

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Навчально-науковий інститут інноватики, природокористування
та інфраструктури
Кафедра економічної експертизи та землевпорядкування

ШАХОВАЛ Ганна Дмитрівна
Дослідження якості прогнозу цін на нерухомість зробленого
сучасними технологіями з використанням штучного інтелекту /
Study of the quality of the forecast of real estate prices made by modern
technologies using artificial intelligence

спеціальність: 193 – Геодезія та землеустрій
освітньо-професійна програма – Експертна оцінка землі та нерухомого майна

Кваліфікаційна робота

Виконала студентка групи ГЕОЗ-41
Г. Д. Шаховал (Ковальчук)

Науковий керівник:
к.т.н., Ю. О. Лук'янченко

Кваліфікаційну роботу допущено
до захисту:
" ____ " _____ 20 ____ р.
Завідувач кафедри
_____ **Б. О. Язлюк**

ТЕРНОПІЛЬ – 2026

ЗМІСТ

ВСТУП	3
РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ ПРОГНОЗУВАННЯ ЦІН НА ЖИТЛОВУ НЕРУХОМОСТЬ	5
1.1. Поняття та особливості ринку нерухомості	5
1.2. Фактори, що впливають на формування вартості нерухомості	7
1.3. Методи прогнозування вартості нерухомості	10
1.4. Характеристика ринку нерухомості Шотландії	13
Висновки до розділу 1	16
РОЗДІЛ 2. АНАЛІЗ ДАНИХ РИНКУ НЕРУХОМОСТІ ТА МЕТОДИКА ДОСЛІДЖЕННЯ	17
2.1. Застосування технологій штучного інтелекту у прогнозуванні економічних процесів	17
2.2. Аналіз часових рядів для прогнозування вартості нерухомості	20
2.3. Методи моделювання та прогнозування часових рядів	22
2.4. Опис моделей ARIMA та SARIMA з використанням python, MAE, RMSE, MAPE	25
Висновки до розділу 2	30
РОЗДІЛ 3. ПРОГНОЗУВАННЯ ЦІН НА НЕРУХОМОСТЬ ТА ОЦІНКА ЯКОСТІ ПРОГНОЗУ	31
3.1. Опис та підготовка вихідних даних (1968–2025)	31
3.2. Програмна реалізація моделей у середовищі Python	34
3.3. Побудова моделі прогнозування цін на нерухомість	36
3.4. Реалізація прогнозу за допомогою моделі ARIMA	38
3.5. Реалізація прогнозу за допомогою моделі SARIMA	44
3.6. Порівняння результатів моделей прогнозування	50
3.7. Практичне застосування отриманих результатів у сфері експертної оцінки нерухомості	55
Висновки до розділу 3	57
ВИСНОВКИ	58
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	60
ДОДАТКИ	65

ВСТУП

Ринок житлової нерухомості є однією з ключових складових національної економіки, що відображає загальний рівень соціально-економічного розвитку країни, інвестиційну активність та добробут населення. Динаміка цін на житлову нерухомість має суттєвий вплив як на окремих інвесторів і домогосподарства, так і на державну політику у сфері будівництва, кредитування та регулювання ринку. У зв'язку з цим особливої актуальності набуває проблема прогнозування вартості нерухомості, що дозволяє приймати обґрунтовані управлінські рішення та знижувати ризики в умовах невизначеності.

Сучасний розвиток інформаційних технологій і методів обробки даних відкриває нові можливості для аналізу економічних процесів. Зокрема, застосування методів аналізу часових рядів та моделей штучного інтелекту дає змогу підвищити точність прогнозів і врахувати складні закономірності зміни цін на ринку нерухомості. Особливе місце серед таких методів займають моделі ARIMA та SARIMA, які широко використовуються для моделювання та прогнозування економічних показників.

Актуальність теми дослідження зумовлена необхідністю вдосконалення підходів до прогнозування цін на житлову нерухомість, зокрема із використанням сучасних математичних моделей та програмних засобів. Вибір ринку нерухомості Шотландії як об'єкта дослідження обумовлений наявністю відкритих статистичних даних, стабільністю економічного середовища та можливістю побудови довгострокових часових рядів.

Метою роботи є розробка та реалізація моделей прогнозування цін на житлову нерухомість на основі аналізу часових рядів із використанням моделей ARIMA та SARIMA, а також оцінка їхньої ефективності.

Для досягнення поставленої мети необхідно вирішити такі **завдання**:

- дослідити теоретичні основи функціонування ринку нерухомості;

- визначити основні фактори, що впливають на формування вартості житла;
- розглянути сучасні методи прогнозування економічних показників;
- здійснити аналіз часових рядів цін на нерухомість;
- реалізувати моделі ARIMA та SARIMA у середовищі Python;
- оцінити точність прогнозів за допомогою показників MAE, RMSE, MAPE;
- провести порівняльний аналіз результатів прогнозування;
- визначити можливості практичного застосування отриманих результатів.

Об'єктом дослідження є ринок житлової нерухомості.

Предметом дослідження є методи та моделі прогнозування цін на житлову нерухомість.

Методи дослідження включають аналіз і узагальнення наукових джерел, методи економічного аналізу, статистичні методи обробки даних, аналіз часових рядів, а також економіко-математичне моделювання із застосуванням моделей ARIMA та SARIMA.

Наукова новизна одержаних результатів полягає у вдосконаленні підходів до прогнозування цін на житлову нерухомість шляхом порівняльного застосування моделей ARIMA та SARIMA для аналізу часових рядів ринку нерухомості Шотландії. У роботі здійснено оцінювання точності прогнозування на основі сучасних статистичних показників та визначено переваги використання сезонних моделей для прогнозування динаміки цін на житло.

Практичне значення отриманих результатів полягає у можливості їх використання для підвищення точності оцінки вартості нерухомості, прийняття інвестиційних рішень, а також у діяльності експертів-оцінювачів та аналітиків ринку.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ОСНОВИ ПРОГНОЗУВАННЯ ЦІН НА ЖИТЛОВУ НЕРУХОМІСТЬ

1.1 Поняття та особливості ринку нерухомості

Ринок нерухомості має свої унікальні характеристики, які відрізняють його від інших ринків товарів та послуг. Насамперед, нерухомість є фізично стаціонарним та невідтворюваним ресурсом, що призводить до обмеженості пропозиції в певних локаціях. Крім того, угоди з нерухомістю зазвичай вимагають значних фінансових вкладень та складне юридичного оформлення. Тривалий період володіння та експлуатації об'єктів, також впливає інвестиційну привабливість.

До особливостей ринку нерухомості можна віднести:

- низьку ліквідність та тривалі терміни продажу;
- значне державне регулювання;
- зростання вартості об'єктів нерухомості з часом;
- високий рівень трансакційних витрат.

Попит на об'єкти нерухомості індивідуалізований, що ускладнює досягнення ринкового рівноваги та пропорційності. Автори [1-3], виділяють такі характеристики ринку нерухомості:

- локальний характер ринку;
- низька взаємозамінність та стандартизованість об'єктів;
- закритий (приватний) характер угод;
- специфіка інформації на ринках;
- вплив нерациональних факторів на угоди;
- нееластичність пропозиції.

Ринок нерухомості характеризується обмеженою пропозицією та унікальними властивостями об'єкта. Процес купівлі проходить у закритому режимі та часто супроводжується недоліком інформації, у своїй немає двох однакових об'єктів нерухомості. Знайти сприятливі пропозиції на цьому ринку

буває складно, і в екстрених ситуаціях швидко знайти потрібний варіант практично неможливо через велику кількість часу для пошуку.

При створенні об'єктів нерухомості необхідно враховувати та дотримуватися стандартів, правил та норми, встановлені державою. У цьому контексті держава вимагає суворого виконання норм, підкреслюючи їх обов'язковість, точність та неухильність, без можливості вибору альтернативного варіанта поведінки.

Ринок нерухомості має свою особливість: середньостатистичний покупець часто не володіє необхідним досвідом у придбанні, знаннями про інвестиції в нерухомість та інформацією про реальні ціни угод. Ця відсутність кваліфікації призводить до невідповідності покупців та, як наслідок, до прийняття невірних рішень. Перед покупкою нерухомості важливо ретельно зібрати інформацію про пропоновані об'єкти з різних джерел і приділити цьому багато часу. Однак через зайнятість основною роботою у покупців не завжди є можливість займатися збором, розглядом, зіставленням, аналізом варіантів [4].

Форуми з продажу нерухомості найчастіше стають майданчиками для спілкування представників конкуруючих проектів, що призводить до накопичення необ'єктивної інформації. Деколи навіть професіонали не можуть відрізнити чесні відгуки від зацікавлених думок. Різні дослідження показують, що емоції та сприйняття цін відіграють ключову роль у процесі ухвалення рішення. На рішення покупців впливають різні фактори, включаючи розташування та стан об'єкта.

Законодавство грає ключову роль розвитку ринку нерухомості. Воно впливає на поведінку учасників, регулюючи відносини, що виникають у процесі підготовки, оформлення та реєстрації угод. Це включає питання оподаткування, ліцензування та захисту прав споживачів. Держава встановлює правові рамки для діяльності не лише у сфері ринку нерухомості, а й регулює взаємодію між суб'єктами ринками, що стосуються майнових аспектів угод. Крім того, воно забезпечує захист прав власників, гарантуючи охорону

приватної власності. Закон захищає право власності, і ніхто не може бути позбавлений свого майна без рішення суду [5].

Багато людей недостатньо обізнані щодо ситуації у різних сферах нерухомості. Важливо розуміти, що цей простір включає кілька взаємопов'язаних елементів: інформаційні; економічні; правові; соціальні та екологічні аспекти [6]. Інформаційне забезпечення ринку нерухомості включає в себе кілька ключових етапів:

- збір даних, який здійснюють кадастрові інженери, веб-портали, ріелтори та інші учасники ринку;
- системне структурування, верифікацію та кластеризацію даних, що виконуються аналітиками;
- моніторинг та аналіз ринку, при якому експерти досліджують більш складні та закриті області, спираючись на якісні аналітичні та первинні дані;
- прогнозування, що ґрунтується на результатах досліджень, які допомагають інвесторам, девелоперам, оцінювачам та іншим учасникам формувати та моделювати економічні, просторові та параметричні характеристики об'єктів, територій та всього ринку [7].

Ефективне управління інвестиціями в комерційну нерухомість є важливим для успішного функціонування даного сектора. Важливо розуміти, що ефективне управління комерційною нерухомістю - це комплексний підхід, заснований на різних показниках, прогнозах та тенденціях, оцінці ризиків, оптимізації витрат та багатьох інших складових компонентів.

1.2 Фактори, що впливають на формування вартості нерухомості

Однією з характерних рис ринку нерухомості є невідповідність очікувань власників та покупців, а також брак аналітичної та експертної інформації про ринок на фоні зростаючого попиту. У таких умовах власники об'єктів нерухомості піднімають ціни до невизначених рівнів, що призводить до кількості зниження кількості продажів, а попит зупиняється в очікуванні можливого зниження цін. Якщо різниця в очікуваннях сторін складає 10-15%, її

можна подолати, але коли вона досягає 50%, можна з упевненістю стверджувати, що угода не відбудеться.

Окрім інших особливостей ринку нерухомості можна виділити такі фактори, які впливають на вартість нерухомості [8]:

По-перше, його обмеженість та розділеність, що знаходить відображення у високому ступені різноманітності цін. Ринки нерухомості локалізовані не тільки в масштабах країни й по регіонах, а й у межах одного населеного пункту.

По-друге, ринки нерухомості індивідуалізовані, оскільки, на відміну більшості інших ринків, один єдиний покупець чи продавець здатний значно впливати на ціну.

По-третє, ринки нерухомості менш структуровані порівняно з звичайними ринками товарів, оскільки практично відсутні оптові продажі за зразками, немає єдиних цін, а інформація про ціни часто важкодоступна та ненадійна.

По-четверте, цей ринок характеризується високим рівнем транзакційних витрат. Транзакційні витрати включають усі витрати, пов'язані з обміном та захистом прав, які несуть як продавці, і покупці. Для того, щоб здійснити угоду з нерухомістю, покупцю необхідно зібрати інформацію не тільки про доступні на ринку об'єкти, відповідних його вимогам, а й про продавців цих об'єктів. Продавець, у свою чергу, повинен розмістити оголошення про продаж, підготувати необхідні правовстановлюючі документи, скласти контракт і так далі, або ж звернутися до ріелторської компанії, яка візьме продаж його нерухомості. Витрати, пов'язані з продажем нерухомості, можуть складати 6-7% вартості об'єкта.

По-п'яте, ринок нерухомості, на відміну від інших ринків, характеризується значною невідповідністю між попитом та пропозицією. Це зумовлено, з одного боку, тривалим процесом будівництва, а з іншого – відмінністю в очікуваннях продавців та покупців. Така невідповідність, у свою чергу, виникає через брак об'єктивної інформації про поточний стан ринку та його майбутні перспективи.

Кадастрова вартість землі в Україні – це офіційно встановлена вартість земельної ділянки, визначена на основі кадастрової оцінки. Кадастрова вартість використовується для цілей обліку та оподаткування земельних ділянок. Визначення кадастрової вартості землі в Україні здійснюється державним органом із земельних ресурсів або спеціалізованими кадастровими установами. Вони застосовують певну методику та критерії оцінки земельних ділянок. Кадастрова вартість землі враховує різні фактори, такі як місце розташування, категорія землі, родючість ґрунту, доступність інфраструктури та інші характеристики ділянки. Ці фактори допомагають визначити ринкову вартість земельної ділянки у конкретному районі.

Кадастрова вартість землі періодично оновлюється відповідно до зміни ринкових умов та розвитку інфраструктури. Оновлення може включати переоцінку земельних ділянок та перегляд кадастрової вартості. Кадастрова вартість землі є основою для розрахунку податків та зборів, пов'язаних із землею. Вона також використовується при угодах купівлі-продажу земельних ділянок, оренду або використання землі в комерційних цілях. Точні процедури та методологія визначення кадастрової вартості землі в Україні можуть бути уточнені у державного органу із земельних ресурсів або у кадастрових установах. Вихідна інформація для отримання кадастрових відомостей забезпечується і проведенням космоаерознімальних, топографо-геодезичних, ґрунтових, геоботанічних та інших обстежень та пошуків.

Для оперативного оновлення земельно-кадастрових відомостей використовуються матеріали поточних обстежень. Зйомок та моніторингу земель. Програма створення ринку нерухомості ставить проблему термінової організації надійної, відкритої, суспільної системи кадастру, яка займатиметься угодами з нерухомістю, чітко визначаючи тип угоди, сторони правочину, права на нерухомість. Для роботи земельного ринку потрібна чітка ідентифікація прав власності та меж ділянок. Оцінка ділянки вимагає ясності з ідентифікації ділянок та повних карток зазначенням кордонів, площ та накладених обмежень. Чітке формулювання права власності є неодмінною умовою при заставі землі.

1.3 Методи прогнозування вартості нерухомості

Аналіз вартості комерційної нерухомості вимагає використання різноманітних джерел даних, які можна класифікувати за походженням та наданою інформацією. Основні джерела включають офіційні статистичні дані, ринкові звіти аналітичних агентств, дані платформ для оренди та продажу нерухомості, а також інформацію від банків та фінансових установ. Кадастр надає офіційну статистику про ринок нерухомості, що робить його важливим ресурсом для аналізу.

Застосовуються такі системи, як геоінформаційні системи (ГІС), які допомагають в управлінні нерухомістю. Ці системи надають дані про місцезнаходження об'єктів, транспортні потоки та демографічні характеристики, що дозволяє визначити вартість, привабливість та конкурентоспроможність щодо об'єктів комерційної нерухомості. Використання цих інструментів допомагає власникам та керуючим ефективно адаптуватися до змін на ринку та оптимізувати свої стратегії.

Ринок комерційної нерухомості являє собою важливу частину економіки, схильну до впливу множини факторів. Одним із ключових аспектів, визначальних його розвиток, є обсяг інвестицій. Говорячи про компанію «CBRE», яка надала звіт за 2025 р., обсяг інвестицій у комерційну нерухомість у Європі склав близько 300 млрд євро, що наголошує на значущості цього сегменту як для інвесторів, так для економіки загалом. Формуванню уявлення про ринкові ціни сприяють звіти компаній, які займаються аналітичною діяльністю в галузі комерційної нерухомості. Наприклад, згідно з [7], можна спертися на дані про транзакції, що надаються агентствами "CBRE" та "Knight Frank". Високу достовірність даних підтверджує наявність глибокого вивчення ринкової активності та проведених транзакцій, про що у 2022 р. заявила компанія "CBRE". Таким чином, уявлення про вартість комерційної нерухомості може бути отримано в результаті комплексного аналізу та об'єднання відомостей від різних досліджуваних джерел.

Особливу важливість при дослідженні вартості комерційної нерухомості має визначення методів аналізу, які б дозволили отримати інформацію, необхідну для відстеження змін. Одним із поширених підходів є метод тимчасових рядів. За допомогою вивчення динаміки цін у часі у США з 2000 по 2020 р.р. був визначений 10-ти річний період відновлення рівня цін після кризової ситуації, що відбулася у 2008 р. [11].

Використання методу допомагає спрогнозувати зміни, розпізнати періодичність коливань, що повторюються, і встановити наявність аномалій, що важливо при формуванні уявлення про зміни вартості в наступних часових періодах. Для визначення впливу різних факторів на вартість застосовується регресійний аналіз. В рамках цього підходу дослідження може бути виявлено зв'язок з такими аспектами, як площа, місце розташування, різні показники макроекономіки. Можна зробити висновок про те, що за застосування цих методів виявляються фактори, що впливають на ринок нерухомості у сучасному світі.

Використання наведених методів показує свою необхідність при досягненні мети, спрямованої вивчення ринку нерухомості, що може бути показано на прикладах. Так, при аналізі ринку офісної нерухомості в Нью-Йорку з використанням часових рядів було виявлено тренд на щорічне збільшення вартості оренди у центральних районах на 2% у період з 2010 по 2020 рр., що дозволило інвесторам та девелоперам коригувати свої стратегії [12]. Такі приклади наголошують на важливості використання наукових методів для прийняття обґрунтованих рішень у сфері нерухомості.

Як зазначається в [13], «що склалися у світовій практиці технології оцінки ринкової вартості базуються на трьох методах»:

1. Порівняльний підхід. Вартість об'єкта моделюється шляхом зіставлення цін продажів, що недавно відбулися аналогічного майна у подібних ринкових умовах.

2. Прибутковий підхід. Оцінка проводиться виходячи з можливості об'єкта генерувати дохід, причому вартість вимірюється поточною вартістю очікуваних доходів від об'єкта, що оцінюється в майбутньому.

3. Витратний підхід. Виходить із припущення, що сума коштів, які треба витратити на будівництво об'єкта, що оцінюється, в його існуючому стані або відтворення споживчих властивостей даного об'єкта, еквівалентна його ринкової вартості.

Зростання попиту на преміальні об'єкти великих містах світу значно впливає на вартість офісних приміщень. Це пов'язано зі зміною переваг орендарів та інвесторів, які прагнуть придбання високоякісних об'єктів у стратегічно вигідних локаціях. Спостерігається загальний тренд до зростання цін на якісні комерційні площі у різних регіонах, що підтверджує актуальність інвестицій у висококласні об'єкти.

Макроекономічні чинники надають значний вплив ринку комерційної нерухомості, оскільки вони визначають загальні умови економічного середовища. У 2025 р. глобальне зростання ВВП, згідно з звітом Міжнародного валютного фонду, що знизився до 3,2%. Це уповільнення економічного зростання створило тиск на ринки нерухомості, особливо в країнах з економікою, що розвивається, де комерційна нерухомість часто виступає важливим інвестиційним активом. Крім того, рівень інфляції в США досяг 8,6%, що призвело до збільшення витрат на будівництво та зниження інвестиційної активності.

Ці зміни демонструють, як макроекономічна нестабільність може вплинути на ринок нерухомості, знижуючи його привабливість для інвесторів. Поточні ринкові умови істотно впливають формування вартості комерційної нерухомості. Середня доходність від оренди комерційних об'єктів у Європі знизилася до 4,7%, що нижче за рівень 2023 р., що становив 5,1%. Це зниження пояснюється зростанням процентних ставок, що збільшує вартість позикових коштів і, як наслідок, знижує прибутковість інвестицій.

Водночас обсяг інвестицій у комерційну нерухомість в Азіатсько-Тихоокеанському регіоні у 2025 р. зменшився на 20% порівняно з попереднім роком, що також пов'язано із нестабільністю ринкових умов. Таким чином, поточні економічні та ринкові зміни чинять значний тиск на вартість та прибутковість комерційної нерухомості.

1.4 Характеристика ринку нерухомості Шотландії

Ринок нерухомості Шотландії є важливою складовою економічної системи Великої Британії та характеризується відносною стабільністю, регіональною диференціацією та помірними темпами розвитку. Незважаючи на інтеграцію в загальнонаціональний ринок, він має низку специфічних особливостей, зумовлених економічними, соціальними та інституційними чинниками. Передусім, слід зазначити, що ринок нерухомості Шотландії відрізняється більшою доступністю житла порівняно з іншими регіонами Великої Британії, зокрема Лондоном. Це пояснюється нижчим рівнем цін, меншою щільністю населення та відносно рівномірним розвитком територій. Водночас попит на житло залишається стабільно високим, що пов'язано зі зростанням чисельності населення у великих містах та активними процесами урбанізації [14].

Шотландський ринок нерухомості (SPRS) швидко зростає протягом останніх 25 років і тепер відіграє важливу роль у забезпеченні житлом Шотландії та її економіки. Відновлення, яке відображає зміни в решті Великої Британії, стало можливим завдяки дерегуляції (серед інших рушійних сил) та зумовлене великою групою гетерогенних приватних інвесторів, які, як правило, володіють невеликою кількістю об'єктів нерухомості. Хоча багато було написано про характеристики та мотивацію цих «орендодавців», шотландські дані зараз застаріли, і мало що відомо про інвестиційну поведінку орендодавців у контексті нормативних та описових теорій інвестування. Це викликає занепокоєння, враховуючи, що ці характеристики, мотивація та поведінка пояснюють вибір орендодавцем SPRS як стратегії соціального забезпечення так

і майбутнім виходом з інвестицій. Вони також у більш широкому сенсі відповідають за формування досвіду.

Таким чином, дефіцит даних має далекосяжні наслідки, які є особливо суттєвими для політиків, які прагнуть подальшого регулювання цього сектору, мінімізуючи при цьому непередбачувані наслідки. Що є суттєвою характеристикою ринку і його регіональною неоднорідністю. Найвищі ціни на нерухомість спостерігаються у великих економічних центрах, таких як Единбург та Глазго, де зосереджені фінансові, освітні та культурні ресурси. Водночас у менш розвинених регіонах, зокрема у західній частині країни, рівень цін значно нижчий, що формує нерівномірність розвитку ринку [15].

Структурно ринок нерухомості Шотландії поділяється на кілька основних сегментів: житлова, комерційна та орендна нерухомість. Найбільшу частку займає житловий сектор, який включає квартири, індивідуальні житлові будинки та таунхауси. Останніми роками спостерігається зростання попиту на приватні будинки, що пояснюється зміною пріоритетів населення, зокрема поширенням дистанційної роботи та прагненням до підвищення якості життя.

Важливу роль у функціонуванні ринку відіграють економічні фактори, зокрема рівень доходів населення, доступність іпотечного кредитування та процентні ставки. Крім того, на ринок впливають соціальні чинники, такі як демографічні зміни, внутрішня міграція та урбанізаційні процеси. Інституційні фактори, включаючи державне регулювання, податкову політику та житлові програми, також суттєво визначають умови розвитку ринку. На ринки землі впливало і зараз впливає включення «вартості» природного капіталу, що часто призводить до вищих вартостей передачі або інвестицій. Однак нещодавнє дослідження, проведене в Шотландії, висвітлює, як невизначеність та відсутність прозорості у визначенні вартості природного капіталу призвели до коригування вартості землі протягом 2023 та 2024 років [16].

Близько 84% усієї зареєстрованої нерухомості SPRS належать приватним особам, і 95% з них володіють від однієї до п'яти об'єктів нерухомості. Це свідчить про те, що в SPRS переважають дрібні інвестори. Цих інвесторів, які

схильні розглядати SPRS як стратегію соціального забезпечення, часто називають «аматорами» через неповний робочий день у своїй інвестиційній діяльності та вторинний дохід, який вона часто забезпечує. Однак, враховуючи «професійний» статус, який зазвичай надається орендодавцям у секторі будівництва для здачі в оренду (BTR) (Homes for Scotland), цей термін також можна тлумачити як питання експертизи, досвіду та компетентності (табл. 1.1).

Інвестиційні рішення цих «аматорських» приватних орендодавців відіграють вирішальну роль у формуванні SPRS (Системи будівництва для оренди). Важливо, що хоча SPRS може запропонувати орендодавцям більший ступінь захисту капіталу порівняно з деякими альтернативами, такими як акції та паї, він залишається за своєю суттю ризикованою інвестицією, прибутковість якої залежить від широкого спектру систематичних та несистематичних ризиків. Важливо, що це також має наслідки для стабільності інвестиційних моделей та умов, з якими стикаються орендарі, які проживають у цьому секторі.

Таблиця 1.1

Основні показники ринку нерухомості Шотландії		
Показник	Значення (2024–2025 рр.)	Характеристика
Середня ціна житла	£190 000 – £195 000	Нижча, ніж у середньому по Великій Британії
Річне зростання цін	5–6%	Стабільна позитивна динаміка
Зростання за 5 років	≈ 25–28%	Довгострокове зростання
Рівень попиту	Високий	Перевищує пропозицію
Основні покупці	Домогосподарства, інвестори	Активний внутрішній попит
Доступність житла	Середня/вища	Краща, ніж у Лондоні
*розроблено автором на основі [17]		

Сучасний етап розвитку ринку нерухомості Шотландії характеризується низкою тенденцій. Зокрема, спостерігається зростання попиту на доступне житло, дефіцит пропозиції на первинному ринку, а також активізація будівництва у передмістях великих міст. Водночас ринок демонструє стійкість до економічних коливань, що свідчить про його відносну зрілість та адаптивність. Разом з тим, існують певні проблеми, що стримують розвиток ринку нерухомості Шотландії. До них належать недостатній обсяг нового житлового будівництва, регіональна диспропорційність розвитку, а також обмежена доступність житла для окремих категорій населення, зокрема молоді. Крім того, ринок залишається чутливим до змін макроекономічної ситуації, включаючи інфляційні процеси та коливання фінансових ринків [18].

Висновки до розділу 1

Отже, ринок нерухомості Шотландії є важливим елементом економіки країни, що характеризується стабільністю, помірним зростанням та значною регіональною диференціацією. Його подальший розвиток залежатиме від ефективності державної політики, рівня інвестиційної активності та загальної економічної ситуації в країні.

РОЗДІЛ 2

АНАЛІЗ ДАНИХ РИНКУ НЕРУХОМОСТІ ТА МЕТОДИКА ДОСЛІДЖЕННЯ

2.1 Застосування технологій штучного інтелекту у прогнозуванні економічних процесів

Використання штучного інтелекту дозволяє значно підвищити точність ціноутворення об'єктів комерційної нерухомості. Прикладом може бути думка експертів у цій області про те, що застосування ШІ в аналізі ринку комерційної нерухомості може збільшити точність оцінки приблизно на 15%. Це значні цифри, оскільки ще кілька років тому це було з області фантастики і розглядалося лише теоретично, але тепер вплив ШІ все змінив. Така ефективність досягається за рахунок аналізу великих обсягів даних та обліку безлічі факторів, що впливають на вартість, включаючи ринкові тенденції, фізичний стан об'єкта нерухомості, його місцезнаходження тощо.

Оцінки ефективності будівельних проектів повинні включати не лише одноразові витрати (сама будівля), а й витрати на експлуатацію. Такий підхід у даний час актуальний для всього світу. У Європі найпоширенішою методикою оцінки ефективності будівельних проектів є методика LCC (Life Cycle Cost) для будівель, ґрунтуючись на відповідних стандартах [6]. У існуючих на даний час методиках оцінки та прогнозування вартість життєвого циклу будівлі враховуються лише будівельні, експлуатаційні, характеристики, але не розглядаються фактори мінливого зовнішнього економічного середовища, які також дуже сильно впливають на сукупну вартість об'єкта, а також географічні характеристики. Отже, з метою отримання більш точної методики виникає потреба у вдосконаленні існуючих методичних підходів.

Рішенням є створення комплексної методики оцінки та прогнозування вартості житла з використанням машинного навчання на основі нейромережевих технологій. Алгоритми нейронних технологій широко застосовується в різних галузях науки, в тому числі в економіці. Застосування

методів, з використанням машинного навчання на нейромережевих технологіях дозволяє вирішити деякі проблеми економіко-статистичного моделювання, що дозволяє підвищити адекватність математичних моделей наблизити їх до економічної реальності.

Машинне навчання (*machine learning*) – метод аналізу даних, заснований на комплексному застосуванні статистики для пошуку закономірностей даних і створення на його основі потрібних прогнозів, де відбивається не пряме рішення задачі, а навчання у процесі застосування рішень множини подібних завдань. Навчання за допомогою штучних нейронних мереж одна із алгоритмів машинного навчання.

Розглянемо деякі дослідження застосування машинного навчання на основі нейромережевих технологій для оцінки та прогнозування вартості життєвого циклу об'єктів нерухомості. Автори [10] розробили нову схему оцінки вартості LCC з використанням штучної нейронної мережі для спрощення та покращення процесів оцінки. У роботі авторів Mohammed Kishk, Assem Al-Hajj зазначили, що розрахунок вартості життєвого циклу не повністю укладається у рамки теорії ймовірностей та статистики. Очевидним напрямом підвищення якості LCC як інструменту підтримки прийняття рішень є використання методів штучних нейронних мереж та теорії нечітких множин. Ці методи можуть моделювати нелінійні характеристики, узагальнювати та мати справу з ситуаціями, включаючи неповну інформацію, людське судження та невизначеність.

Young, з колегами [31] провели аналіз вартості життєвого циклу освітніх установ з урахуванням машинного навчання. Були зібрані та проаналізовано 121 освітніх закладів. Дослідження моделі виявили значні входні параметри, що впливають на LCC будівлі. Слід зазначити наявні переваги в оцінці вартості життєвого циклу об'єктів, які були відображені у роботах вищезгаданих авторів:

- використовуються дійсно необхідні інформації та дані, які суттєво впливає на вартість життєвого циклу об'єктів нерухомості. Це знизить вартість проведення аналізу LCC;
- дозволяють особам, які приймають рішення, систематично та швидко отримувати оцінки та рекомендації щодо прийняття рішень;
- зменшення суб'єктивних елементів, які переважають в оцінці вартості життєвого циклу.

Очікується, що, зробивши процес аналізу більш об'єктивним, простим і менш дорогим, методика оцінки та прогнозування вартості будівлі може знайти широке практичне застосування у будівельній галузі. Розрахунок вартості життєвого циклу житлової будівлі здійснюється у чотири етапи. Алгоритм розрахунку сукупної вартості житлової будівлі із застосуванням машинного навчання на основі нейромережових технологій представлений нижче.

Перший етап: збирання вихідних даних

- перелік та елементи будівлі;
- визначення тривалості життєвого циклу кожного елемента розрахунку;
- визначення кількості та періодичність проведення ремонтів та заміни обладнання.

Другий етап: розрахунок одноразових витрат

- одноразові витрати до введення в експлуатацію (на придбання земельних ділянок, проектування на підключення до інженерних мереж; вартість матеріалів та обладнання, витрати на будівельно-монтажні роботи та ін.)
- одноразові витрати на знесення (утилізація).

Третій етап: розрахунок періодичних витрат із застосуванням машинного навчання та нейромережових технологій протягом планового періоду експлуатації пов'язаних з:

- витратами на експлуатацію;
- витратами на поточний ремонт;

- витратами на капітальний ремонт;
- витратами на утримання майна.

Четвертий етап: розрахунок витрат життєвого циклу шляхом підсумовування результатів другого та третього етапів. Як ми бачимо, останні дослідження в галузі використання машинного навчання на основі нейромережових технологій в рамках оцінки життєвого циклу будівлі визначили, що система є надійною і представляє цінність для зацікавлених осіб у будівельній сфері. В результаті проведеного дослідження можна зробити висновок, що методики оцінки вартості життєвого циклу будівель із застосуванням машинного навчання на основі нейромережових технологій на етапі проектування має велику значимість та потенціал у будівельній сфері України.

Компанії у поточних реаліях намагаються оптимізувати свої процеси та адаптуватися до багатьох змін на ринку комерційної нерухомості. Використовуючи просунуті алгоритми та аналіз цифрових даних, штучний інтелект та нові технології дозволяють значно впливати на ринок комерційної нерухомості, що наголошує на явному потенціалі нових технологій, а це виражається в ефективному управлінні комерційного об'єкта нерухомості. Застосування технологій великих даних також відкриває нові можливості для аналізу ринкових реалій та споживчої поведінки.

2.2 Аналіз часових рядів для прогнозування вартості нерухомості

Аналіз часових рядів є одним із ефективних статистичних методів прогнозування вартості нерухомості. Він базується на дослідженні змін показників у часі та дозволяє виявити закономірності розвитку ринку нерухомості. Основною метою цього підходу є визначення тенденцій, сезонних коливань та випадкових змін, які впливають на динаміку цін.

Часовий ряд представляє собою послідовність значень певного показника, зафіксованих у різні моменти часу. У випадку ринку нерухомості такими показниками можуть бути середня ціна квадратного метра житла, кількість

укладених угод купівлі-продажу, рівень попиту та пропозиції на ринку. Під час аналізу часових рядів виділяють кілька основних компонентів. До них належать тренд, який відображає довгострокову тенденцію зміни цін; сезонні коливання, що виникають у певні періоди року; циклічні коливання, пов'язані з економічними процесами; а також випадкові фактори, які можуть спричиняти короткострокові зміни вартості нерухомості.

Для прогнозування цін на нерухомість за допомогою часових рядів використовують різні методи. Серед них найбільш поширеними є метод ковзної середньої, експоненційне згладжування, а також економетричні моделі, зокрема моделі ARIMA та SARIMA. Ці моделі дозволяють оцінити майбутні значення цін на основі попередніх статистичних даних. Використання аналізу часових рядів дає можливість враховувати історичну динаміку розвитку ринку нерухомості та будувати обґрунтовані прогнози щодо зміни вартості житла. Це має важливе значення для інвесторів, забудовників, банківських установ і потенційних покупців, оскільки допомагає приймати більш ефективні економічні рішення. Таким чином, аналіз часових рядів є важливим інструментом дослідження ринку нерухомості, який дозволяє прогнозувати зміну цін та оцінювати перспективи розвитку цього сектору економіки.

Впровадження сучасних технологій управління орендарями також істотно впливає збільшення заповнюваності. Дослідження компанії «CBRE», проведене 2021 р., показало, що використання таких технологій може підвищити заповнюваність об'єктів на 15–20%. Ці методи дозволяють не тільки мінімізувати простой, а й забезпечувати стабільний дохід від оренди, що робить їх важливою складовою стратегії управління комерційною нерухомістю [1].

Для аналізу комерційної нерухомості все частіше використовують програмні продукти, які точно та ефективно можуть аналізувати ті чи інші показники для компанії, вважати доходи та планувати витрати, прогнозувати майбутні значення. Програмні продукти, такі як «Yardi» та «MRI Software», дозволяють власникам нерухомості грамотно керувати нерухомістю з урахуванням цих програм. Ці ІТ-системи приймають виважені та обґрунтовані

рішення, враховуючи першоджерело інформації. Інші програмні рішення, такі як "Tableau" і "Power BI", забезпечують обробку великих обсягів даних, таких як ринкові тенденції та показники ефективності об'єктів Глибокий аналіз після роботи даних програм точно забезпечений, оскільки дані системи вже давно були протестовані і можуть використовуватись у найбільших бізнесах. У сучасних реаліях багато компаній спираються на штучний інтелект та Big Data (великі дані), які, в свою чергу, допомагають прогнозувати зміни в ринкових умовах, наприклад, динаміку орендних ставок чи попит певні типи об'єктів.

2.3 Методи моделювання та прогнозування часових рядів

Як відомо, основний принцип торгівлі на фінансовому ринку визначається «золотим правилом інвестування»: «купуй дешево, продавай дорого» [18]. Даний принцип створює необхідність у проведенні аналізу ринку, метою якого є прогнозування майбутніх значень курсів активів, що торгуються. Можна виділити два підходи до моделювання ринкових даних: моделювання самих нестационарних курсів або одноперехідних пристроїв та їх подальше моделювання. У рамках другого підходу так чи інакше всі методи передбачають використання автокореляції.

Альтернативний підхід, що полягає у моделюванні нестационарних часових рядів курсів, що не може спиратися на підходи, що ґрунтуються на використанні лінійних закономірностей динаміки нестационарних часових рядів, оскільки нестационарні часові ряди автокореляції прагнуть одиниці. Це зумовлює необхідність використання інших нелінійних моделей та комплексних підходів, які здатні розкрити сутність складної динаміки.

В роботі пропонується та досліджується метод моделювання та прогнозування однопериодних змін вихідного часового ряду. В основі даного методу лежить ідея використання ансамблю моделей на основі декомпозиції вихідного часового ряду для завдання моделювання курсів, описана у [2]. Він використовує алгоритм CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise), який здійснює декомпозицію вихідного

тимчасового ряду на осцилюючі компоненти, різних порядків частот, які називаються модами. За результатами розкладання високочастотних компонентів є очевидним статистична модель ARMA.

Частина низькочастотних компонентів, які в певному сенсі описують криву тренду, не є стаціонарними і вимагають моделювання за допомогою інших нелінійних методів. Зокрема автори статті [2] пропонують для їх моделювання використовувати нейронну мережу архітектури LSTM. Сума прогнозів у всіх компонентах утворює комбінований прогноз ансамблю моделей. На практиці з моделюванням нестационарних компонентів пов'язані основні недоліки описаного підходу. Велика кількість нестационарних компонентів погіршує якість сукупного прогнозу внаслідок низької точності можливого прогнозу нестационарних компонентів.

У зв'язку з цим точність прогнозування при аналізі одноперіодних збільшень курсів буде вища точність прогнозування при аналізі самих курсів, оскільки кількість нестационарних компонентів у розкладанні стаціонарних одноперіодних збільшень курсів менше, ніж у розкладанні нестационарних курсів, які динаміка і форма простіше. Крім цього, за результатам експериментів відомо [18], що якість прогнозу нейронної мережі LSTM для нестационарних компонентів трохи гірше якості прогнозу моделі Arima. При цьому використання та навчання моделі ARIMA значно швидше за використання та навчання нейронної мережі LSTM. Таким чином, трохи втрачаючи в точності можна досягти значного приросту в швидкості роботи алгоритму, що є суттєвим у контексті використання в реальних додатках.

Розроблений метод в два етапи: декомпозиція тимчасового ряду за допомогою алгоритму CEEMDAN та побудова комбінованого прогнозу ансамблю моделей ARMA та SARIMA для одержаних компонентів розкладання. Алгоритм розкладання CEEMDAN базується на використанні процедури EMD (Empirical Mode Decomposition) Процедура EMD декомпозиції тимчасових рядів полягає в розкладанні тимчасового ряду у функції внутрішніх

мод IMF (Intrinsic Mode Function). Функції IMF задовольняють наступним властивостям [19]:

1. Кількість локальних мінімумів та максимумів функцій IMF має дорівнювати кількості перетинів функцією нуля, або відрізнятися не більше ніж на одиницю.

2. У будь-якій точці середнє значення між лінією, що обгинає, локальний мінімум і лінію локального максимуму має дорівнювати нулю.

Такі властивості функцій IMF забезпечують їм наявність сильнішої автокореляції порівняно з вихідним тимчасовим рядом, що гарантує високу ефективність прогнозів, заснованих на даній властивості. За рахунок цього сукупний прогноз щодо компонентів розкладання є більш точним, ніж прогноз по вихідному часовому ряду.

Алгоритм декомпозиції EDM є ітеративним і включає наступні кроки [20]:

1. Визначаємо локальні мінімуми та максимуми аналізованого часового ряду. Вони з'єднуються лінією кубічного сплайну, утворюючи 2 огинальні лінії, для яких знаходимо середню лінію m_1 .

2. Визначаємо перший компонент процедури просіювання: $h_{01} = x - m_1$.

3. Процес ітеративно повторюється доки k -ий компонент не стане функцією IMF: $h_{1k} = h_{1(k-1)} - m_{1k}$.

4. k -ий компонент процесу “просіювання” стає компонентом розкладання: $c_1 = h_{1k}$.

5. Визначається низка залишків: $r_1 = x - c_1$.

6. Процес повторюється починаючи з першого пункту для низки залишків r_1 .

Такий алгоритм дозволяє розкласти вихідний часовий ряд у компоненти різних порядків частот, при цьому зберігаючи його початкову структуру. Підхід EMD має деякі проблеми [4]: в рамках одного компонента можуть бути осциляції амплітуд, що відрізняються, у свою чергу подібні осциляції можуть

спостерігатися у різних компонентах, що називається “змішенням мод” (Mode mixing). Описані проблеми декомпозиції можуть бути вирішені з допомогою побудови ансамблю алгоритмів розкладання та додавання адаптивного шуму на початку кожної ітерації до аналізованого тимчасового ряду. Такий алгоритм називається CEEMDAN, особливості його роботи представлені у роботах [21-22].

2.4 Опис моделей ARIMA та SARIMA з використанням python, MAE, RMSE, MAPE

З традиційних статистичних моделей для аналізу часових рядів обрані:

ARIMA (Autoregressive Integrated Moving Average) є моделлю прогнозування тимчасових рядів, яка поєднує три основні компоненти: авторегресію (AR), інтеграцію (I) та ковзне середнє (MA). Вона призначена для аналізу одновимірних тимчасових рядів, які є стаціонарними або які можна призвести до стаціонарного шляхом диференціювання. AR(p) – частина авторегресії припускає, що поточне значення ряду можна пояснити через його попередні значення. I(d) – розділ інтеграції означає, що дані можуть вимагати диференціювання d разів, щоб стати стаціонарними. MA(q) - компонент ковзного середнього передбачає, що поточне значення ряду можна пояснити через попередні помилки прогнозу.

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2.1)$$

де Y_t – оператор лага;

c – константа;

ϕ_i – коефіцієнти авторегресії;

θ_i - Коефіцієнти ковзного середнього;

d – порядок диференціювання;

ε_t – помилка прогнозу в момент часу t.

SARIMA (Seasonal ARIMA) є розширенням ARIMA моделі для обліку сезонності у даних. Оскільки тимчасові ряди мають сезонні патерни, то

SARIMA додає сезонні елементи в ARIMA для моделювання часових рядів. SARIMA позначається як $ARIMA(p, d, q)(P, D, Q)[S]$, де P, D, Q – сезонні аналоги параметрів p, d та q , а S – довжина сезонного періоду.

$$\Phi(B^m)\phi(B)(1-B)^d(1-B^m)^D y_t = \Theta(B^m)\theta(B)\varepsilon_t \quad (2.2)$$

де y_t – тимчасовий ряд;

$\Phi(B^m)$ – поліном сезонної авторегресії (SAR) порядку P ;

$\phi(B)$ – поліном авторегресії (AR) порядку p ;

B – оператор зсуву назад (лаг оператор), такий, що $B^k y_t = y_{t-k}$;

$\Theta(B^m)$ – поліном сезонної ковзної середньої (SMA) порядку Q ;

$\theta(B)$ – поліном ковзної середньої (MA) порядку q ;

d – порядок несезонного диференціювання;

D – порядок сезонного диференціювання;

ε_t – білий шум (випадкова помилка).

Модель Хольта-Вінтерса є методом експоненційного згладжування для аналізу часових рядів із сезонними компонентами. Вона розширює модель експоненційного згладжування Хольта, додаючи компонент для сезонності. Модель включає три рівняння: рівень, тренд та сезонність. Кожен компонент має свій коефіцієнт згладжування: α (рівень), β^* (тренд) та γ (сезонність). Модель може бути адаптована для додавання (адитивної) або множення (мультиплікативної) сезонності в залежності від характеру даних.

Формула (3) для згладженого значення тренда представлена нижче:

$$l_t = \alpha(y_t - s_{t-S}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (2.3)$$

де y_t – Фактичне значення в момент часу (t);

s_{t-S} – сезонний компонент у момент часу ($t-S$) (де S – період сезонності);

l_{t-1} – згладжене значення тренду у попередній момент часу;

b_{t-1} – швидкість зміни тренда попереднього часу;

α – параметр згладжування для тренду.

Формула (4) для згладженого значення сезонного компонента представлена нижче:

$$s_t = \gamma(y_t - l_{t-1}) + (1 - \gamma)s_{t-S} \quad (2.4)$$

де s_{t-S} – сезонний компонент у момент часу;

γ – параметр згладжування для сезонного компонента, $0 \leq \gamma \leq 1$.

Формула (5) для прогнозу (m) періодів уперед $\hat{y}_{t+m}$:

$$\hat{y}_{t+m} = l_t + m b_t + s_{t+m-S} + (m - 1)b_S \quad (2.5)$$

де l_t – згладжене значення тренду на момент часу (t);

b_t – швидкість зміни тренду в момент часу (t);

s_{t+m-S} – сезонний компонент для моменту часу (t+m-S);

b_S – середня зміна тренду за період сезонності (S).

З методів машинного навчання буде досліджено:

Лінійна регресія одна з найголовніших і широко використовуваних типів прогнозних моделей через її простоту та ефективність в певних випадках. Вона будується на припущенні про лінійну залежність між однією або декількома незалежними змінними та залежною змінною (обсягом продажів). У разі однієї незалежної змінної вона називається простою лінійною регресією, а при кількох – множинною лінійною регресією.

У ході використання різних інструментів та технологій для побудови моделі машинного навчання з метою прогнозування продажів було ухвалено рішення скористатися мовою програмування Python. Цей вибір ґрунтується на низці ключових факторів, які роблять Python найбільш підходящим для реалізації завдань машинного навчання, будучи високорівневою мовою програмування загального призначення, відомою своєю динамічною строгою типізацією та автоматичним управлінням пам'яттю. Ці характеристики значно спрощують процес розробки, підвищують продуктивність та забезпечують високу читальність та якість коду. Python сприяє переносимості програм, що робить його ідеальним для розробки масштабованих та підтримуваних моделей машинного навчання. Практично всі сучасні моделі машинного навчання підтримуються у ньому. Це забезпечує легкість інтеграції та адаптації моделей у різних середовищах та платформах.

Словник Python дозволив легко застосовувати різні моделі та методи обробки даних до кожного окремого часового ряду, що зробить модель більш гнучкою та адаптивною. Це особливо важливо для масштабованості рішення, так як зі збільшенням числа магазинів модель можна легко адаптувати без значних змін у коді. Також це спростить процес валідації та тестування моделі, оскільки дані можна буде обробляти та аналізувати незалежно. Загалом, застосування словника Python для роботи з множинними тимчасовими рядами в наборі даних M2 виявилось успішним рішенням, яке дозволило значно покращити якість прогнозів та спростить процес розробки моделі. Для оцінки якості моделей були використані: середньоквадратичні помилки (MSE), середня абсолютна помилка (MAE), коефіцієнт детермінації (R^2) та MAPE.

MAPE (Mean Absolute Percentage Error) – це метрика оцінки помилки прогнозу, яка вимірює середню абсолютну процентну помилку між фактичними та прогнозованими значеннями, що допомагає оцінити, наскільки модель точно прогнозує щодо фактичного значення. MAPE буде використана як бізнес-метрика для надання аналітики, наскільки прогнозовані значення відповідають реальним показникам. Формула даної метрики представлена нижче:

$$\text{MAPE} = \frac{\sum_{i=1}^n \left| \frac{y_{pred,i} - y_{true,i}}{y_{true,i}} \right|}{n} \times 100\% \quad (2.6)$$

де y_{pred} – передбачені значення моделі;

y_{true} – реальні значення (справжні мітки);

n – загальна кількість спостережень та прогнозів;

i – індекс, конкретне спостереження.

Середня абсолютна помилка (MAE) є середнім абсолютним значенням різниці між прогнозованими та реальними значеннями. Ця метрика дозволяє оцінити величину помилки прогнозу в вихідних одиницях виміру.

Середньоквадратична помилка (MSE) є середнім квадратом різниці між прогнозованими та реальними значеннями. Вона враховує не тільки величину помилки, а й її варіабельність.

MAE відображає середню абсолютну помилку прогнозу в одиницях і балах. Менші значення MAE вказують на точніші прогнози. MAE розраховується за формулою [9]:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.7)$$

MAPE визначає середню похибку у відсотках і визначається як [10]:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| / y_i \times 100\% \quad (2.8)$$

RMSE враховує квадрати похибок, що робить цю метрику чутливою до великих відхилень. Для його визначення використано наступну формулу [12]:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.9)$$

На основі проведених розрахунків було отримано наступні значення показників точності для різних моделей прогнозування (табл. 2.1) [36].

Таблиця 2.1

Порівняння точності моделей прогнозування

Модель	MAE	RMSE	MAPE (%)
ARIMA	13 407.54	14 344.26	7.44
SARIMA	4065.63	5 276.67	2.29

У ході дослідження було встановлено, що сучасні методи машинного навчання, зокрема, SARIMA забезпечують значно вищу точність прогнозування порівняно з класичними статистичними моделями. Отримані значення показників точності дозволяють оцінити адекватність побудованих моделей та їх придатність для практичного використання.

Коефіцієнт детермінації (R^2) є основною статистичною метрикою оцінки точності та пояснюваної частини мінливості в модельному прогнозі. Використовується для оцінки, наскільки добре модель пояснює залежність між прогнозами та вихідними або контролюючими даними (фактичним попитом). Розраховуємо також дві метрики оцінки продуктивності моделі за допомогою поділу даних на навчальні та тестові набори з розрахунком середнього значення R^2 (Avg, R^2 (Cross-Validation)) та середнього значення MSE (Avg, MSE (Cross-

Validation)). При розрахунку Avg, MSE, R^2 у контексті крос-валідації, вираховуємо значення цих метрик для кожного фолду (частини даних) і потім усереднюємо по всіх фолдах для здобуття загального значення оцінки продуктивності моделі. Це дозволяє отримати більш надійні результати оцінки моделі на основі різних частин даних.

Для підвищення якості аналізу та моделювання часових рядів застосовані різні методи та критерії. Інформаційний критерій Акаїці стане інструментом для порівняльної оцінки відносного якості моделей та вибору найбільш підходящої, враховуючи як точність, так і складність моделі. Перетворення Бокса-Кокса дозволить стабілізувати дисперсію та наблизить розподіл даних до нормального, що позитивно позначиться на якості моделі та точності прогнозування. Логарифмування даних буде використано для зменшення розкиду значень та лінеаризації трендів, особливо у випадках експоненційного зростання чи великого розкиду значень. Тест Дарбіна-Уотсона допоможе виявити наявність автокореляції першого порядку залишки регресійної моделі, що може вказувати на неадекватність моделі та необхідність її доопрацювання.

Висновки до розділу 2

У другому розділі було розглянуто особливості застосування технологій штучного інтелекту для прогнозування економічних процесів на ринку нерухомості. Встановлено, що аналіз часових рядів є ефективним інструментом дослідження динаміки вартості нерухомості та дозволяє виявляти тенденції зміни ринкових показників.

Було охарактеризовано основні методи моделювання часових рядів та визначено переваги моделей ARIMA і SARIMA для прогнозування економічних даних. Також розглянуто основні метрики оцінювання точності прогнозу – MAE, RMSE та MAPE. Використання Python і сучасних бібліотек аналізу даних дозволяє автоматизувати процес побудови моделей та підвищити ефективність прогнозування вартості нерухомості.

РОЗДІЛ 3

ПРОГНОЗУВАННЯ ЦІН НА НЕРУХОМІСТЬ ТА ОЦІНКА ЯКОСТІ ПРОГНОЗУ

3.1 Опис та підготовка вихідних даних (1968–2025)

Для наукового дослідження було використано сайт, що містить статичний набір даних про ціни на нерухомість у Великій Британії. Джерело даних – відкритий файл історичних цін на нерухомість, доступний для завантаження з Інтернету ([Індекс цін на житло у Великій Британії: завантаження даних серпень 2025 - GOV.UK](https://www.kaggle.com/datasets/swarupsudulaganti/uk-house-price-prediction-dataset-2015-to-2024/data)) та даних: (<https://www.kaggle.com/datasets/swarupsudulaganti/uk-house-price-prediction-dataset-2015-to-2024/data>).

Файл охоплює тривалий період, починаючи з 1968х років, що дозволяє аналізувати динаміку змін цін на нерухомість протягом багатьох років. Це забезпечує достатньо даних для тренування та тестування моделей прогнозування. Дані [8] збираються щомісячно, що забезпечує високу точність та деталізацію для аналізу та побудови прогнозних моделей.

Кожен запис у файлі складається з:

- дата (Date): стовпчик, що містить дату спостереження у форматі YYYY-MM-DD.
- ціна (Price): стовпчик, що містить середню ціну на нерухомість у відповідний місяць.
- регіон (Region): стовпчик, що містить назву регіону.

Для дослідження процесу прогнозування цін на нерухомість було використано набір даних, що містить часові спостереження за ринком житлової нерухомості. Дані представлені у вигляді таблиці, де кожен запис відповідає певному часовому періоду.

Структура набору даних

Набір даних включає такі основні змінні:

- Date – дата спостереження;

- RegionName – регіон дослідження;
- AveragePrice – середня ціна нерухомості;
- Index, IndexSA – індекси цін (з урахуванням сезонності та без);
- 1m%Change, 12m%Change – зміни цін за 1 місяць та 12 місяців;
- SalesVolume – обсяг продажів;

Додатково представлено деталізацію за типами житла:

- DetachedPrice – ціна окремих будинків;
- SemiDetachedPrice – напівокремі будинки;
- TerracedPrice – таунхауси;
- FlatPrice – квартири;

Також містяться дані за способом придбання:

- CashPrice – ціни при купівлі за готівку;
- MortgagePrice – ціни при іпотечному кредитуванні;

Окремо виділені групи покупців:

- FTBPrice – покупці вперше;
- FOOPrice – інші власники;

А також:

- NewPrice – новобудови;
- OldPrice – вторинний ринок.

Аналіз динаміки цін на нерухомість за період 1968–2025 рр. свідчить про наявність чітко вираженої тенденції до зростання. Упродовж досліджуваного періоду спостерігається поступове збільшення вартості нерухомості, що обумовлено впливом інфляційних процесів, урбанізації, зростання попиту та загального розвитку економіки. На початковому етапі (1968–2000 рр.) зростання цін відбувалося відносно повільно та рівномірно. Проте починаючи з 2000-х років темпи зростання значно прискорюються, що пов'язано з активізацією ринку нерухомості та збільшенням інвестицій у цей сектор.

Окремі періоди характеризуються підвищеною нестабільністю, що проявляється у коливаннях цін. Такі коливання можуть бути пов'язані з економічними кризами, змінами макроекономічної ситуації, а також

зовнішніми факторами, такими як фінансові потрясіння чи глобальні події. Після 2015 року спостерігається найбільш інтенсивне зростання цін, що свідчить про активний розвиток ринку та зростання попиту на нерухомість. Водночас збільшується і варіативність значень, що ускладнює процес прогнозування та вимагає використання більш складних моделей аналізу (рис. 3.1).

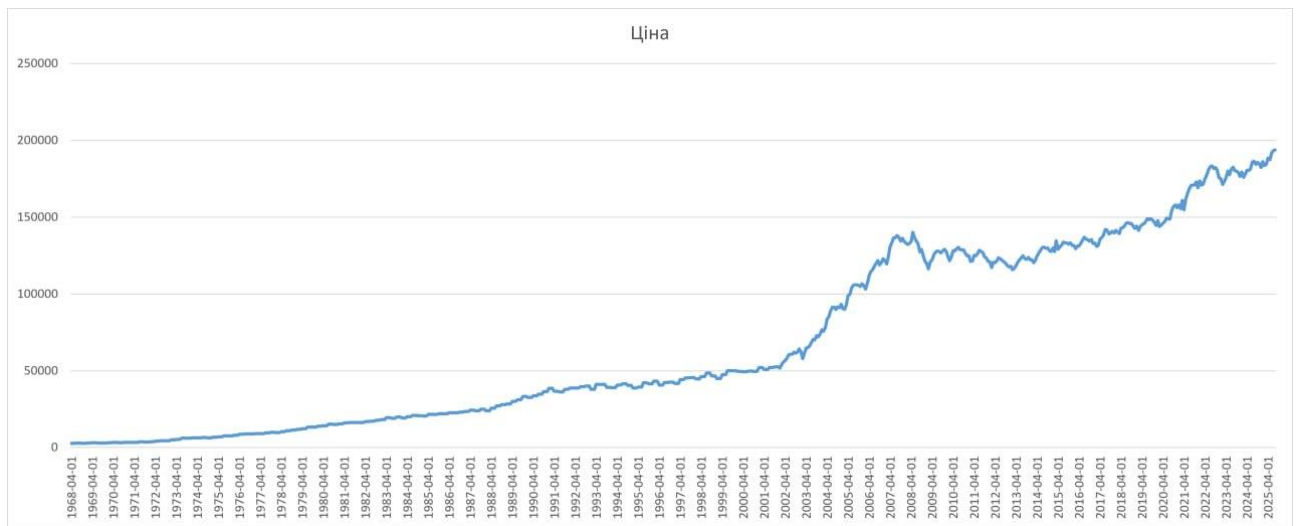


Рис. 3.1. Аналіз динаміки цін у фунтах на нерухомість в Шотланії (1968-2025)

На рис. 3.1 представлено динаміку зміни цін на нерухомість за період 1968–2025 рр.

Основні тенденції

Аналіз графіка дозволяє виділити кілька ключових етапів розвитку ринку:

1. 1968–1990 рр. – період повільного зростання
 - Спостерігається стабільне, поступове підвищення цін.
 - Ринок характеризується відносною рівновагою між попитом і пропозицією.
2. 1990–2003 рр. – помірне зростання
 - Темпи зростання збільшуються.
 - Формується стійкий висхідний тренд.
3. 2003–2008 рр. – різкий підйом (ціновий бум)

- Відбувається стрімке зростання цін.
- Можливі причини: активізація кредитування, інвестиційний попит.
- 4. 2008–2012 рр. – спад і нестабільність
- Чітко видно падіння після піку.
- Пов'язано з глобальною фінансовою кризою.
- 5. 2012–2019 рр. – відновлення ринку
- Поступове відновлення і стабілізація.
- Ціни знову демонструють висхідний тренд.
- 6. 2020–2025 рр. – новий етап зростання
- Спостерігається значне підвищення цін.
- Можливий вплив інфляції, змін у попиті та економічної нестабільності.

На графіку видно кілька періодів нестабільності:

- 2008–2010 рр. – різке падіння або сповільнення росту (світова фінансова криза);
- 2020 р. – коливання, пов'язані з пандемією;
- після 2022 р. – нестабільність із подальшим відновленням росту.

Після 2000 року значно зростає варіативність цін, що означає:

- більшу залежність від економічних факторів;
- підвищений ризик для прогнозування;
- необхідність використання складніших моделей (LSTM, XGBoost).

Досліджуваний часовий ряд має нелінійний характер, включає тренд і випадкові коливання, що підтверджує доцільність застосування як статистичних, так і сучасних методів машинного навчання для прогнозування.

3.2 Програмна реалізація моделей у середовищі Python

Для реалізації моделей прогнозування цін на нерухомість було використано мову програмування Python, яка є однією з найпоширеніших у сфері аналізу даних та машинного навчання. Основною перевагою Python є

наявність великої кількості бібліотек для роботи з часовими рядами, статистикою та нейронними мережами.

У процесі дослідження було застосовано наступні бібліотеки:

pandas – для обробки та аналізу даних;

numpy – для математичних операцій;

matplotlib – для побудови графіків;

scikit-learn – для оцінки якості моделей;

statsmodels – для реалізації моделі ARIMA;

xgboost – для побудови моделі градієнтного бустингу;

tensorflow / keras – для створення нейронної мережі LSTM.

На першому етапі завантажуюмо дані:

```
import pandas as pd
# Завантаження даних
df = pd.read_csv("data.csv")
# Перетворення дати
df['Date'] = pd.to_datetime(df['Date'])
# Сортвання
df = df.sort_values('Date')
# Встановлення індексу
df.set_index('Date', inplace=True)
# Вибір цільової змінної
series = df['AveragePrice']
```

На другому етапі за допомогою Python робимо побудову моделі ARIMA:

```
from statsmodels.tsa.arima.model import ARIMA
# Створення та навчання моделі
model_arima = ARIMA(series, order=(1,1,1))
model_arima_fit = model_arima.fit()
# Прогноз
forecast_arima = model_arima_fit.forecast(steps=12)
```

На третьому етапі за допомогою коду у Python робимо оцінку якості прогнозу:

```
from sklearn.metrics import mean_absolute_error, mean_squared_error
mae = mean_absolute_error(y, forecast_xgb)
rmse = mean_squared_error(y, forecast_xgb, squared=False)
print("MAE:", mae)
print("RMSE:", rmse)
```

Отримані за допомогою Python результати підтверджують доцільність використання сучасних інструментів аналізу даних для прогнозування цін на нерухомість.

3.3 Побудова моделі прогнозування цін на нерухомість

У дослідженні взяли участь статистичні моделі ARIMA, SARIMA, метод Хольта-Вінтерса, Prophet для кожного ряду, моделі машинного навчання лінійна регресія, градієнтний бустинг (Gradient Boosting – метод ансамблевого машинного навчання), до-найближчих сусідів, бегінг (begging – жебрацтво, випрошування коштів), включаючи моделі нейронних мереж архітектури LSTM, BiLSTM та TimesNet. У дослідженні моделей часових рядів із використанням статистичних моделей було представлено загальний середній результат по метрикам для зведеної таблиці. Оцінка моделей включає обчислення MSE, MAE, R^2 та MAPE. Структурна схема системного підходу до вирішення задач моделювання та прогнозування в задачах машинного навчання представлена на рисунку 3.2.

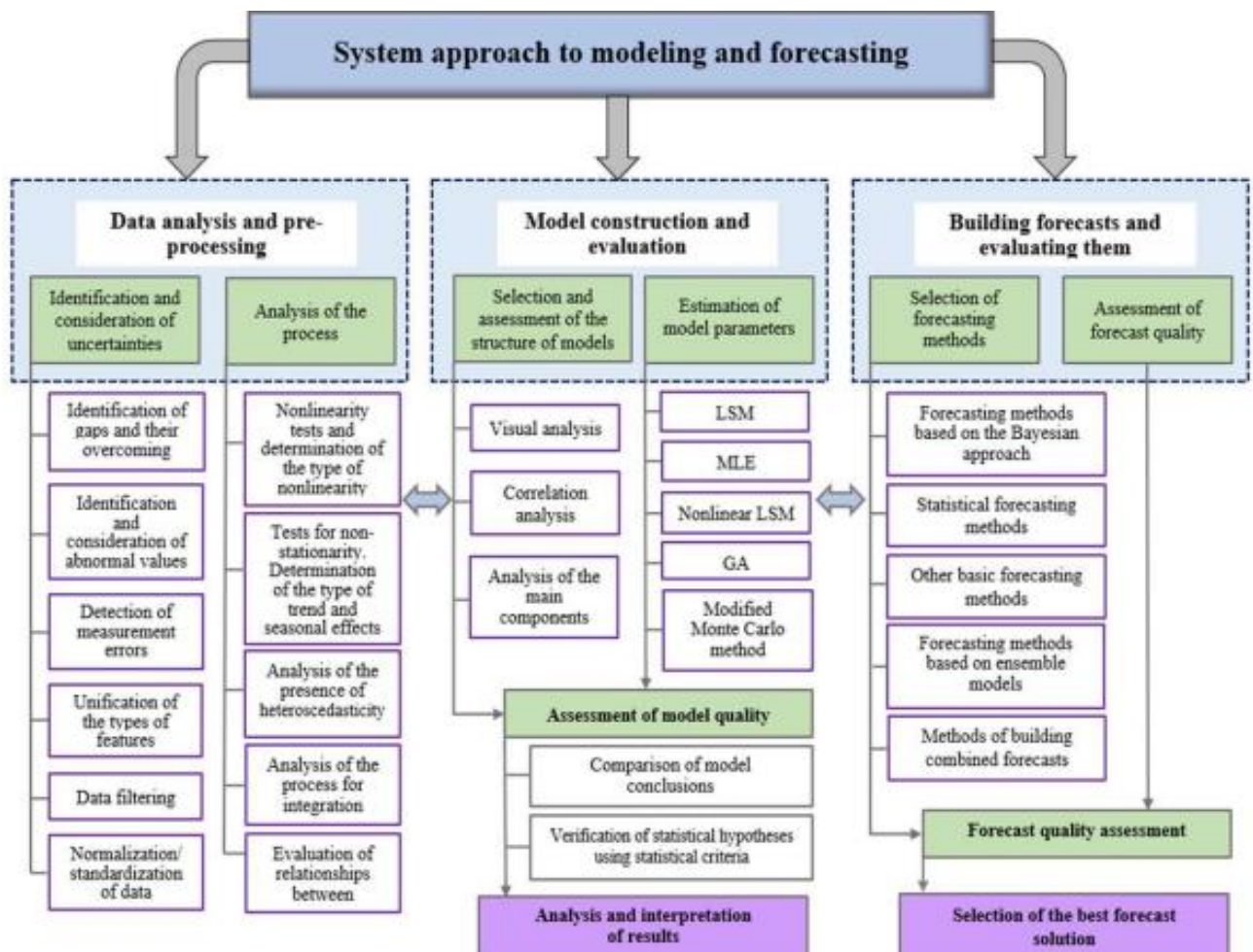


Рис. 3.2. Структура системного підходу до моделювання та прогнозування)[32]

Проведемо побудову моделі, використовуючи різні алгоритми машинного навчання, та проаналізуємо їхню точність.

Багатовимірна лінійна регресія

Математична постановка багатовимірної лінійної регресії:

$$price = \sum_1^N \beta_i X_i + \varepsilon$$

де *price* – ціна за квадратний метр;

β_i – Коефіцієнт регресії;

x_i – значення і-ого атрибуту;

ε – випадкова змінна, що відповідає випадковій, непрогнозованій помилці моделі;

N – кількість атрибутів у наборі даних.

Часовий ряд цін на нерухомість можна описати як функцію попередніх значень та випадкової складової:

$$y_t = f(y_{t-1}, y_{t-2}, \dots, \varepsilon_t) \quad (3.1)$$

де

y_t – значення ціни у момент часу t ;

ε_t – випадкова похибка.

Перед побудовою моделей було виконано:

- аналіз стаціонарності ряду;
- застосування операції диференціювання;

Ці етапи дозволили визначити параметри моделей прогнозування. Імпортуємо клас `LinearRegression`, створюємо його екземпляр і викликаємо метод `fit` з обліком нашого навчального набору даних. Заключним кроком є оцінка продуктивності алгоритму. Зробимо розрахунок середньоквадратичної помилки та коефіцієнта детермінації: $RMSE = 230066,024992$ та $R^2 = 0,3556$.

У разі оцінки передбачення значень за допомогою багатовимірної лінійної регресії коефіцієнт детермінації близький до 0, що вказує на погану якість апроксимації даних, що спостерігаються побудованою моделлю. Регресія пояснює 35,56% розкиду середнього значення цільової ознаки. Ця модель зовсім не придатна для використання.

3.4 Реалізація прогнозу за допомогою моделі ARIMA

Для моделі ARIMA застосовано 2 підходи, які дозволяють оцінити ефективність моделі. У першому підході кожен тимчасовий ряд аналізується на стаціонарність, і, якщо необхідно, виконується різницеве перетворення. Потім для даних застосовується декомпозиція, виділяючи тренд та сезонність, і будується модель ARIMA, параметри якої підбираються з мінімізацією AIC.

Модель ARIMA дозволяє враховувати залежність поточного значення від попередніх значень часового ряду. Інформація про якість та ефективність прогнозних моделей для топ-5 років представлено таблиці 1. MSE в 01.04.1969

високий, що вказує на наявність помилки в моделі. Низький R^2 вказує на те, що моделі пояснюють частину варіації тимчасових рядів, але не ідеально. Далі було використано логарифмування для набору даних у статистичних моделях, щоб стабілізувати дисперсію, так як дані мають ненормальний характер розподілу. Особливу увагу при виборі моделі треба звертати на визначення коефіцієнта як основної метрики.

Середньоквадратична помилка (MSE) варіюється від 11 000 до 858 000, що говорить про значну відмінність у точності передбачень між різними моделями. Так само і абсолютна помилка (MAE) коливається від 51 до 166 000, вказуючи на різний ступінь точності різних моделей прогнозування даних. Коефіцієнт детермінації (R^2), який оцінює частку варіації у залежній змінній, пояснену незалежними змінними, знаходиться в межах від 0,130 до 0,925. Показуючи значні відмінності у пояснювальній здатності моделей. Деякі значення MAPE досягають 383%, що може вказувати на значні проблеми у деяких моделях під час роботи з даними, схильними до помилок чи аномалій (рис. 3.3-3.11).

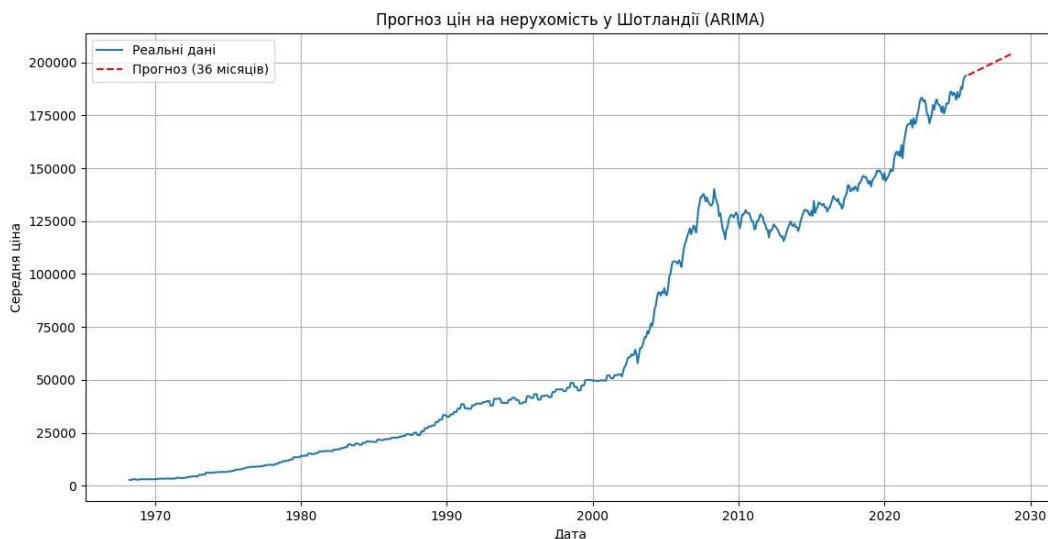


Рис. 3.3. Модель ARIMA 36 m (1; 2; 1)

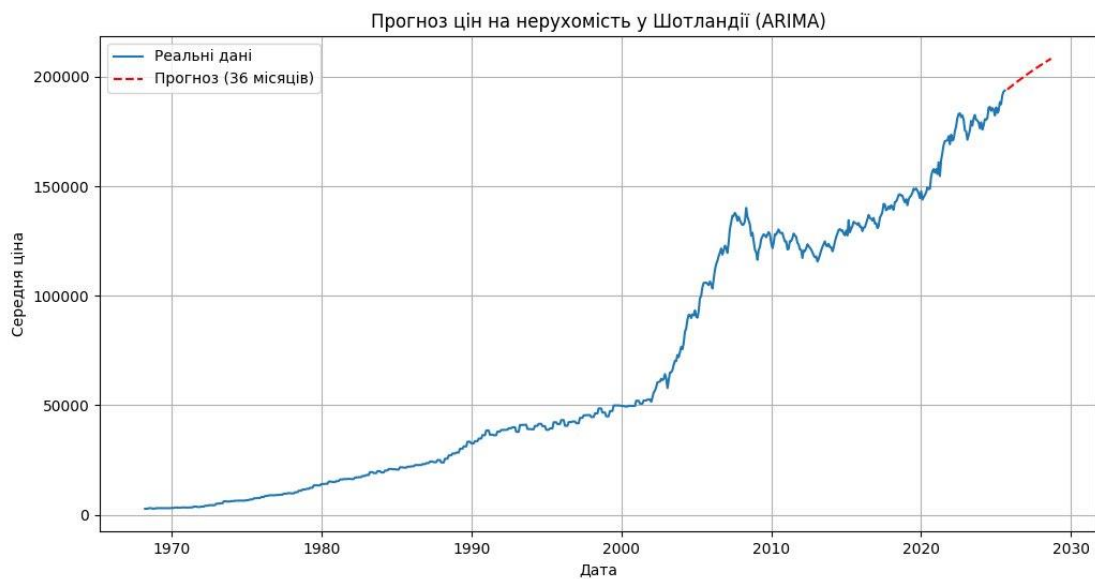


Рис. 3.4. Модель ARIMA 36 m (1; 1; 1)

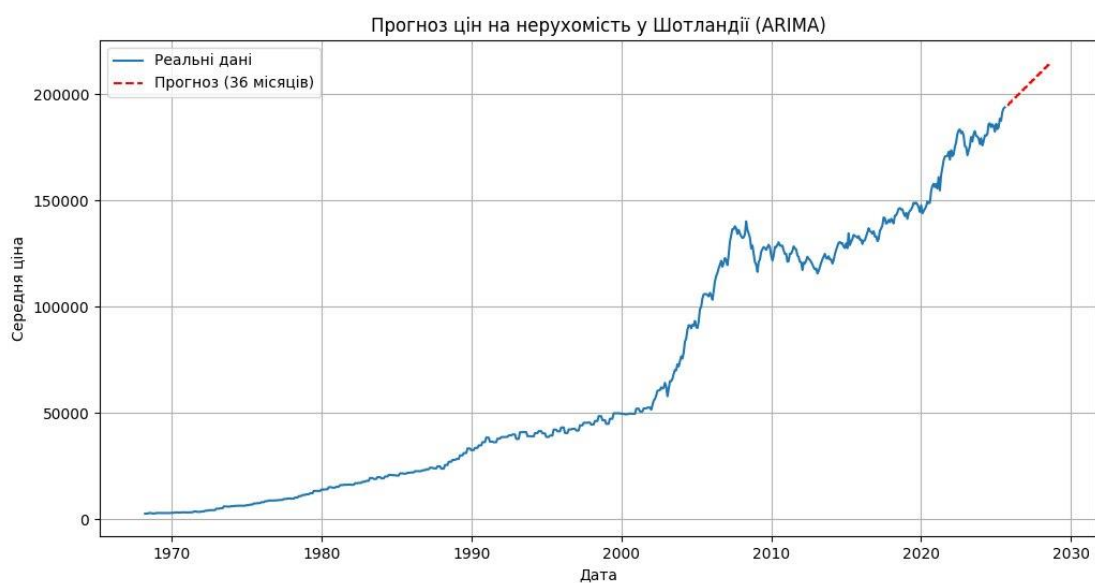


Рис. 3.5. Модель ARIMA 36 m (1; 2; 2)



Рис. 3.6. Модель ARIMA 60 m (1; 1; 1)

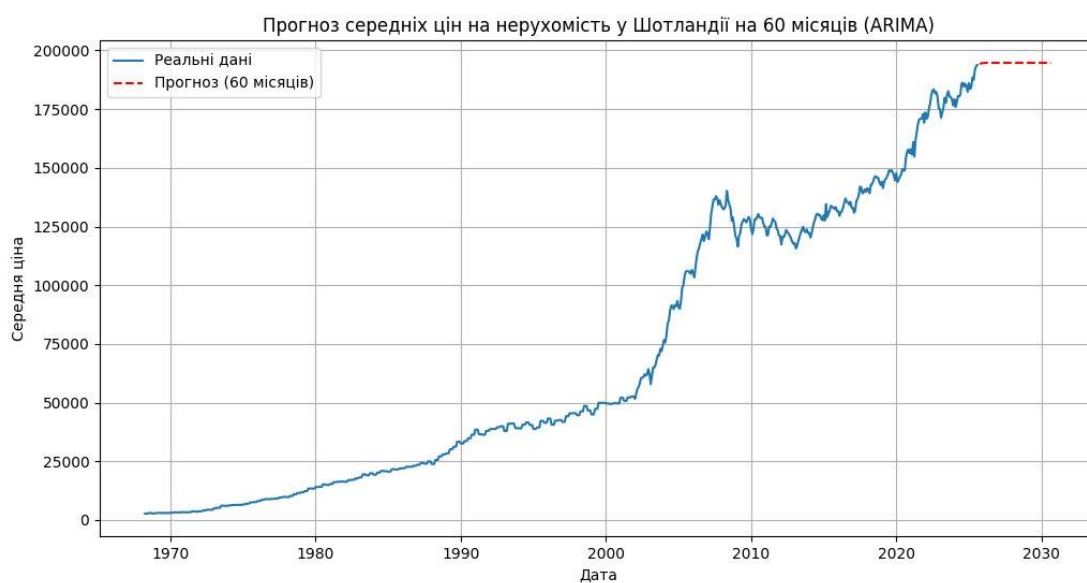


Рис. 3.7. Модель ARIMA 60 m (2; 1; 1)



Рис. 3.8. Модель ARIMA 60 m (2; 2; 1)

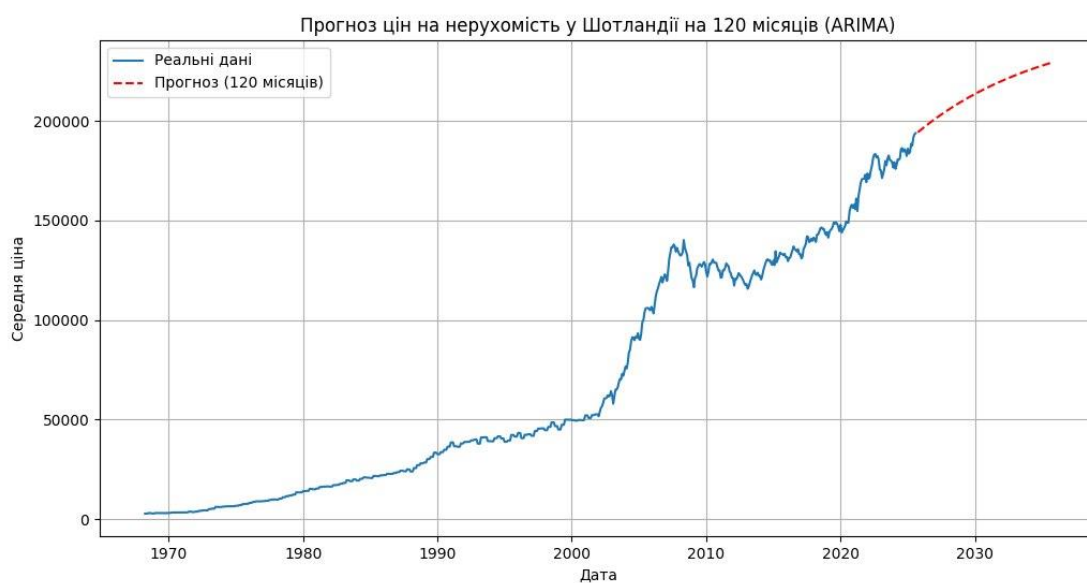


Рис. 3.9. Модель ARIMA 120 m (1; 1; 1)

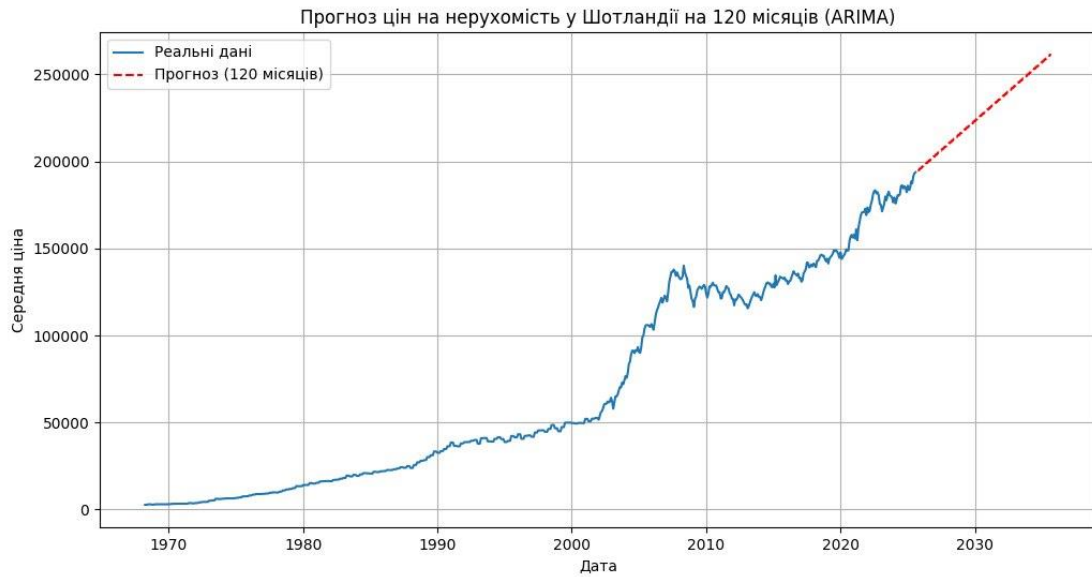


Рис. 3.10. Модель ARIMA 120 m (2; 2; 2)

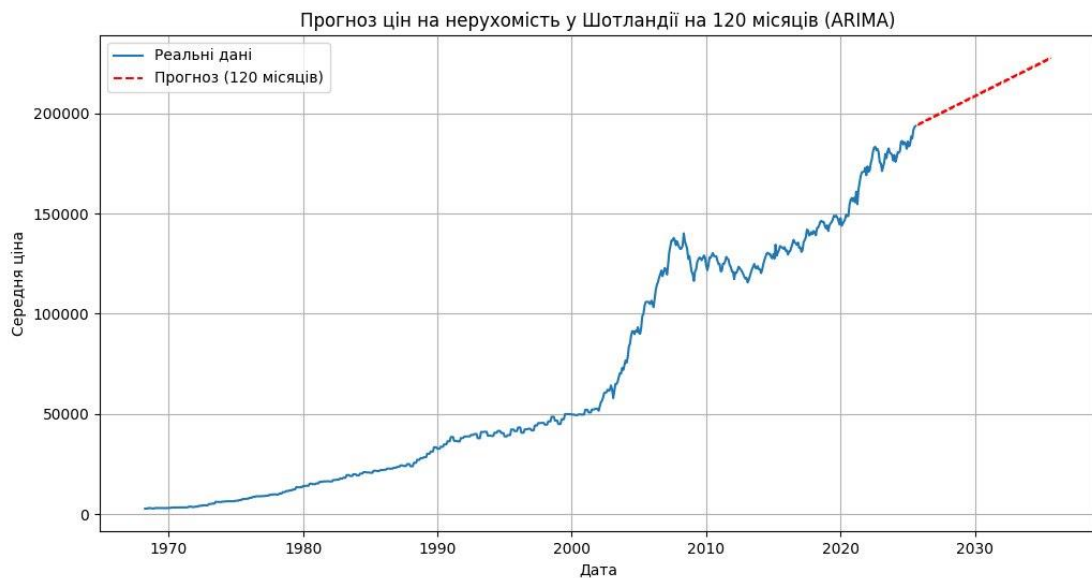


Рис. 3.11. Модель ARIMA 120 m (1; 2; 3)

На рисунках 3.3–3.11 представлено результати прогнозування часових рядів за допомогою моделей ARIMA з різними параметрами (p , d , q) та різною довжиною вибірки (36 m, 60 m, 120 m). Перш за все, слід відзначити, що зміна параметрів моделі істотно впливає на характер прогнозної кривої. Для моделей із параметром диференціювання $d = 2$ (рис. 3.3, 3.5, 3.7, 3.10, 3.11)

спостерігається більш різка зміна траєкторії прогнозу, що проявляється у підвищеній варіативності та меншій стабільності результатів. Це може свідчити про переускладнення моделі та її надмірну чутливість до випадкових коливань у вихідних даних.

Натомість моделі з параметром $d = 1$ (рис. 3.4, 3.6, 3.8, 3.9) демонструють більш згладжену та стабільну динаміку прогнозу. У цих випадках простежується або плавний тренд, або поступова стабілізація значень, що свідчить про більш адекватне відображення основної тенденції часового ряду.

Аналіз моделей із різною довжиною вибірки показує, що:

- для 36 місяців характерна більша нестабільність прогнозів, оскільки обсяг даних є обмеженим, що знижує точність оцінки параметрів;
- для 60 місяців моделі демонструють більш збалансовану поведінку, де поєднується чутливість до змін і відносна стабільність;
- для 120 місяців прогнози є найбільш згладженими та логічно узгодженими, оскільки більша кількість спостережень дозволяє краще виявити загальну тенденцію ряду.

Окремо слід звернути увагу на моделі з більшими значеннями параметрів p та q (наприклад, (2;2;2) або (1;2;3)), які формують більш “гнучкі” криві. У деяких випадках це призводить до перенавчання моделі (overfitting), коли вона добре описує наявні дані, але формує менш надійний прогноз.

3.5 Реалізація прогнозу за допомогою моделі SARIMA

Для більш точного врахування сезонних коливань у часовому ряді цін на нерухомість було застосовано сезонну модель SARIMA (Seasonal ARIMA). На відміну від класичної ARIMA, дана модель дозволяє враховувати періодичні зміни, що є характерними для ринку нерухомості (рис. 3.12-3.21).

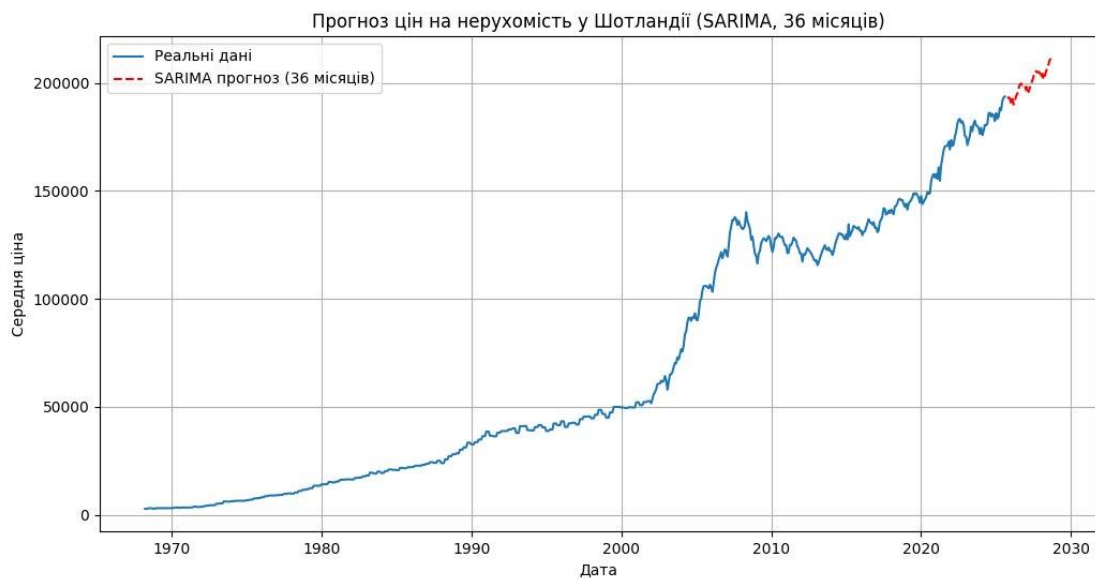


Рис. 3.12. Модель SARIMA 36 m (1; 1; 1) (1; 1; 1; 12)

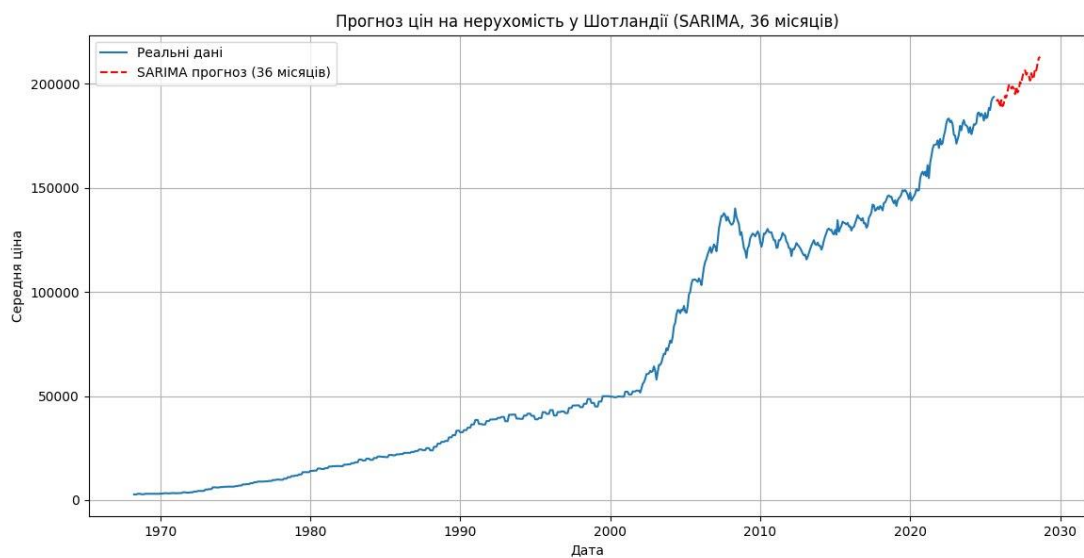


Рис. 3.13. Модель SARIMA 36 m (1; 1; 1) (1; 2; 1; 12)

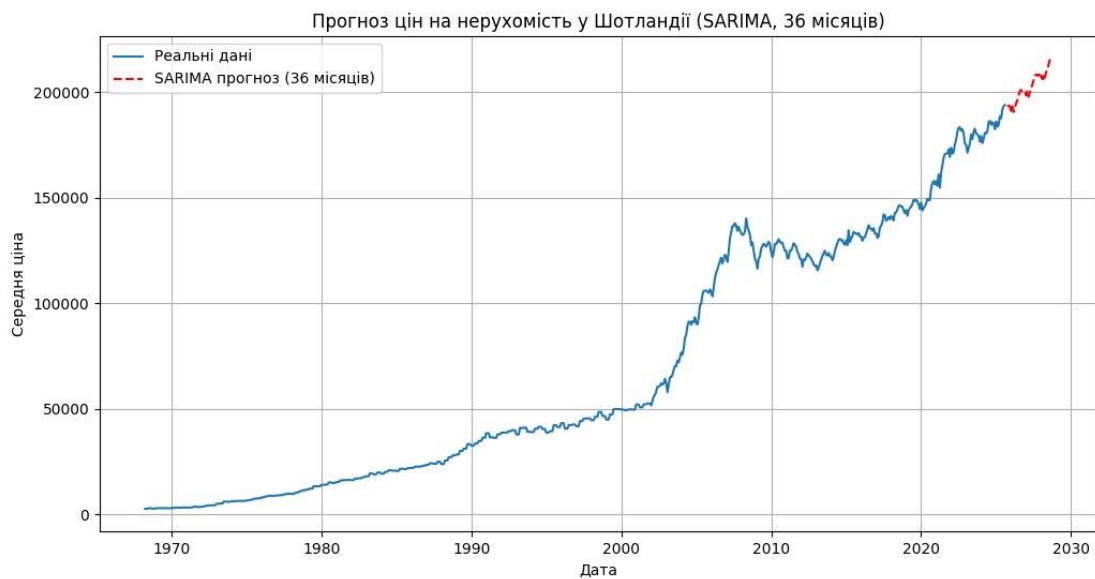


Рис. 3.14. Модель SARIMA 36 m (1; 2; 1) (1; 1; 1; 12)

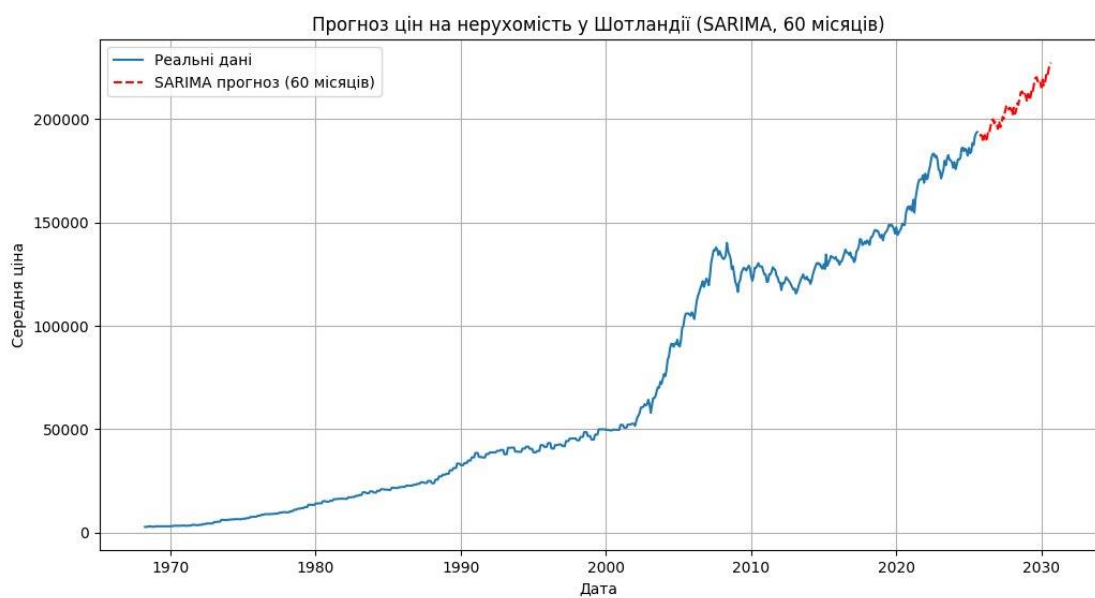


Рис. 3.15. Модель SARIMA 60 m (2; 1; 1) (1; 2; 1; 12)

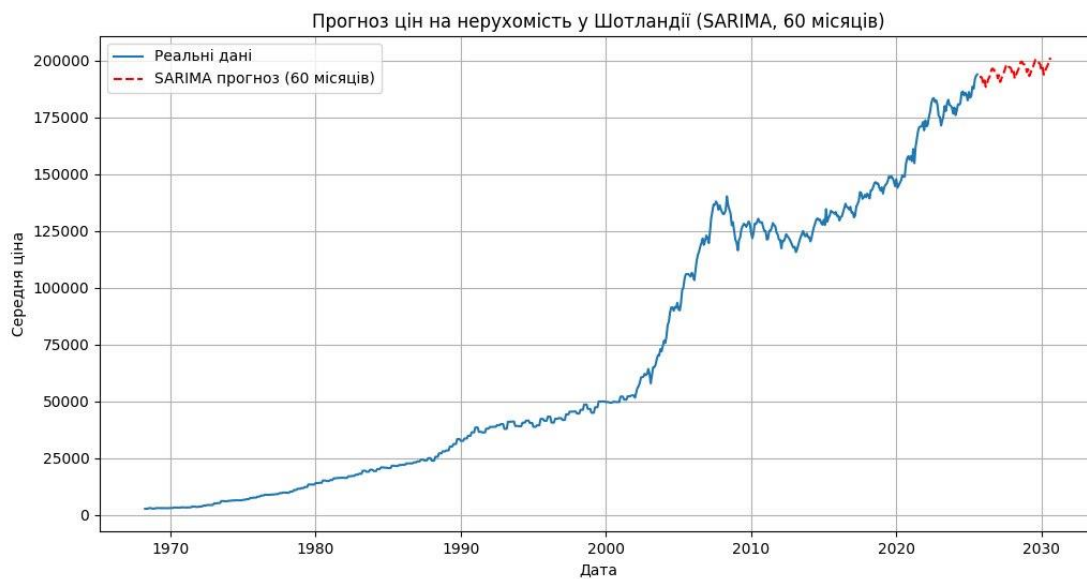


Рис. 3.16. Модель SARIMA 60 m (1; 3; 1) (1; 1; 1; 12)

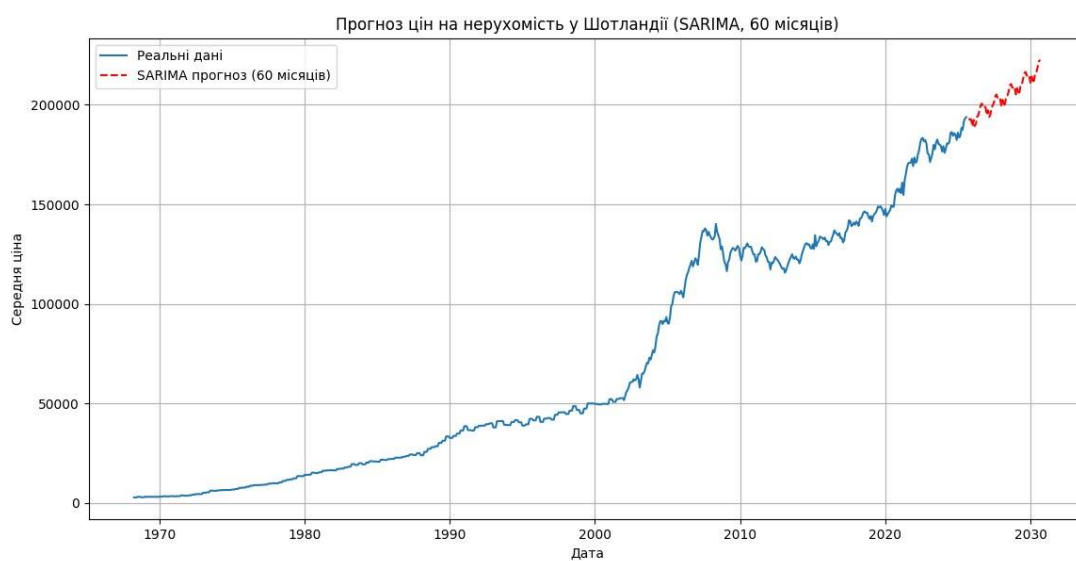


Рис. 3.17. Модель SARIMA 60 m (1; 1; 1) (3; 2; 1; 12)

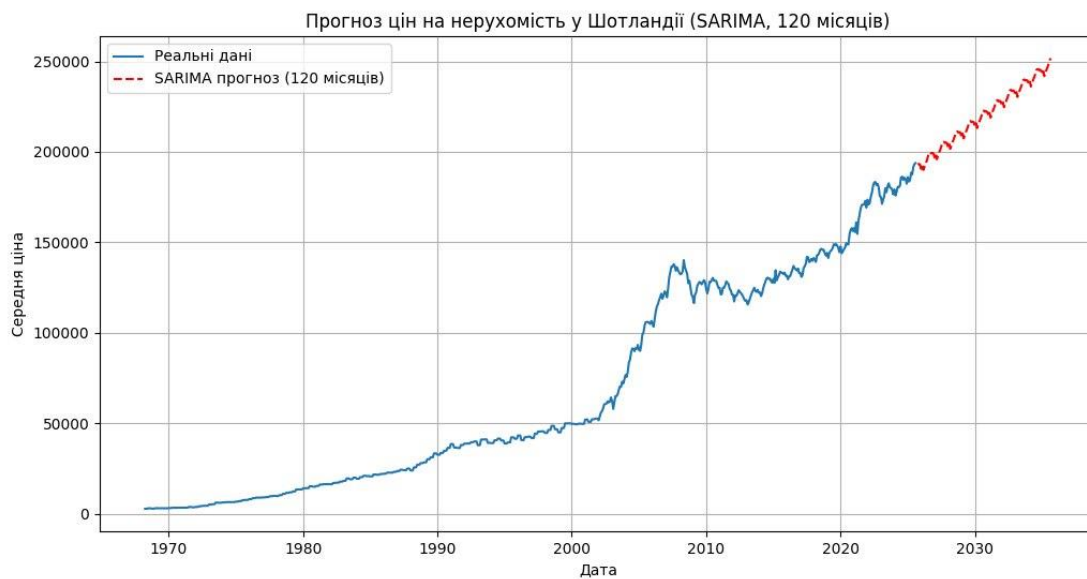


Рис. 3.18. Модель SARIMA 120 m (1; 1; 1) (1; 1; 1; 12)

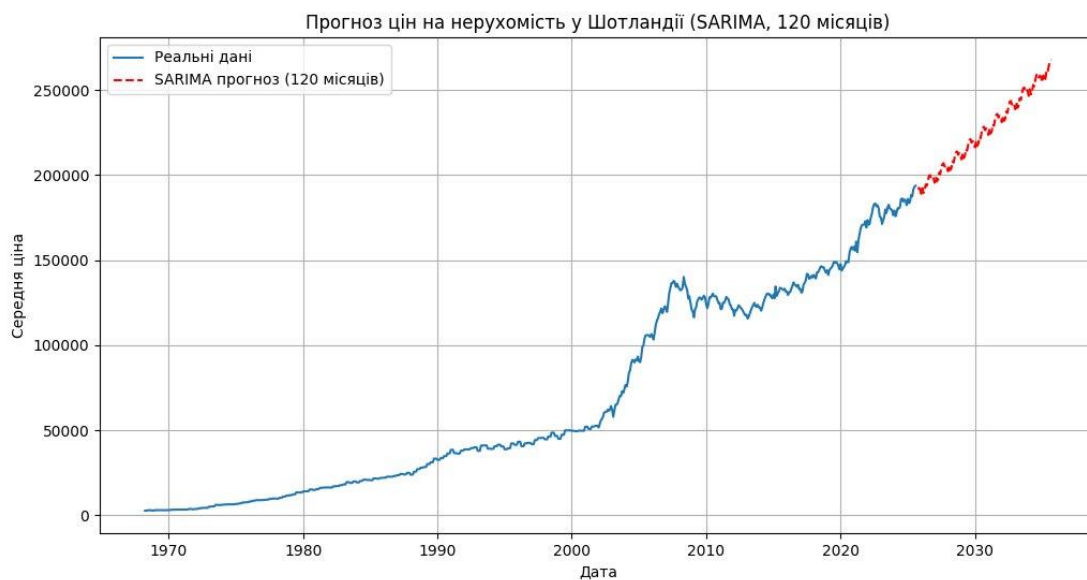


Рис. 3.19. Модель SARIMA 120 m (1; 2; 1) (1; 2; 1; 12)

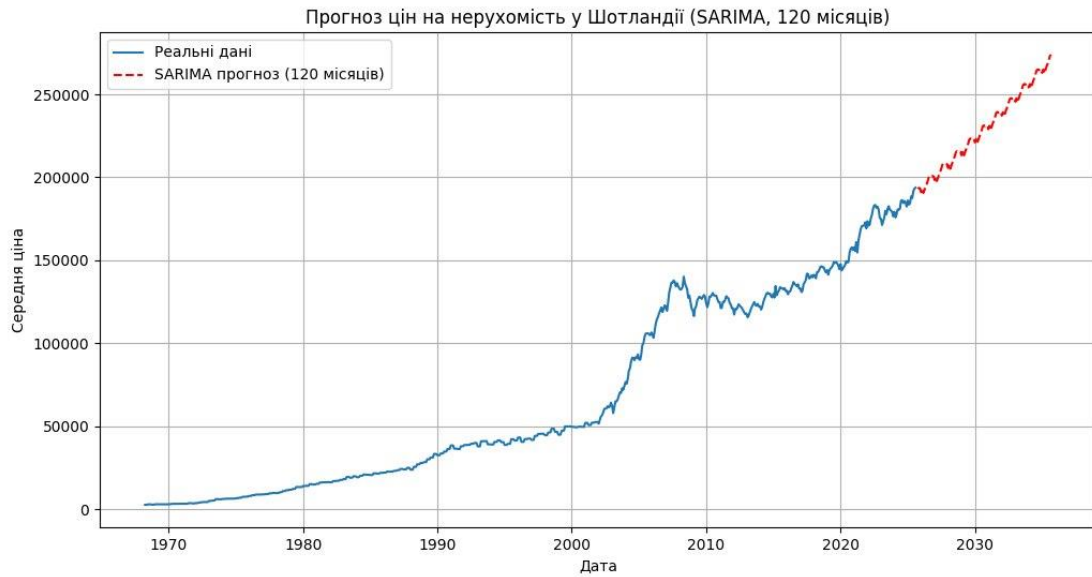


Рис. 3.20. Модель SARIMA 120 m (1; 2; 1) (1; 1; 3; 12)



Рис. 3.21. Модель SARIMA 120 m (1; 3; 1) (1; 2; 1; 12)

На рисунках 3.12–3.21 представлено результати прогнозування із використанням сезонних моделей SARIMA з різними комбінаціями параметрів (p, d, q) та сезонних складових $(P, D, Q, s=12)$ для часових інтервалів 36, 60 та 120 місяців. На відміну від моделей ARIMA, у даному випадку чітко простежується вплив сезонної компоненти, що проявляється у вигляді

періодичних коливань прогнозової кривої. Це підтверджує доцільність використання SARIMA для аналізу ринку нерухомості, де сезонність є одним із визначальних факторів.

Аналіз показує, що:

- моделі з сезонною складовою $(1;1;1;12)$ демонструють найбільш збалансовану поведінку, де поєднуються тренд і помірна сезонність;
- збільшення параметра сезонного диференціювання $D = 2$ призводить до надмірної варіативності та іноді до “переломів” у прогнозній кривій, що може свідчити про переускладнення моделі;
- підвищення порядків P та Q робить модель більш чутливою до сезонних коливань, однак у деяких випадках це призводить до нестабільності прогнозу.

Аналогічно до моделей ARIMA, довжина часового ряду має суттєвий вплив:

- 36 місяців – спостерігається недостатня стабільність сезонного патерну, що знижує надійність прогнозу;
- 60 місяців – моделі краще відтворюють сезонність, однак можуть бути чутливими до короткострокових коливань;
- 120 місяців – найбільш адекватне відображення як тренду, так і сезонної складової, що забезпечує більш узгоджений і логічний прогноз.

Отримані результати не є точним прогнозом цін на нерухомість, а відображають поведінку моделей SARIMA при різних налаштуваннях параметрів. Проведене дослідження має експериментальний характер і спрямоване на оцінку можливостей моделі щодо врахування сезонності. Таким чином, використання SARIMA є доцільним наступним кроком після ARIMA у процесі побудови більш точних прогнозних моделей.

3.6 Порівняння результатів моделей прогнозування

Для перевірки адекватності побудованої моделі ARIMA було здійснено розподіл даних на навчальну та тестову вибірки. Навчання моделі проводилось

на даних за період 1968–2020 рр., тоді як перевірка точності прогнозу здійснювалась на тестових даних за 2021–2025 рр. (рис. 3.22-3.23).

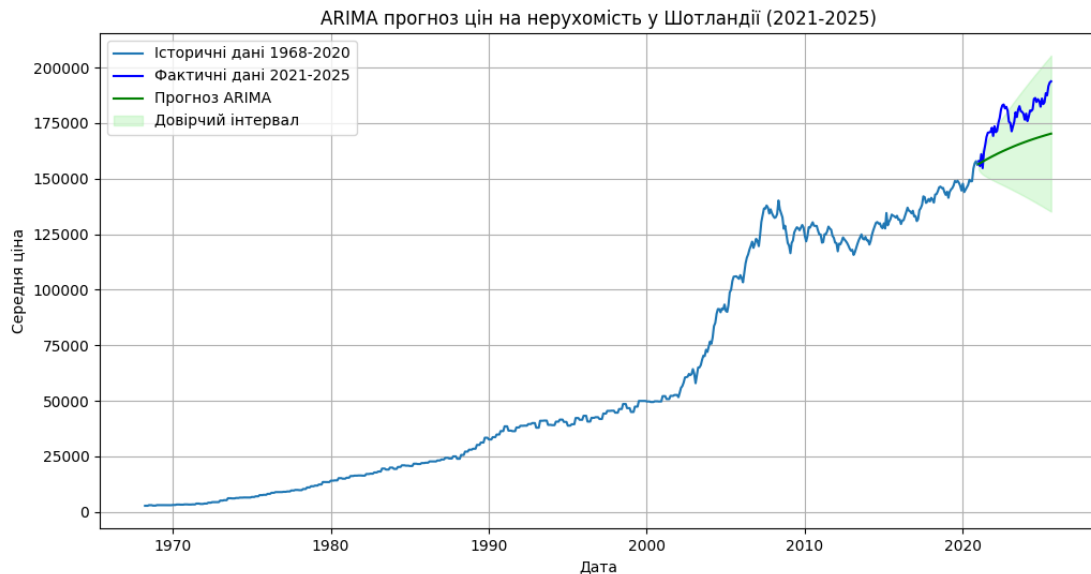


Рис. 3.22. ARIMA перевірка

Скрипт

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.arima.model import ARIMA
import numpy as np

# =====
# 1. Завантаження даних
# =====
data = pd.read_csv("data.csv") # твій файл 1968-2025
data['Date'] = pd.to_datetime(data['Date'])
data = data.set_index('Date')
# Перевірка
print("Перші 5 рядків:")
print(data.head())
print("Останні 5 рядків:")
print(data.tail())
# =====
# 2. Розділення на навчання та тест
# =====
train = data[:'2020-12-31'] # 1968-2020
test = data['2021-01-01':] # 2021-2025 для перевірки прогнозу
# =====
# 3. Навчання ARIMA
# =====
# order=(p,d,q) - параметри можна змінювати
model = ARIMA(train['AveragePrice'], order=(1,1,1))
```

```

results = model.fit()
print(results.summary())
# =====
# 4. Прогноз на 2021-2025
# =====
forecast_steps = len(test) # кількість місяців/років у тесті
forecast = results.get_forecast(steps=forecast_steps)
forecast_mean = forecast.predicted_mean
conf_int = forecast.conf_int()
# =====
# 5. Обчислення RMSE
# =====
rmse = np.sqrt(np.mean((test['AveragePrice'] - forecast_mean)**2))
print(f"RMSE прогнозу ARIMA на 2021-2025: {rmse:.2f}")
# =====
# 6. Візуалізація
# =====
plt.figure(figsize=(12,6))
plt.plot(train.index, train['AveragePrice'], label="Історичні дані 1968-2020")
plt.plot(test.index, test['AveragePrice'], label="Фактичні дані 2021-2025",
color="blue")
plt.plot(forecast_mean.index, forecast_mean, color="green", label="Прогноз ARIMA")
plt.fill_between(
    conf_int.index,
    conf_int.iloc[:,0],
    conf_int.iloc[:,1],
    color="lightgreen",
    alpha=0.3,
    label="Довірчий інтервал"
)
plt.title("ARIMA прогноз цін на нерухомість у Шотландії (2021-2025)")
plt.xlabel("Дата")
plt.ylabel("Середня ціна")
plt.legend()
plt.grid(True)
plt.show()

```

Для перевірки ефективності моделі SARIMA було здійснено розподіл даних на навчальну та тестову вибірки. Навчання моделі проводилось на історичних даних за період 1968–2020 рр., тоді як тестування – на даних 2021–2025 рр.

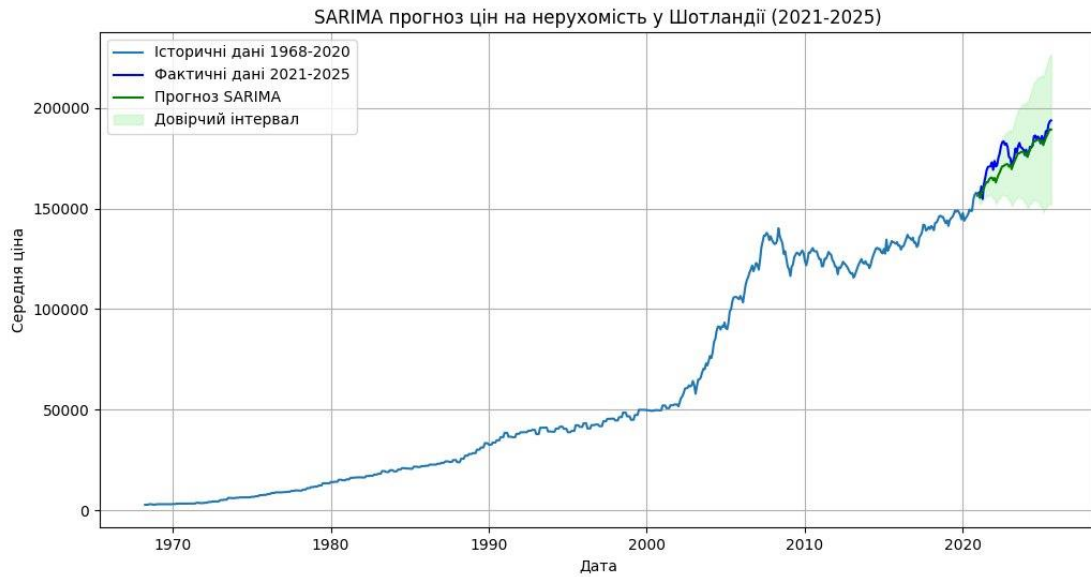


Рис. 3.23. SARIMA перевірка

Скрипт

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.statespace.sarimax import SARIMAX
from sklearn.metrics import mean_squared_error
import numpy as np
# =====
# 1. Завантаження даних
# =====
data = pd.read_csv("data.csv") # твій файл 1968-2025
data['Date'] = pd.to_datetime(data['Date'])
data = data.set_index('Date')
# Перевірка
print("Перші 5 рядків:")
print(data.head())
print("Останні 5 рядків:")
print(data.tail())
# =====
# 2. Розділення на навчання та тест
# =====
train = data[:'2020-12-31'] # 1968-2020
test = data['2021-01-01':] # 2021-2025 для перевірки прогнозу
# =====
# 3. Навчання SARIMA
# =====
model = SARIMAX(
    train['AveragePrice'],
    order=(1,1,1),
    seasonal_order=(1,1,1,12), # сезонність = 12 місяців
    enforce_stationarity=False,
    enforce_invertibility=False
)
```

```

results = model.fit()
print(results.summary())
# =====
# 4. Прогноз на 2021-2025
# =====
forecast_steps = len(test) # автоматично підбирає к-сть місяців у тесті
forecast = results.get_forecast(steps=forecast_steps)
forecast_mean = forecast.predicted_mean
conf_int = forecast.conf_int()
# =====
# 5. Обчислення RMSE
# =====
rmse = np.sqrt(mean_squared_error(test['AveragePrice'], forecast_mean))
print(f"RMSE прогнозу на 2021-2025: {rmse:.2f}")
# =====
# 6. Візуалізація
# =====
plt.figure(figsize=(12,6))
plt.plot(train.index, train['AveragePrice'], label="Історичні дані 1968-2020")
plt.plot(test.index, test['AveragePrice'], label="Фактичні дані 2021-2025", color="blue")
plt.plot(forecast_mean.index, forecast_mean, color="red", label="Прогноз SARIMA")
plt.fill_between(
    conf_int.index,
    conf_int.iloc[:,0],
    conf_int.iloc[:,1],
    color="pink",
    alpha=0.3,
    label="Довірчий інтервал"
)
plt.title("SARIMA прогноз цін на нерухомість у Шотландії (2021-2025)")
plt.xlabel("Дата")
plt.ylabel("Середня ціна")
plt.legend()
plt.grid(True)
plt.show()

```

За всіма трьома показниками похибки модель SARIMA продемонструвала вищу точність прогнозування порівняно з моделлю ARIMA (табл. 3.1)

Таблиця 3.1

Порівняння моделей ARIMA та SARIMA

Метрика	ARIMA	SARIMA	Краща модель
MAE	13407,54	4065,63	SARIMA
RMSE	14344,26	5276,67	SARIMA
MAPE, %	7,44	2,29	SARIMA

Найменші значення MAE, RMSE та MAPE свідчать про кращу здатність SARIMA відтворювати фактичну динаміку цін на нерухомість у Шотландії, що

підтверджує доцільність врахування сезонної складової при прогнозуванні часових рядів.

3.7 Практичне застосування отриманих результатів у сфері експертної оцінки нерухомості

У сукупності результати статистичної діагностики та специфікації дозволяють зробити висновок, що оптимальною з погляду прогнозної точності є базова лінійна модель OLS із застосуванням стандартних помилок (HC3), що коригують вплив на оцінку коефіцієнтів (рисунок 3.24). Ця модель забезпечує мінімальні значення помилок прогнозування ($RMSE \approx 775$ USD; $MAE \approx 621$ USD при $n = 223$) та високу якість апроксимації ($R^2 \approx 0.9997$).

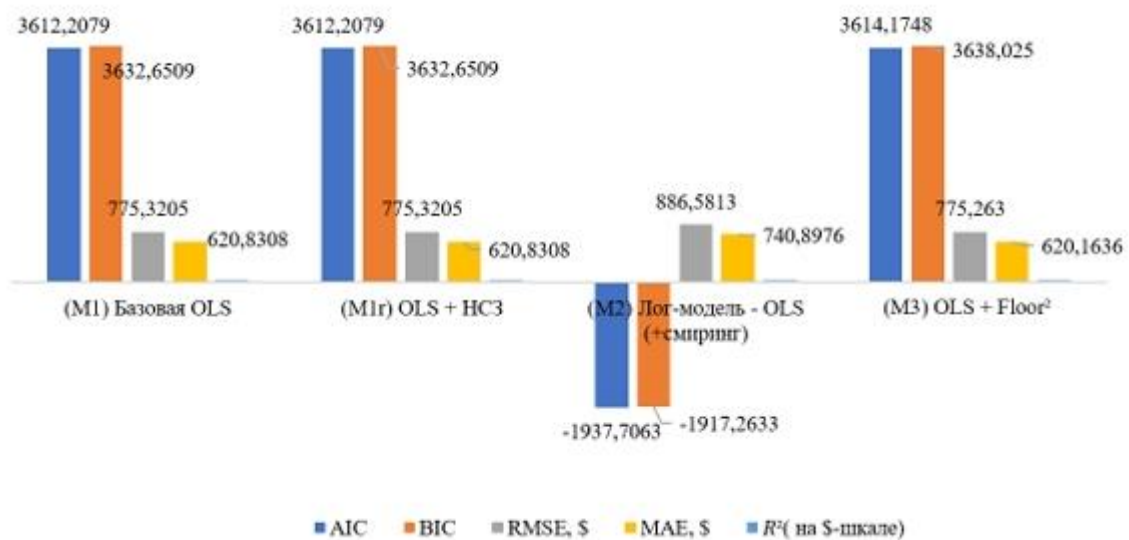


Рис. 3.24. Порівняльний аналіз моделей [33]

Інтерпретація коефіцієнтів також підтвердила очікувані ефекти структурних характеристик житла в Шотландії. Запропонована модель надійно пояснює та прогнозує ціни в межах об'єкта, що досліджується, надаючи практичні орієнтири для прийняття управлінських рішень. Однак для забезпечення широкої узагальнюваності та підвищення аналітичної цінності потрібне розширення бази даних просторовим та тимчасовим зрізам, а також включення додаткових факторів, відображають характеристики попиту та локалізаційних особливостей. Таким чином, інвестиції в нерухомість можуть

бути ефективними, але їх успішність залежить від багатьох факторів. Зростання вартості нерухомості, дохід від оренди, капіталізація доходу, управління ризиками та стабільність прибутковості – все це впливає ефективність вкладень. Крім того, податкові вигоди можуть покращити фінансові результати інвестицій.

Однак, слід пам'ятати, що інвестиції в нерухомість також несуть певні ризики, пов'язані з змінами на ринку, управлінням об'єктом та економічною ситуацією. Важливо провести ретельний аналіз, оцінити потенціал об'єкта та прогнозувати його прибутковість перед прийняттям рішення про інвестиції у нерухомість. Ринок нерухомості є організаційною системою, що об'єднує покупців та продавців для визначення цін та здійснення угод. Він також функціонує як сфера інвестицій та формує широкий спектр економічних відносин, пов'язаних з купівлею-продажем, іпотекою, орендою та іншими операціями. Унікальність кожного об'єкта нерухомості визначає локалізований характер даного ринку, де вартість визначається як внутрішніми характеристиками об'єкта, а й сукупністю зовнішніх факторів, включаючи політичну, економічну, соціальну та екологічну ситуацію в країні та регіоні.

Значення ринку нерухомості для національної економіки Шотландії важко переоцінити, оскільки нерухомість традиційно становить значну частину національного багатства. Крім того, стан даного ринку відображає загальні тенденції макроекономічного розвитку та впливає на фінансову стабільність, інвестиційну привабливість та рівень соціального добробуту. Проведений аналіз показав, що динаміка цін та доступність об'єктів на первинному ринку багато в чому залежать від поєднання макроекономічних умов, інституційного середовища, заходів державної політики та локальних факторів попиту. Включення державних програм та механізмів регулювання до системи аналізу підтверджує їх значну роль у забезпеченні стійкості та збалансованості розвитку будівельної галузі.

Висновки до розділу 3

Таким чином, результати дослідження дозволяють стверджувати, що ринок нерухомості не тільки виконує функцію механізму розподілу житла, а й виступає стратегічним чинником соціально-економічного розвитку. Подальші дослідження доцільно направити на розширення бази емпіричних даних, поглиблений аналіз просторових відмінностей, а також облік інституційних та соціокультурних чинників, які впливають функціонування ринку нерухомості. Застосування моделей ARIMA та SARIMA дозволяє підвищити точність прогнозів, зменшити невизначеність та покращити якість управлінських рішень.

ВИСНОВКИ

У ході дослідження було реалізовано модель прогнозування цін на нерухомість Шотландії. Поставлені цілі були досягнуті в такий спосіб. Для освоєння теми оцінки комерційної нерухомості було вивчено та проаналізовано безліч наукової література. Для розуміння актуальності застосування машинного навчання прогнозування вартості комерційної нерухомості було розглянуто статті, що використовують машинне навчання як методологію дослідження. Була дана характеристика ринку комерційної нерухомості та сформульована необхідність застосування методів машинного навчання.

Продемонстровано підходи до опису та побудови прогнозної моделі, а також розглянуто основні методи машинного навчання та завдання, які вони можуть вирішувати. Вивчено основні категорії помилок при розгляді моделей прогнозування та описані метрики для оцінки передбачуваної спроможності моделі. Сформульовано завдання для побудови моделей, а також способи побудови алгоритмів машинного навчання. В результаті аналізу інструментів побудови моделей була обрана найбільш підходяща для цього завдання мова програмування «Python».

Основним середовищем розробки було обрано ARIMA і SARIMA, оскільки ці середовища дозволяють здійснювати реалізацію моделі прогнозування, вимагаючи менших зусиль. Було сформовано вибірку даних шляхом заповнення пропущених змінних та підготовлена для побудови вибраними алгоритмами машинного навчання. Для подальшого навчання моделі було здійснено поділ набору даних на навчальний та тестовий набір відповідно. Проведено побудову моделі, використовуючи різні алгоритми машинного навчання та метод оцінки точності прогнозування (MAE, RMSE, MAPE). Проаналізовано та зіставлено їхню передбачувану точність.

Реалізація моделі підтвердила допустимість застосування методу машинного навчання для прогнозування цін комерційної нерухомості. Завдяки машинному навчанню вдалося вирішити завдання, у яких звичайні алгоритмічні рішення виявились би менш ефективними. За результатами тестування зроблено висновки про застосування алгоритму: застосування методу машинного навчання у зв'язку з оцінкою цін на комерційну нерухомість виявилось дуже ефективним, але потребує великих ресурсів та суворого підходу, орієнтованого на дані. Без правильного та глибокого розуміння даних основ структур було неможливо розробити адекватну прогнозну модель.

Зважаючи на все вищеперелічене можна підбити підсумок даної роботи. Усі цілі, поставлені перед початком дослідження можна вважати досягнутими, завдання виконаними, а результатом дослідження є побудована модель машинного навчання, за допомогою якої можна прогнозувати вартість комерційної нерухомості.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Challu C. et al. Nhits: Neural hierarchical interpolation for time series forecasting / C. Challu, Kin G. Olivares, B. N. Oreshkin, F. Garza // Proceedings of the AAAI Conference on Artificial Intelligence. – 2023. – Т. 37. – №. 6. – P. 6989-6997. DOI: <https://doi.org/10.1609/aaai.v37i6.25854>
2. Chen S. A. et al. Tsmixer: An all-mlp architecture for time series forecasting/ S. A. Chen // arXiv preprint arXiv:2303.06053. – 2023. – P.1-24. <https://doi.org/10.48550/arXiv.2303.06053>
3. Dai Y., Huang J. A sales prediction method based on lstm with hyperparameter search/ Y. Dai, J. Huang //Journal of Physics: Conference Series. IOP Publishing. –2021. – №. 1. – Т. 1756. – P. 012015. DOI 10.1088/1742-6596/1756/1/012015
4. Dairu X. Machine Learning Model for Sales Forecasting by Using XGBoost / X. Dairu, Z. Shilong // 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE). – 2021. – P. 480-483. DOI: 10.1109/ICCECE51280.2021.9342304
5. Das A. et al. Long-term forecasting with tide: Time-series dense encoder / A. Das // arXiv preprint arXiv:2304.08424. – 2023. – P. 1-21. <https://doi.org/10.48550/arXiv.2304.08424>
6. Gray, S., Hardley, S. & McQuillan, R. (2026). Leadership for wellbeing: How a Scottish secondary school embeds whole school health and wellbeing promotion. Health Education. <https://doi.org/10.1108/HE-06-2025-0093>
7. Global Commercial Real Estate Services. URL: <https://corees.pl/en/area-calculator/>
8. Global Real Estate Consultants // Knight Frank. URL: <https://www.knightfrank.com/>

9. Joseph R. V. A hybrid deep learning framework with CNN and Bidirectional LSTM for store item demand forecasting / R. V. Joseph // Computers and Electrical Engineering. –2022. –Т. 103. – P. 108358. <https://doi.org/10.1016/j.compeleceng.2022.108358>
10. Kishk, M.; and Al-Hajj, A., 2000, “Handling Linguistic Assessments In Life Cycle Costing - A Fuzzy Approach”, The Construction And Building Research Conference Of The RICS Research Foundation COBRA2000, The University of Greenwich, 30 August - September, 2000, pp. 228-243. <https://www.scribd.com/document/59043155/RM-Kishk-Revised>
11. Li X., Kang Y., Li F. Forecasting with time series imaging / Li X., Y. Kang, F. Li // Expert Systems with Applications. – 2020. – Т. 160. – P. 113680. <https://doi.org/10.1016/j.eswa.2020.113680>
12. Life cycle costing (LCC) as a contribution to sustainable construction: a common methodology [Electronic resource]. – Date of access: 12.03.2026. DOI:10.1088/1755-1315/1084/1/012023
13. Niu Y. Data Prediction Based on Support Vector Machine (SVM) — Taking Soil Quality Improvement Test Soil Organic Matter as an Example / Y. Niu, S. Ye // IOP Conference Series: Earth and Environmental Science. – 2019. –Vol. 295. № 2. – P. 012021. DOI 10.1088/1755-1315/295/2/012021
14. Phyu M. M. Retail Demand Forecasting Using Sequence to Sequence Long Short-Term Memory Networks / M.M. Phyu, M.T. Khine // 2023 IEEE Conference on Computer Applications (ICCA). –2023. – P. 208-213. DOI: 10.1109/ICCA51723.2023.10181450
15. Rasul K. et al. Lag-llama: Towards foundation models for time series forecasting/ K. Rasul // arXiv preprint arXiv:2310.08278. – 2023. – P. 1-23. <https://doi.org/10.48550/arXiv.2310.08278>
16. Reflections from MIPIM: Leveraging data, analytics, digital, and AI in real estate – McKinsey & Company. URL: <https://www.mckinsey.com/uk/our-insights/uk-blog/mipim-reflections>

17. Rental Market Tracker: Rents Rise 14% in December – Biggest Jump in Over Two Years // Redfin News. URL: <https://www.redfin.com/news/redfin-rental-report-december-2021/>

18. Singh K. E-commerce system for sale prediction using machine learning technique/ K. Singh, P. Booma, M., U. Eaganathan // Journal of Physics: Conference Series. – IOP Publishing, 2020. – T. 1712. – №. 1. – P. 012042. DOI 10.1088/1742-6596/1712/1/012042

19. Soaita, A. M., Searle, B. A., McKee, K., & Moore, T. (2016). Becoming a landlord: strategies of property-based welfare in the private rental sector in Great Britain. *Housing Studies*, 32(5), 613–637. <https://doi.org/10.1080/02673037.2016.1228855>

20. Stephens, M. (2020). How Housing Systems are Changing and Why: A Critique of Kemeny's Theory of Housing Regimes. *Housing, Theory and Society*, 37(5), 521–547. <https://doi.org/10.1080/14036096.2020.1814404>

21. Tang, S.; Shelden, D.R.; Eastman, C.M.; Pishdad-Bozorgi, P.; Gao, X. (2019). A Review of Building Information Modeling (BIM) and the Internet of Things (IoT) Devices Integration: Present Status and Future Trends. *Autom. Constr.*, 101, 127–139. <https://doi.org/10.1016/j.autcon.2019.01.020>

22. Thaler, R. H. (2016). Behavioral Economics: Past, Present, and Future. *American Economic Review*, 106(7), 1577–1600. DOI: 10.1257/aer.106.7.1577

23. Van Dessel, G. (2015). What's the best time to send a survey? CheckMarket. Retrieved 15/08/2020 from: <https://www.checkmarket.com/blog/survey-invitations-best-time-send/>

24. Scottish Government. (2021). Housing to 2040. Edinburgh: Scottish Government Retrieved from <https://www.gov.scot/publications/housing-2040-2/>.

25. Watson, Andrew Robert (2024) Canny Investors? A mixed methods investigation into the investment behaviours of Scottish private rented sector landlords. PhD thesis. <https://theses.gla.ac.uk/84408/>

26. Watson, A. R. (2021). The Pandemic Arrears Crisis: Private landlord perspectives on the cessation of the 'eviction ban'. CaCHE. Retrieved 15/12/2021

from: https://housingevidence.ac.uk/wp-content/uploads/2024/02/20211028_CaCHE_ThePandemicRentArrearsCrisis_FINA_L.pdf

27. Watson, A. R. (2025). 2025 UK Housing Review Autumn Briefing Paper: Reforming the private rented sector: compliance and enforcement challenges from Glasgow: <https://www.ukhousingreview.org.uk/ukhr25/docs/UKHR-Autumn-Briefing-2025.pdf>

28. Watts, B., & Stephenson, A. (2017). 'No DSS': A review of evidence on landlord and letting agent attitudes to tenants receiving Housing Benefit. Shelter Scotland. Retrieved 08/02/2021 from: https://scotland.shelter.org.uk/professional_resources/policy_library/no_dss_report

29. Wen Q. et al. Transformers in time series: A survey/ Q. Wen //arXiv preprint – 2022. – P. 6778-6786. <https://doi.org/10.48550/arXiv.2202.07125>

30. Wu H. et al. Timesnet: Temporal 2d-variation modeling for general time series analysis / H. Wu // The eleventh international conference on learning representations. – 2022. – P. 1-23. <https://doi.org/10.48550/arXiv.2210.02186>

31. Young, J. C., Rose, D. C., Mumby, H. S., Benitez- Capistros, F., Derrick, C. J., Finch, T., Mukherjee, N. (2018). A methodological guide to using and reporting on interviews in conservation science research. *Methods in Ecology and Evolution*, 9(1), 10–19. <https://doi.org/10.1111/2041-210X.12828>

32. Yue Z. et al. Ts2vec: Towards universal representation of time series/ Z. Yue, Z. Yue, Y. Wang, J. Duan, T. Yang //Proceedings of the AAAI Conference on Artificial Intelligence. – 2022. – T. 36. – №. 8. – P. 8980-8987. <https://doi.org/10.1609/aaai.v36i8.20881>

33. Zeng A. et al. Are transformers effective for time series forecasting? / A. Zeng, M. Chen, L. Zhang, Q. Xu // Proceedings of the AAAI conference on artificial intelligence. – 2023. – T. 37. – №. 9. – P. 11121-11128. <https://doi.org/10.1609/aaai.v37i9.26317>

34. Zhang Y. et al. Multi-resolution time-series transformer for long-term forecasting/ Y. Zhang // International Conference on Artificial Intelligence and Statistics. – PMLR, 2024. – P. 4222-4230. <https://proceedings.mlr.press/v238/zhang24l/zhang24l.pdf>

35. Zohdi M. Demand forecasting based machine learning algorithms on customer information: an applied approach / M. Zohdi // International Journal of Information Technology. – 2022. – Т.14. – № 4. – P. 1937-1947. <https://doi.org/10.1007/s41870-022-00875-3>

36. Застосування штучного інтелекту для прогнозування часового ряду руху полюсів землі: використання нейронної моделі LSTM / Лук`янченко Ю.О, Гуменний М.І, Лопушанський О.М., // матеріали XI Міжнародної науково-практичної конференції «Сучасні технології землеустрою, кадастру та управління земельними ресурсами».

ДОДАТКИ

Додаток А

Характеристика ринку нерухомості Шотландії



ARIMA модель

```

IDLE Shell 3.14.0
File Edit Shell Debug Options Window Help
Python 3.14.0 (tags/v3.14.0:ebf955d, Oct 7 2025, 10:15:03) [MSC v.1944 64 bit (AMD64)] on win32
Enter "help" below or click "Help" above for more information.
>>>
==== RESTART: C:\Users\User\OneDrive\Рабочий стол\real_estate_forecast\main.py ====
Перші 5 рядків:
      AveragePrice
Date
1968-04-01      2738
1968-05-01      2738
1968-06-01      2738
1968-07-01      3001
1968-08-01      3001
Останні 5 рядків:
      AveragePrice
Date
2025-04-01      188448
2025-05-01      187478
2025-06-01      191657
2025-07-01      193254
2025-08-01      193786

Warning (from warnings module):
  File "C:\Users\User\Hosna nanka\Lib\site-packages\statsmodels\tsa\base\tsa_model.py", line 473
    self._init_dates(dates, freq)
ValueWarning: No frequency information was provided, so inferred frequency MS will be used.

Warning (from warnings module):
  File "C:\Users\User\Hosna nanka\Lib\site-packages\statsmodels\tsa\base\tsa_model.py", line 473
    self._init_dates(dates, freq)
ValueWarning: No frequency information was provided, so inferred frequency MS will be used.

Warning (from warnings module):
  File "C:\Users\User\Hosna nanka\Lib\site-packages\statsmodels\tsa\base\tsa_model.py", line 473
    self._init_dates(dates, freq)
ValueWarning: No frequency information was provided, so inferred frequency MS will be used.

SARIMAX Results
=====
Dep. Variable:      AveragePrice      No. Observations:      633
Model:              ARIMA(1, 1, 1)    Log Likelihood         -5468.392
Date:              Tue, 27 Jan 2026   AIC                    10942.784
Time:              14:53:56           BIC                    10956.131
Sample:            - 12-01-2020
=====

```

```

IDLE Shell 3.14.0
File Edit Shell Debug Options Window Help

Warning (from warnings module):
  File "C:\Users\User\Hosna nanka\Lib\site-packages\statsmodels\tsa\base\tsa_model.py", line 473
    self._init_dates(dates, freq)
ValueWarning: No frequency information was provided, so inferred frequency MS will be used.

Warning (from warnings module):
  File "C:\Users\User\Hosna nanka\Lib\site-packages\statsmodels\tsa\base\tsa_model.py", line 473
    self._init_dates(dates, freq)
ValueWarning: No frequency information was provided, so inferred frequency MS will be used.

Warning (from warnings module):
  File "C:\Users\User\Hosna nanka\Lib\site-packages\statsmodels\tsa\base\tsa_model.py", line 473
    self._init_dates(dates, freq)
ValueWarning: No frequency information was provided, so inferred frequency MS will be used.

SARIMAX Results
=====
Dep. Variable:      AveragePrice      No. Observations:      633
Model:              ARIMA(1, 1, 1)    Log Likelihood         -5468.392
Date:              Tue, 27 Jan 2026   AIC                    10942.784
Time:              14:53:56           BIC                    10956.131
Sample:            - 12-01-2020           HQIC                    10947.968
=====
Covariance Type:    opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ar.L1          0.9862        0.007    149.425      0.000      0.973      0.999
ma.L1         -0.9545        0.012   -80.468      0.000     -0.978     -0.931
sigma2        1.918e+06    6.17e+04    31.085      0.000    1.8e+06    2.04e+06
=====
Ljung-Box (L1) (Q):      0.90    Jarque-Bera (JB):      453.05
Prob(Q):                0.34    Prob(JB):              0.00
Heteroskedasticity (H):  74.05    Skew:                  0.26
Prob(H) (two-sided):    0.00    Kurtosis:              7.11
=====
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
RMSE прогнозу ARIMA на 2021-2025: 14344.26

```

SARIMA модель

```

IDLE Shell 3.14.0*
File Edit Shell Debug Options Window Help

1968-05-01      2738
1968-06-01      2738
1968-07-01     3001
1968-08-01     3001

Warning (from warnings module):
  File "C:\Users\User\Homa nanka\Lib\site-packages\statsmodels\tsa\base\tsa_model.py", line 473
    self._init_dates(dates, freq)
ValueWarning: No frequency information was provided, so inferred frequency MS will be used.

Warning (from warnings module):
  File "C:\Users\User\Homa nanka\Lib\site-packages\statsmodels\tsa\base\tsa_model.py", line 473
    self._init_dates(dates, freq)
ValueWarning: No frequency information was provided, so inferred frequency MS will be used.

=====
SARIMAX Results
=====
Dep. Variable:          AveragePrice      No. Observations:      689
Model:              SARIMAX(1, 1, 1)x(1, 1, 1, 12)    Log Likelihood      -5697.833
Date:                Tue, 27 Jan 2026              AIC              11405.666
Time:                14:10:21                      BIC              11428.142
Sample:              04-01-1968                    HQIC              11414.377
                  - 08-01-2025

Covariance Type:      opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ar.L1          -0.2554      0.178      -1.437      0.151      -0.604      0.093
ma.L1           0.1177      0.185       0.636      0.524      -0.245      0.480
ar.S.L12        0.1582      0.038       4.119      0.000       0.083      0.234
ma.S.L12       -0.8523      0.020     -41.655      0.000     -0.892     -0.812
sigma2         1.74e+06      5.19e+04      33.554      0.000      1.64e+06      1.84e+06
=====
Ljung-Box (L1) (Q):      0.00      Jarque-Bera (JB):      973.34
Prob(Q):      0.95      Prob(JB):      0.00
Heteroskedasticity (H):      56.57      Skew:      -0.47
Prob(H) (two-sided):      0.00      Kurtosis:      8.87
=====

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

```

SARIMA модель перевірка

```

IDLE Shell 3.14.0
File Edit Shell Debug Options Window Help

Перші 5 рядків даних:
AveragePrice
Date
1968-04-01      2738
1968-05-01      2738
1968-06-01      2738
1968-07-01      3001
1968-08-01      3001

Warning (from warnings module):
  File "C:\Users\User\Hosha namka\Lib\site-packages\statsmodels\tsa\base\tsa_model.py", line 473
    self._init_dates(dates, freq)
ValueWarning: No frequency information was provided, so inferred frequency MS will be used.

Warning (from warnings module):
  File "C:\Users\User\Hosha namka\Lib\site-packages\statsmodels\tsa\base\tsa_model.py", line 473
    self._init_dates(dates, freq)
ValueWarning: No frequency information was provided, so inferred frequency MS will be used.

SARIMAX Results
=====
Dep. Variable:          AveragePrice      No. Observations:      689
Model:              SARIMAX(1, 1, 1)x(1, 1, 1, 12)      Log Likelihood      -5697.833
Date:                Tue, 27 Jan 2026      AIC                  11405.666
Time:                14:10:21      BIC                  11428.142
Sample:              04-01-1968      HQIC                 11414.377
                   - 08-01-2025

Covariance Type:      opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ar.L1          -0.2554      0.178      -1.437      0.151      -0.604      0.093
ma.L1           0.1177      0.185       0.636      0.524      -0.245      0.480
ar.S.L12        0.1582      0.038       4.119      0.000      0.083      0.234
ma.S.L12       -0.8523      0.020     -41.655      0.000     -0.892     -0.812
sigma2         1.74e+06      5.19e+04     33.554      0.000     1.64e+06     1.84e+06
=====
Ljung-Box (L1) (Q):      0.00      Jarque-Bera (JB):      973.34
Prob(Q):                0.95      Prob(JB):              0.00
Heteroskedasticity (H):  56.57      Skew:                 -0.47
Prob(H) (two-sided):    0.00      Kurtosis:              8.87
=====

Ln: 15 Col: 0

```

```

IDLE Shell 3.14.0
File Edit Shell Debug Options Window Help

=====
Date:                Tue, 27 Jan 2026      AIC                  11405.666
Time:                14:10:21      BIC                  11428.142
Sample:              04-01-1968      HQIC                 11414.377
                   - 08-01-2025

Covariance Type:      opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ar.L1          -0.2554      0.178      -1.437      0.151      -0.604      0.093
ma.L1           0.1177      0.185       0.636      0.524      -0.245      0.480
ar.S.L12        0.1582      0.038       4.119      0.000      0.083      0.234
ma.S.L12       -0.8523      0.020     -41.655      0.000     -0.892     -0.812
sigma2         1.74e+06      5.19e+04     33.554      0.000     1.64e+06     1.84e+06
=====
Ljung-Box (L1) (Q):      0.00      Jarque-Bera (JB):      973.34
Prob(Q):                0.95      Prob(JB):              0.00
Heteroskedasticity (H):  56.57      Skew:                 -0.47
Prob(H) (two-sided):    0.00      Kurtosis:              8.87
=====

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

= RESTART: C:\Users\User\OneDrive\Рабочий стол\real_estate_forecast\sarima_model.py
Перші 5 рядків:
AveragePrice
Date
1968-04-01      2738
1968-05-01      2738
1968-06-01      2738
1968-07-01      3001
1968-08-01      3001
Останні 5 рядків:
AveragePrice
Date
2025-04-01      188448
2025-05-01      187478
2025-06-01      191657
2025-07-01      193254
2025-08-01      193786

Ln: 48 Col: 0

```

```

IDLE Shell 3.14.0
File Edit Shell Debug Options Window Help

2025-06-01      191657
2025-07-01      193254
2025-08-01      193786

Warning (from warnings module):
  File "C:\Users\User\Homa nanka\Lib\site-packages\statsmodels\tsa\base\tsa_model.py", line 473
    self._init_dates(dates, freq)
ValueWarning: No frequency information was provided, so inferred frequency MS will be used.

Warning (from warnings module):
  File "C:\Users\User\Homa nanka\Lib\site-packages\statsmodels\tsa\base\tsa_model.py", line 473
    self._init_dates(dates, freq)
ValueWarning: No frequency information was provided, so inferred frequency MS will be used.

===== SARIMAX Results =====
Dep. Variable:          AveragePrice      No. Observations:      633
Model:              SARIMAX(1, 1, 1)x(1, 1, 1, 12)    Log Likelihood      -5162.446
Date:                Tue, 27 Jan 2026              AIC              10334.891
Time:                14:43:06                      BIC              10356.926
Sample:              04-01-1968                  HQIC              10343.465
                   - 12-01-2020

Covariance Type:      opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ar.L1          -0.2182      0.432      -0.505      0.613      -1.065      0.628
ma.L1           0.1569      0.439       0.357      0.721      -0.704      1.018
ar.S.L12        0.2034      0.039       5.222      0.000       0.127      0.280
ma.S.L12       -0.8595      0.020     -43.641      0.000      -0.898     -0.821
sigma2         1.457e+06  4.22e+04      34.498      0.000  1.37e+06  1.54e+06
=====
Ljung-Box (L1) (Q):          0.00    Jarque-Bera (JB):      1316.10
Prob(Q):          0.95    Prob(JB):          0.00
Heteroskedasticity (H):      78.59    Skew:          -0.53
Prob(H) (two-sided):          0.00    Kurtosis:         10.14
=====

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
RMSE прогноза на 2021-2025: 8372.99

```

Ln: 34 Col: 78



Поиск



1°C Cloudy

ENG

14:46

27.01.2026

