кроком є перевірка ефективності на тестовому наборі. Було визначено, що R-квадрат становить 0,8371, або приблизно 84% дисперсії, знайденої у залежній змінній, можна пояснити змінною незалежних змінних.

Далі було виконано розрахунок точності для моделі прогнозу. Було отримано такі результати:

- RMSE = 85054,44
- MAE = 61814.5
- MAPE = 0,1456575
- SMAPE = 0,143053
- MASE = 0,2652008

Зазвичай, найкращими оцінками для перевірки ефективності моделі є RMSE та MAPE. RMSE говорить про те, що в середньому, прогнозовані значення відхиляються від експериментальних на 85054,44 доларів. Показник MAPE, говорить про те, що, в середньому, прогнозовані значення відхиляються від експериментальних на 14,5%, що є не поганим показником.

Наступною побудованою моделлю стала регуляризація. У своїй роботі я використала такі методи як: рідж регресія, ласо регресія та еластична сітка.

Гіперпараметр λ – параметр налаштування, який визначає, наскільки ми хочемо покарати гнучкість нашої моделі. На рис. 10 можна побачити, як в залежності від значення показника λ штрафуються коефіцієнти моделі. Можна зробити висновок, що коли $\lambda = 0$, то ніякого ефекту немає, і наша цільова функція дорівнює нормальній цільовій функції регресії, а чим більший цей параметр, тим більший штраф накладається на коефіцієнти.

Також варто зауважити, що рідж регресія не штрафує змінні рівно до нуля, тому всі вони залишаються в моделі. Штраф за ласо фактично зсовує коефіцієнти до нуля. Таким чином, модель ласо та еластична сітка не тільки вдосконалює модель за допомогою регуляризації, але також здійснює автоматизований вибір змінних.

Рисунок 10. Штрафування коефіцієнтів
Джерело: розроблено автором на основі даних [30]

Для того, щоб ідентифікувати оптимальну лямбду для моделі, нам потрібно виконати перехресну перевірку. Перша та друга вертикальні пунктирні лінії на рис. 11 представляють значення λ з мінімальним значенням MSE та найбільшим значенням λ в межах однієї стандартної помилки мінімального значення MSE.

З огляду на ці значення можна зробити висновок, що краще використовувати лямбду однієї стандартної похибки, оскільки лямбда мінімального MSE не має майже жодного впливу на змінні.


```
hyper_grid.h2o <- list(
  ntrees      = seq(50, 500, by = 50),
  mtries      = seq(5, 15, by = 10),
  max_depth   = seq(20, 40, by = 5),
  min_rows    = seq(1, 5, by = 1),
  nbins       = seq(10, 30, by = 5),
  sample_rate = c(.55, .632, .75)
)
```

Рисунок 15. Параметри моделі для пошуку по сітці

Джерело: розроблено автором на основі даних [30]

У результаті, було обрано найкращу комбінацію гіперпараметрів, за якими значення MSE є найменшим (рис. 16).

```
best_model <- h2o.randomForest(
  x = x,
  y = y,
  training_frame = train.h2o,
  ntrees      = 450,
  mtries      = 15,
  max_depth   = 20,
  min_rows    = 1,
  nbins       = 30,
  sample_rate = .55)
)
```

Рисунок 16. Результат пошуку по сітці

Джерело: розроблено автором на основі даних [30]

Після тестування випадкового лісу, природно запитати, які змінні мають найбільшу прогнозовану силу. Змінні, що мають велике значення, є рушіями результату, і їх значення мають значний вплив на значення результату. Навпаки, змінні з низькою важливістю можуть бути опущені з моделі, що робить її простішою та швидшою для підгонки та прогнозування.

На рис. 17 можна побачити відносну важливість змінних у моделі. Цей показник допомагає дослідникам даних оцінити, які змінні мають найбільший вплив. Найбільш важливими змінними є sqft_living (житлова площа) та lat (широта), а найменш важливими – waterfront (вихід на набережну), yr_renovated (рік останнього ремонту), sqft_basement (площа підвалу).

Рисунок 17. Відносна важливість змінних

Джерело: розроблено автором на основі даних [30]

Після оцінки продуктивності на тестовому наборі даних, були отримані наступні показники ефективності:

- R квадрат = 0,864
- RMSE = 76899,5
- MAPE = 0,11122
- MAE = 53012,82

RMSE говорить про те, що в середньому, прогнозовані значення відхиляються від експериментальних на 76899,5 доларів. Показник MAPE говорить про те, що, в середньому, прогнозовані значення відхиляються від експериментальних на 14%.

На останок, була побудована модель з використанням алгоритму машинного навчання XGBoost. Можна налаштувати модель, змінюючи значення гіперпараметрів. Пошук по сітці допоможе встановити ті значення параметрів, за якими оцінка RMSE буде найкращою (рис. 18.)

```
hyper_grid <- expand.grid(
  eta = c(.01, .05, .1, .3),
  gama = c(0,.25,1),
  lambda = c(0,1,10),
  max_depth = c(1, 3, 5, 7),
  min_child_weight = c(1, 3, 5, 7),
  subsample = c(.65, .8, 1),
  colsample_bytree = c(.75, .9, 1),
  optimal_trees = 0,
  min_RMSE = 0
)
```

Рисунок 18. Параметри моделі для пошуку по сітці

Джерело: розроблено автором на основі даних [30]

Після пошуку, були визначені параметри, які можна побачити на рис. 19, які зменшили значення RMSE до 70812,76 доларів. Після цього можна запустити модель на тренувальних даних з використанням 2018 дерев та обраних параметрів.

```
params <- list(
  eta = 0.01,
  gama = 0.25,
  lambda = 5,
  max_depth = 7,
  min_child_weight = 1,
  subsample = 0.65,
  colsample_bytree = 0.75
)
# train final model

xgb.fit.final <- xgboost(
  params = params,
  data = x_train,
  label = y_train,
  nrounds = 2018,
  objective = "reg:linear",
  verbose = 0
)
```

Рисунок 19. Результат пошуку по сітці

Джерело: розроблено автором на основі даних [30]

На рис. 20 можна побачити відносну важливість змінних у моделі. Найбільш важливими змінними є lat (широта), sqft_living (житлова площа) та grade (індекс за відповідністю будівельному та дизайнерському рівням) а найменш важливими – waterfront (вихід на набережну), yr_renovated (рік останнього ремонту), floors (кількість поверхів).

Рисунок 20. Відносна важливість змінних

Джерело: розроблено автором на основі даних [30]

Порівняння із відотною важливістю змінних у моделі випадкового лісу показує, що важливість змінних різниться. Це є логічним, оскільки метод пристосування до даних для кожної моделі різний. Однак,

незалежно від моделі, змінні, що стосуються широти та площі будинку, мабуть, мають важливий вплив на ціну. Цей факт свідчить про те, що ці дві змінні є основними складовими ціни на житло.

Останнім кроком є перевірка ефективності на тестовому наборі даних. Після розрахунків можна обчислити наступні метрики ефективності:

- R квадрат = 0,893
- RMSE = 67669,15
- MAE = 47063,49
- MAPE = 0,1034903
- SMAPE = 0,1048065
- MASE = 0,2188252

RMSE говорить про те, що в середньому, прогнозовані значення відхиляються від експериментальних на 67669,15 доларів. Показник MAPE говорить про те, що, в середньому, прогнозовані значення відхиляються від експериментальних на 10%.

У результаті проведених досліджень, з рис. 21 можна зробити висновок, що:

1. за повної лінійної регресійної моделі приблизно 83,4% дисперсії, знайденої у залежній змінній, можна пояснити змінною незалежних змінних; в середньому, прогнозовані значення відхиляються від експериментальних на 85054,44 доларів, або на 15%;
2. за моделлю еластичної сітки приблизно 69,1% дисперсії, знайденої у залежній змінній, можна пояснити змінною незалежних змінних; в середньому, прогнозовані значення відхиляються від експериментальних на 115863,40 доларів, або на 20%;
3. за алгоритмом випадкового лісу приблизно 86,4% дисперсії, знайденої у залежній змінній, можна пояснити змінною незалежних змінних; в середньому, прогнозовані значення відхиляються від експериментальних на 76961,49 доларів, або на 14%;
4. за алгоритмом XGBoost приблизно 89,3% дисперсії, знайденої у залежній змінній, можна пояснити змінною незалежних змінних; в середньому, прогнозовані значення відхиляються від експериментальних на 67669,15 доларів, або на 10%.

Отже, для прогнозування ціни на цьому наборі даних найкраще буде використання алгоритму машинного навчання XGBoost, оскільки відхилення прогнозованих значень від реальних становить лише 10%, що є гарним показником.

Переглянувши статті українських дослідників на тему моделювання цін на нерухомість, ми зробили висновок, що застосовані методи дуже різняться між собою. Багато українських дослідників використовують саме лінійну регресійну модель. На нашу думку, такий підхід є застарілим, оскільки зараз існує дуже багато нових та потужних алгоритмів машинного навчання, які здатні набагато краще виконувати задачі прогнозування.

Лінійна регресія є чудовим інструментом для аналізу взаємозв'язків між змінними, але вона не рекомендується для більшості практичних застосувань, оскільки надто спрощує реальні проблеми, приймаючи лінійну залежність між змінними. У ту саму чергу, такий алгоритм машинного навчання як XGBoost охоплює як лінійні, так і нелінійні відносини. Окрім цього, він працює на основі ансамблевого навчання, що дає змогу об'єднати сильні сторони різних моделей та дати більш гарний результат.

Висновки.

Було проведено моделювання та прогнозування ціни на будинки в окрузі Кінг, США з використанням 4 методів машинного навчання – лінійна регресія, регресія регуляризації, випадковий ліс та XGBoost.

Після перевірки ефективності моделей на тестовому наборі даних та порівняння результатів між собою, було визначено, що за показниками RMSE та MAPE найкращу точність дає алгоритм XGBoost, а найгіршу – регуляризація.

Було визначено, що багато науковців використовують лінійну регресію для моделювання та прогнозування ціни на нерухомість. Для того, щоб продемонструвати, що сучасні алгоритми краще розв'язують цю задачу, було вирішено порівняти результати прогнозування алгоритму XGBoost та множинної лінійної регресії на основі випадково обраного будинку з вибірки.

Результати показали, що модель XGBoost краще описує дані і ймовірність спрогнозувати справжню ціну є вищою.

Моделювання ціни на житлову нерухомість може допомогти зменшити невизначеність, більш точно контролювати необхідні затрати, встановлювати правильну ціну на житло з урахуванням великої кількості факторів. Запропоновані розробки можуть послужити методологічною базою для прийняття подібних рішень різними підприємствами, допомагаючи їм зробити більш виважені висновки щодо подальших дій.

Список літератури.

1. Асаул А.М., Павлов В.І., Пилипенко І.І., Павліха Н.В., Кривов'язюк І.В. Економіка нерухомості : навч. посіб., 2-ге вид.– К.: Кондор, 2006. 336 с.
2. Стерник Г. М., Стерник С. Г. Методология моделирования и прогнозирования жилищного рынка: монография. Москва: РГ-Пресс, 2018. 588 с.
3. Манцевич Ю. М. Житло: проблеми та перспективи. Київ : Профі, 2004. 360 с.
4. Григорків В. С., Ярошенко О. І., Філіпчук Н. В. Нейронні мережі та їх використання для прогнозування тенденцій ринку нерухомості. *Науковий вісник НЛТУ України*. 2012. Вип. 22.5. С. 324-330.

5. Воронін В. О., Мамчин М. М., Лянце Е. В. Прогнозне моделювання тенденцій розвитку ринку нерухомості. *Вісник Національного університету "Львівська політехніка"*. 2012. № 735 : Логістика. С. 38-46.
6. Максишко Н. К., Шаповалова В. О. Нечітка модель ідентифікації фаз на ринку нерухомості. *Нейро-нечіткі технології моделювання в економіці*. 2014. № 3. С. 94-119.
7. Real Estate - Industry Overview, Types of Real Estate, Careers. *Corporate Finance Institute*. URL: <https://corporatefinanceinstitute.com/resources/careers/jobs/real-estate/> (дата звернення: 09.06.2021 р.).
8. What is the Housing Market? - Definition | Meaning | Example. *My Accounting Course*. URL: <https://www.myaccountingcourse.com/accounting-dictionary/housing-market> (дата звернення: 09.06.2021 р.).
9. Галаган Д. В. Ринок нерухомості як різновид інвестиційного ринку. *Теорія інвестицій*. 2010. №6. С.19-21.
10. Youds C. Supply and demand: the housing market. *The Sloman Economics News Site*. URL: <https://pearsonblog.campaignserver.co.uk/supply-and-demand-the-housing-market/> (дата звернення: 09.06.2021 р.).
11. Russell Cooper, Andrew John. *Macroeconomics: Theory through Applications*. 2011. 524 p.
12. Chandler T. 12 Factors That Affect Property Prices. URL: <https://buyersagent-sydney.com.au/12-factors-that-affect-property-prices/> (дата звернення: 09.06.2021 р.).
13. Whitten R. What factors influence a property's market value? *Finder*. URL: <https://www.finder.com.au/what-influences-a-property-s-value> (дата звернення: 09.06.2021 р.).
14. Gomez J. 8 critical factors that influence a home's value. *Opendoor*. URL: <https://www.opendoor.com/w/blog/factors-that-influence-home-value> (дата звернення: 09.06.2021 р.).
15. Кучеренко В. Р., Засць М. А., Захарченко О. В., Сментина Н. В., Улибіна В. О. Оцінка та управління нерухомістю: навчальний посібник. Одеса: Видавництво ТОВ «Лерадрук», 2013. 272 с.
16. 13 surprising factors that affect property value. *Upside Realty*. URL: <https://upside.com.au/articles/selling-your-property/selling-guide/9-surprising-factors-affect-home-value> (дата звернення: 09.06.2021 р.).
17. 10 factors that affect property value. *The home of smart money*. URL: <https://www.loans.com.au/blog/10-factors-that-affect-property-value> (дата звернення: 09.06.2021 р.).
18. Nguyen J. 4 Key Factors That Drive the Real Estate Market. *Investopedia*. URL: <https://www.investopedia.com/articles/mortgages-real-estate/11/factors-affecting-real-estate-market.asp> (дата звернення: 09.06.2021 р.).
19. Які фактори впливають на формування ціни. URL: <https://vn.com.ua/ua/news/kakie-factory-vlijajut-na-formirovanie-tseny-za-1-m2-v-zhk-i-kg> (дата звернення: 09.06.2021 р.).
20. Кузьміч О.Й., Іванова В.О. Аналіз деяких факторів, які впливають на вартість нерухомості. *Містобудування та територіальне планування*. УДК 528.48. С.251-257.
21. Pham Duc Trung, Nguyen Gia Trung Quan. Factors affecting the price of the real estate: A case of Ho Chi Minck city. *British Journal of Marketing Studies (BJMS)*. 2019. Vol. 7, Issue 6, P.35-45.
22. Bartsch C. 10 Factors Large and Small that Influence Real Estate Property Values. *HomeLight Blog*. URL: <https://www.homelight.com/blog/real-estate-property-value/> (дата звернення: 09.06.2021 р.).
23. Ринок котеждної нерухомості 2020: пропозиція, структура, прогноз. *NAI Ukraine*. URL: <https://naiukraine.com/uk/news/rynok-kotedzhnoyi-neruhomosti-2020-propozytsiya-struktura-prognoz/> (дата звернення: 09.06.2021 р.).
24. Аналіз ринку замиської нерухомості Києва. URL: <https://pro-consulting.ua/ua/issledovanie-rynka/analiz-rynka-zagorodnoj-nedvizhimosti-kieva-2018-god> (дата звернення: 09.06.2021 р.).
25. «МАЙГЕР» Девелопмент і Консалтинг в Україні. Тенденції розвитку ринку котеждної нерухомості. *МАЙГЕР*. URL: <http://www.mayger.ua/uk/analitika/tendentsiyi-rozvitku-rynku-kotedzhnoyi-neruhomosti/> (дата звернення: 09.06.2021 р.).
26. Колесніченко О. Будинки, котежді, таунхауси: чому попит на замиську нерухомість не падає і ціни зростають. *Економічна правда*. URL: <https://www.epravda.com.ua/publications/2021/03/3/671561/> (дата звернення: 09.06.2021 р.).
27. Shivhare P. Predicting King County House Prices. URL: <https://www.slideshare.net/PawanShivhare1/predicting-king-county-house-prices> (дата звернення: 09.06.2021 р.).
28. Parcel Record Assessor extract table. URL: https://www5.kingcounty.gov/sdc/FGDCDocs/PARCEL_EXTR_faqs.htm (дата звернення: 09.06.2021 р.).
29. Glossary for Improved Sales. King County, Washington - King County. URL: <https://kingcounty.gov/Assessor/Reports/ArchivedAreaReports/~media/Assessor/AreaReports/AreaReportGlossary.ashx> (дата звернення: 09.06.2021 р.).
30. House Sales in King County, USA. Kaggle: Your Machine Learning and Data Science Community. URL: <https://www.kaggle.com/harlfoxem/housesalesprediction> (дата звернення: 09.06.2021 р.).
31. Why overpricing your home is a bad idea. *Duet Property*. URL: <https://www.duetproperty.com.au/2020/02/why-overpricing-your-home-is-a-bad-idea/#:~:text=The%20first%20few%20days%20after,buyers,%20significantly%20decreasing%20buyer%20demand> (дата звернення: 09.06.2021 р.).

References.

1. Asaul, A.M., Pavlov, V.I., Pylypenko, I.I., Pavlikha, N.V. and Kryvov'yaziuk, I.V. (2006), *Ekonomika nerukhomosti* [Real estate economics], 2nd ed, Kondor, Kyiv, Ukraine.
2. Sternik, G. M. and Sternik, S. G. (2018), *Metodologija modelirovaniya i prognozirovaniya zhilishnogo rynku* [Housing market modeling and forecasting methodology], RG-Press, Moscow, Russia.
3. Mantsevykh, Yu. M. (2004), *Zhytlo: problemy ta perspektyvy* [Housing: problems and prospects], Profi, Kyiv, Ukraine.
4. Hryhorkiv, V.S., Yaroshenko, O.I. and Filipchuk, N.V. (2012), "Neural networks and their application for real estate market trends forecasting", *Naukovyi visnyk NLTU Ukrainy*, vol. 22.5, pp. 324-330.
5. Voronin, V. O., Mamchyn, M. M. and Liantse, E. V. (2012), "Predictive modeling of real estate market trends", *Visnyk Natsionalnoho universytetu "Lvivska politehnika"*, vol. 735, pp. 38-46.
6. Maksyshko, N. K. and Shapovalova, V. O. (2014), "The fuzzy model of phase identification on the real estate market", *Neiro-nechitki tekhnolohii modeliuвання v ekonomitsi*, vol. 3, pp. 94-119.
7. *Corporate Finance Institute* (2021), "Real Estate - Industry Overview, Types of Real Estate, Careers", available at: <https://corporatefinanceinstitute.com/resources/careers/jobs/real-estate/> (Accessed 09 June 2021).
8. *My Accounting Course* (2021), "What is the Housing Market? - Definition | Meaning | Example", available at: <https://www.myaccountingcourse.com/accounting-dictionary/housing-market> (Accessed 09 June 2021).
9. Halahan, D. V. (2010), "The real estate market as a kind of investment market", *Teoriia investytsii*, vol. 6, pp. 19-21.
10. Youds, C. (2020), "Supply and demand: the housing market", *The Sloman Economics News Site*, available at: <https://pearsonblog.campaignserver.co.uk/supply-and-demand-the-housing-market/> (Accessed 09 June 2021).
11. Cooper, R. and John, A. (2011), *Macroeconomics: Theory through Applications*, Flat World, Boston, MA, United States.
12. Chandler, T. (2015), "12 Factors That Affect Property Prices", available at: <https://buyersagent-sydney.com.au/12-factors-that-affect-property-prices/> (Accessed 09 June 2021).
13. Whitten, R. (2021), "What factors influence a property's market value?", *Finder*, available at: <https://www.finder.com.au/what-influences-a-property-value> (Accessed 09 June 2021).
14. Gomez, J. (2019), "8 critical factors that influence a home's value", *Opendoor*, available at: <https://www.opendoor.com/w/blog/factors-that-influence-home-value> (Accessed 09 June 2021).
15. Kucherenko, V. R., Zaiets, M. A., Zakharchenko, O. V., Smentyna, N. V. and Ulybina, V. O. (2013), *Otsinka ta upravlinnia nerukhomosti* [Real estate appraisal and management], TOV «Leradruk», Odessa, Ukraine.
16. *Upside Realty* (2021), "13 surprising factors that affect property value", available at: <https://upside.com.au/articles/selling-your-property/selling-guide/9-surprising-factors-affect-home-value> (Accessed 09 June 2021).
17. *The home of smart money* (2021), "10 factors that affect property value", available at: <https://www.loans.com.au/blog/10-factors-that-affect-property-value> (Accessed 09 June 2021).
18. Nguyen, J. (2021), "4 Key Factors That Drive the Real Estate Market", *Investopedia*, available at: <https://www.investopedia.com/articles/mortgages-real-estate/11/factors-affecting-real-estate-market.asp> (Accessed 09 June 2021).
19. *vn.com.ua*: website (2019), "What factors affect price formation", available at: <https://vn.com.ua/ua/news/kakie-factory-vlijajut-na-formirovanie-tseny-za-1-m2-v-zhk-i-kg> (Accessed 09 June 2021).
20. Kuzmich, O.Y. and Ivanova, V.O., "Analysis of some factors that affect the value of real estate", *Mistobuduvannia ta terytorialne planuvannia*, pp. 251-257.
21. Pham Duc Trung and Nguyen Gia Trung Quan (2019), "Factors affecting the price of the real estate: A case of Ho Chi Minck city", *British Journal of Marketing Studies (BJMS)*, Vol. 7, Issue 6, pp.35-45.
22. Bartsch, C. (2019), "10 Factors Large and Small that Influence Real Estate Property Values", *HomeLight Blog*, available at: <https://www.homelight.com/blog/real-estate-property-value/> (Accessed 09 June 2021).
23. *NAI Ukraine*: website (2020), "Cottage real estate market 2020: proposal, structure, forecast", available at: <https://naiukraine.com/uk/news/rynok-kotedzhnoyi-nerukhomosti-2020-propozytsiya-struktura-prognoz/> (Accessed 09 June 2021).
24. *Pro Consulting*: website (2019), "Analysis of the suburban real estate market in Kyiv", available at: <https://pro-consulting.ua/ua/issledovanie-rynka/analiz-rynka-zagorodnoj-nedvizhimosti-kieva-2018-god> (Accessed 09 June 2021).
25. «MAIHER» *Development i Konsaltnh v Ukraini*: website (2021), "Trends in the development of the cottage real estate market", available at: <http://www.mayger.ua/uk/analitika/tendentsiyi-rozvitku-rinku-kotedzhnoyi-nerukhomosti/> (Accessed 09 June 2021).
26. Kolesnichenko, O. (2021), "Houses, cottages, townhouses: why the demand for suburban real estate is not falling and prices are rising", *Ekonomichna pravda*, available at: <https://www.epravda.com.ua/publications/2021/03/671561/> (Accessed 09 June 2021).
27. Shivhare, P. (2017), "Predicting King County House Prices", available at: <https://www.slideshare.net/PawanShivhare1/predicting-king-county-house-prices> (Accessed 09 June 2021).
28. King County (2021), "Parcel Record Assessor extract table", available at: https://www5.kingcounty.gov/sdc/FGDCDocs/PARCEL_EXTR_faqs.htm (Accessed 09 June 2021).

29. King County (2021), “Glossary for Improved Sales”, available at: https://kingcounty.gov/Assessor/Reports/ArchivedAreaReports/~/_media/Assessor/AreaReports/AreaReportGlossary.aspx (Accessed 09 June 2021).

30. Kaggle (2021), “Your Machine Learning and Data Science Community”, available at: <https://www.kaggle.com/harlfoxem/housesalesprediction> (Accessed 09 June 2021).

31. *Duet Property* (2021), “Why overpricing your home is a bad idea”, available at: <https://www.duetproperty.com.au/2020/02/why-overpricing-your-home-is-a-bad-idea/#:~:text=The%20first%20few%20days%20after,buyers,%20significantly%20decreasing%20buyer%20demand> (Accessed 09 June 2021).

Стаття надійшла до редакції 10.01.2022 р.