
**СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ/SYSTEM ANALYSIS,
MANAGEMENT AND PROCESSING OF INFORMATION**

DOI: <https://doi.org/10.60797/IRJ.2026.169.123> EDN: TKPERV**МОДЕЛИРОВАНИЕ СТОИМОСТИ ЖИЛОЙ НЕДВИЖИМОСТИ МОСКОВСКОЙ ОБЛАСТИ С
ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ**

Научная статья

Комаров Н.Г.^{1,*}, Марганова О.Н.²^{1,2} Российский технологический университет – МИРЭА, Щёлково, Российская Федерация

* Корреспондирующий автор (nikitakmrv76[at]gmail.com)

Предложена: 05.05.2026; Принята: 09.07.2026; Опубликовано: 17.07.2026

Аннотация

Рынок жилья в Московской области сильно формируется под влиянием непростых нелинейных связей. В этой работе продемонстрирован гибридный метод прогнозирования стоимости жилья, соединяющий машинное обучение и пространственную эконометрику. Для анализа использовалась выборка из 482 672 объектов за 2018–2021 годы. Было подтверждено, что цены связаны между собой в пространстве статистически значимо: $I = 0,56$ при $p = 0,001$. Наилучший результат показала модель Random Forest: MAE = 0,146 в логарифмическом масштабе, MAPE = 0,95%. SHAP-анализ показал, что сильнее всего на цену влияют площадь квартиры, количество комнат, расстояние до центра и локальная плотность. Такой подход может пригодиться в геофинтех-платформах и системах массовой оценки недвижимости.

Ключевые слова: прогнозирование стоимости недвижимости, пространственная автокорреляция, индекс Морана, SHAP-анализ, Random Forest, XGBoost, LightGBM, CatBoost, пространственная лаговая модель.

**MODELLING THE VALUE OF RESIDENTIAL PROPERTY IN MOSCOW OBLAST USING MACHINE
LEARNING METHODS**

Research article

Komarov N.G.^{1,*}, Marganova O.N.²^{1,2} Russian Technological University – MIREA, Schelkovo, Russian Federation

* Corresponding author (nikitakmrv76[at]gmail.com)

Suggested: 05.05.2026; Accepted: 09.07.2026; Published: 17.07.2026

Abstract

The housing market in Moscow Oblast is strongly influenced by complex non-linear relationships. This work presents a hybrid method for forecasting house prices, combining machine learning and spatial econometrics. A sample of 482,672 properties from 2018 to 2021 was used for the analysis. It was confirmed that prices are statistically significantly correlated in space: $I = 0.56$ at $p = 0.001$. The Random Forest model yielded the best results: MAE = 0.146 on a logarithmic scale, MAPE = 0.95%. SHAP analysis showed that the factors most strongly influencing price are the flat's floor area, the number of rooms, the distance to the city centre and local population density. This approach may prove useful in geofintech platforms and mass property valuation systems.

Keywords: property price forecasting, spatial autocorrelation, Moran's index, SHAP analysis, Random Forest, XGBoost, LightGBM, CatBoost, spatial lag model.

Введение

Современный рынок жилой недвижимости характеризуется высокой динамикой, сложной структурой пространственных взаимосвязей между объектами и значительной неоднородностью ценовых показателей. Московская область является одним из наиболее активных регионов Российской Федерации, где стоимость объектов определяется как индивидуальными характеристиками недвижимости, так и географическим расположением, а также доступностью транспортной и социальной инфраструктуры [1].

Традиционные методы оценки, основанные на регрессионных моделях, часто не учитывают пространственные зависимости и локальные особенности рынка, что ограничивает точность прогнозов [2], [3]. В последние годы возрос интерес к применению методов машинного обучения для прогнозирования цен на жилые объекты, включая ансамблевые алгоритмы и методы градиентного бустинга [4], [5].

Цель работы — разработка и апробация моделей прогнозирования стоимости жилой недвижимости Московской области, способных учитывать как структурные характеристики объектов, так и пространственные зависимости.

Задачи исследования включают:

1. Оценку пространственной автокорреляции цен и выявление кластеров объектов с высокими и низкими ценами.
2. Подготовку и очистку данных, включая обработку пропусков, фильтрацию выбросов и формирование пространственных признаков.

3. Прогнозирование стоимости жилья с использованием моделей машинного обучения и оценку точности прогнозов по метрикам MAE.

4. Применение пространственных регрессионных моделей для учета влияния соседних объектов на формирование цен.

5. Интерпретацию влияния факторов на стоимость жилья с использованием SHAP-анализа для выявления ключевых детерминант ценового уровня.

Для оценки прогнозной силы моделей использовалось разделение данных по временному принципу. Наблюдения за 2018–2020 гг. использовались для обучения моделей, а данные за 2021 г. формировали тестовую выборку, что позволяет оценить способность моделей к прогнозированию на будущих периодах в реальных рыночных условиях. Дополнительно на обучающей выборке применялась пятикратная перекрёстная проверка.

Гипотеза исследования заключается в предположении, что использование пространственных характеристик объектов недвижимости и учет пространственной автокорреляции цен позволяют повысить точность прогнозирования стоимости жилой недвижимости по сравнению с моделями, принимающие структурные характеристики объектов, при этом прирост точности составит не менее 5% на тестовой выборке.

Научная новизна работы заключается в интеграции методов машинного обучения и пространственных моделей. Это позволяет одновременно учитывать структурные характеристики объектов и пространственные эффекты при прогнозировании стоимости недвижимости. Реализация предложенного подхода обеспечивает повышение точности прогнозов и формирование практических рекомендаций для органов планирования, инвесторов и разработчиков геофинтех-платформ.

Обзор литературы

Оценка и прогнозирование стоимости жилой недвижимости является актуальной задачей как экономической науки, так и практической деятельности геофинтех-платформ. Традиционные регрессионные модели учитывают структурные характеристики объектов, такие как площадь квартиры, количество комнат или этажность здания. Однако эти модели часто не отражают влияние локальной среды и пространственной зависимости между объектами, что ограничивает точность прогнозирования [3], [6].

Пространственные модели позволяют выявлять локальные закономерности и количественно оценивать влияние соседних объектов на формирование цен. К наиболее распространённым подходам относятся:

- Spatial Lag (пространственная лаговая модель) — учитывает влияние цен соседних объектов через матрицу пространственных весов [2];

- Spatial Error — моделирует пространственную зависимость случайной ошибки [7];

- Geographically Weighted Regression (GWR) – локальная регрессия, позволяющая учитывать неоднородность факторов по территории [10].

Методы географически взвешенной регрессии широко применяются в исследованиях пространственной дифференциации цен жилья и позволяют выявлять территориальные вариации факторов [7].

Применение показателей пространственной автокорреляции, таких как индекс Moran's I, позволяет выявлять локальные кластеры объектов с высокой и низкой стоимостью, что является важным инструментом анализа структуры рынка [10], [13].

Параллельно с развитием пространственных моделей наблюдается активное внедрение методов машинного обучения. Ансамблевые алгоритмы и градиентный бустинг (Random Forest, XGBoost, LightGBM, CatBoost) демонстрируют высокую эффективность при моделировании нелинейных зависимостей между характеристиками объектов и их стоимостью [11], [12], [14], [18]. Эти методы позволяют строить точные прогнозные модели на больших массивах данных, учитывая как структурные характеристики объектов, так и влияние локальной среды [15].

Кроме сложных моделей, в литературе рассматривается применение линейной регрессии для анализа ключевых факторов формирования стоимости, подчёркивая важность очистки данных, отбора признаков и корректного построения моделей [8]. Дополнительно предлагаются классификации систем машинного обучения для систематизации подходов к анализу экономических данных и прогнозированию рыночных показателей [5], [16].

Сравнительный анализ различных моделей прогнозирования рыночной стоимости жилой недвижимости показывает, что выбор алгоритма существенно влияет на точность оценок, а динамика рыночных условий требует регулярной калибровки моделей, особенно при сопоставлении с кадастровой стоимостью [9].

Таким образом, обзор литературы выделяет два ключевых направления анализа рынка жилой недвижимости:

1. Пространственные модели — Moran's I, Spatial Lag, GWR, выявление пространственных закономерностей распределения цен;

2. Методы машинного обучения — высокоточные прогнозные модели на основе больших массивов данных с учетом как структурных характеристик объектов, так и влияния локальной среды.

При этом следует отметить, что каждый из рассмотренных подходов имеет ограничения: Spatial Lag эффективен при глобальной пространственной зависимости, но не учитывает локальную нестационарность факторов. Ансамблевые методы машинного обучения демонстрируют высокую точность на больших данных, однако их «чёрный ящик» затрудняет интерпретацию, что частично компенсируется применением SHAP-анализа [16]. Однако важно учитывать, что интерпретация SHAP-значений может быть затруднена при наличии мультиколлинеарных признаков, поскольку в таких условиях вклад коррелирующих переменных распределяется между ними неоднозначно [22].

Таким образом, синтез этих подходов позволяет компенсировать недостатки каждого из них. Синтез этих направлений и их интеграция в геофинтех-платформы представляют перспективное направление исследований, направленное на повышение точности прогнозирования, выявление локальных закономерностей и более глубокое понимание структуры рынка недвижимости [17], [19], [20].

Данные и подготовка к анализу

Задача прогнозирования стоимости жилой недвижимости формально представляется как моделирование функции (1):

$$P=f(X, S) \quad (1)$$

где X включает структурные характеристики объектов, а S – пространственные факторы, такие как локальная плотность, расстояние до центра и инфраструктурная доступность.

Эмпирической основой настоящего исследования является набор данных, размещённый в открытом доступе на платформе Hugging Face [21]. Данный датасет содержит информацию о рынке жилой недвижимости Российской Федерации за период 2018–2021 гг. и включает как структурные характеристики объектов, так и их географические координаты. Общий массив данных охватывает более 5 миллионов наблюдений по всем регионам страны. Это делает его одним из крупнейших открытых источников данных по рынку жилой недвижимости России.

Для целей данного исследования была выделена выборка объектов, расположенных на территории Московской области. Выбор региона обусловлен высокой активностью рынка недвижимости, значительным объёмом предложений и выраженной пространственной дифференциацией цен. После фильтрации по региональному признаку и предварительной очистки данных итоговый набор наблюдений составил около 480 тыс. записей. Это обеспечивает статистическую репрезентативность и позволяет проводить корректный эконометрический и пространственный анализ.

Для оценки прогнозной силы моделей использовалось разделение данных по временному принципу. Наблюдения за 2018–2020 гг. использовались для обучения моделей, а данные за 2021 г. формировали тестовую выборку. Дополнительно на обучающей выборке применялась пятикратная перекрёстная проверка.

Каждое наблюдение в наборе данных соответствует отдельному объекту жилой недвижимости и включает параметры, характеризующие как физические свойства объекта, так и его пространственное положение. К основным переменным относятся:

1. price — цена объекта недвижимости;
2. geo_lat, geo_lon — географические координаты объекта (широта и долгота);
3. area — общая площадь квартиры;
4. rooms — количество комнат;
5. level — этаж расположения квартиры;
6. levels — этажность здания;
7. kitchen_area — площадь кухни;
8. building_type / object_type — тип здания и тип объекта недвижимости.

Перед построением прогнозных моделей была выполнена процедура предварительной обработки данных, включающая следующие этапы:

1. Очистка данных — удаление наблюдений с пропущенными значениями ключевых признаков (цена, площадь, координаты).
2. Фильтрация выбросов — исключение аномальных значений площади и цены, выходящих за пределы реалистичных диапазонов рынка.
3. Удаление дубликатов — исключение повторяющихся объявлений объектов недвижимости.
4. Преобразование целевой переменной — для стабилизации дисперсии и уменьшения асимметрии распределения цены было применено логарифмическое преобразование переменной price.
5. Нормализация признаков — масштабирование числовых переменных для повышения устойчивости алгоритмов машинного обучения.

Дополнительно были сформированы новые признаки, учитывающие пространственные характеристики объектов недвижимости. В частности, был рассчитан показатель расстояния до центра города, используемый как прокси-переменная доступности инфраструктуры и транспортных узлов. Расстояние вычислялось на основе географических координат объектов.

Кроме того, введён показатель локальной плотности объектов недвижимости, отражающий степень концентрации объектов в пространстве. Этот показатель рассчитывался на основе среднего расстояния до ближайших соседних объектов. Данный показатель рассчитывался как величина, обратная среднему расстоянию до 10 ближайших соседей.

Использование как традиционных характеристик объектов недвижимости, так и пространственных признаков создаёт расширенный набор факторов, отражающих структурные и географические особенности рынка.

Корреляционный анализ признаков

Для выявления закономерностей взаимосвязей между переменными и анализа структуры данных был проведён корреляционный анализ признаков. В качестве основной меры линейной зависимости использовался коэффициент Пирсон, который широко применяется в эконометрических и пространственных исследованиях для оценки силы и направления взаимосвязи между количественными переменными.

Результаты корреляционного анализа выявили несколько ключевых закономерностей. Наибольшая положительная корреляция наблюдается между ценой объекта и общей площадью квартиры, что соответствует экономической логике рынка: более крупные квартиры, как правило, обладают более высокой стоимостью. Значительная положительная зависимость выявлена также между количеством комнат и площадью квартиры, что отражает структурную взаимосвязь характеристик жилых объектов.

Умеренная корреляция отмечена между ценой недвижимости и площадью кухни, что указывает на тенденцию увеличения площади кухни в более дорогих квартирах. В то же время признаки, связанные с этажом расположения

квартиры и этажностью здания, демонстрируют относительно слабую связь с ценой объекта, что согласуется с наблюдениями других региональных исследований рынка жилья.

Особое внимание привлекают пространственные признаки. Показатель расстояния до центра города демонстрирует отрицательную корреляцию с ценой, подтверждающую классическую гипотезу урбанистической экономики о снижении стоимости жилья по мере удаления от центральных районов. Показатель локальной плотности объектов отражает степень концентрации объектов недвижимости в определённой зоне, что также влияет на формирование цен. На рисунке 1 представлена матрица корреляции признаков, наглядно демонстрирующая линейные зависимости между основными переменными.

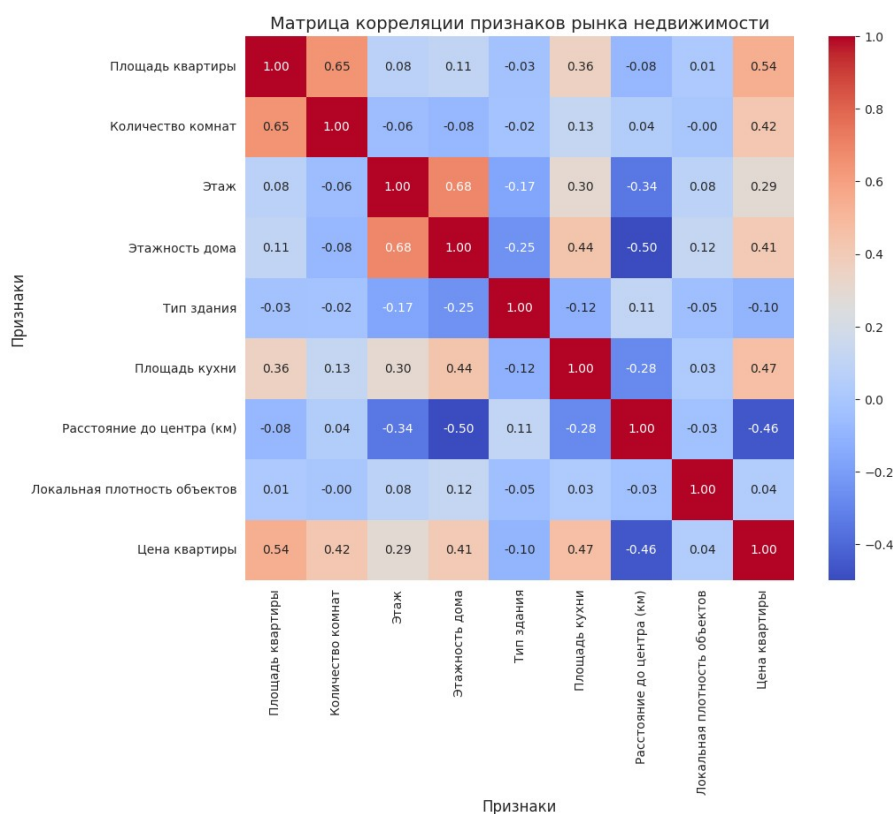


Рисунок 1 - Матрица корреляции признаков
DOI: <https://doi.org/10.60797/IRJ.2026.169.123.1>

Для дальнейшей интерпретации влияния отдельных факторов на прогноз стоимости жилой недвижимости был применён SHAP-анализ (SHapley Additive exPlanations) на модели XGBoost. Данный подход позволяет количественно оценить вклад каждого признака в прогноз модели [16]. Результаты SHAP показали, что наибольший вклад в прогноз вносят площадь квартиры и количество комнат, существенное влияние оказывают пространственные характеристики — расстояние до центра и локальная плотность объектов, а остальные признаки, включая этаж, этажность и площадь кухни, имеют умеренное воздействие. На рисунке 2 показан вклад признаков в прогноз стоимости жилья по SHAP-анализу модели XGBoost.

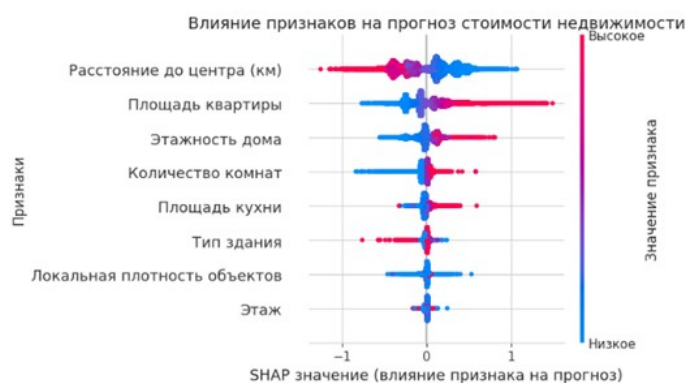


Рисунок 2 - Влияние признаков на прогноз стоимости недвижимости
DOI: <https://doi.org/10.60797/IRJ.2026.169.123.2>

Корреляционный анализ в сочетании с SHAP-интерпретацией демонстрирует согласованность структурных и пространственных факторов при формировании стоимости недвижимости. Полученные результаты позволяют выделить ключевые детерминанты ценового уровня, что повышает точность прогнозных моделей и создаёт основу для применения методов машинного обучения и пространственной эконометрики

Пространственный анализ рынка недвижимости

Одной из особенностей рынка жилой недвижимости является выраженная пространственная зависимость цен объектов. Стоимость жилья формируется не только на основе характеристик конкретного объекта, но и под влиянием факторов окружающей среды, таких как инфраструктура района, транспортная доступность, плотность застройки и уровень развития территории. В связи с этим для анализа структуры рынка недвижимости был проведён пространственный анализ, направленный на выявление пространственной автокорреляции цен и локальных кластеров стоимости жилья.

Для проверки наличия пространственной зависимости цен на недвижимость был рассчитан индекс Морана, который является одним из наиболее распространённых показателей пространственной автокорреляции [10], [13]. Данный индекс позволяет определить, наблюдается ли пространственная кластеризация значений исследуемой переменной.

Значения индекса Морана высчитывают следующим образом (2):

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

где n — количество наблюдений, x_i — значение исследуемой переменной в точке i , x_j — значение переменной в точке j , $\bar{x}$ — среднее значение переменной, w_{ij} — элемент матрицы пространственных весов, отражающий степень пространственной близости объектов i и j .

Результат интерпретируется следующим образом:

- $I > 0$ — положительная пространственная автокорреляция (схожие значения группируются в пространстве);
- $I \approx 0$ — пространственная случайность;
- $I < 0$ — отрицательная пространственная зависимость.

Для построения матрицы пространственных весов был выбран принцип k-ближайших соседей с $k=8$ [23]. Выбор данного параметра обусловлен результатами предварительного анализа чувствительности: при $k < 5$ матрица становится разреженной и не отражает значимых пространственных связей, а при $k > 15$ происходит избыточное сглаживание, снижающее чувствительность индекса Морана. Значение $k=8$ обеспечивает устойчивую оценку пространственной автокорреляции.

В ходе расчётов было получено значение индекса Морана $I = 0,56$ с $p\text{-value} = 0,001$. Значение индекса указывает на положительную пространственную автокорреляцию цен на жилую недвижимость, а крайне низкое $p\text{-value}$ показывает, что такая кластеризация статистически значима и маловероятна при случайном распределении цен. Это означает, что объекты с высокой стоимостью имеют тенденцию располагаться рядом друг с другом, формируя пространственные кластеры высоких цен, тогда как более дешёвые объекты также склонны группироваться в отдельных зонах.

В целях визуализации пространственной дифференциации стоимости недвижимости на территории Московской области было построено распределение цен на жилую недвижимость, представленное на рисунке 3.

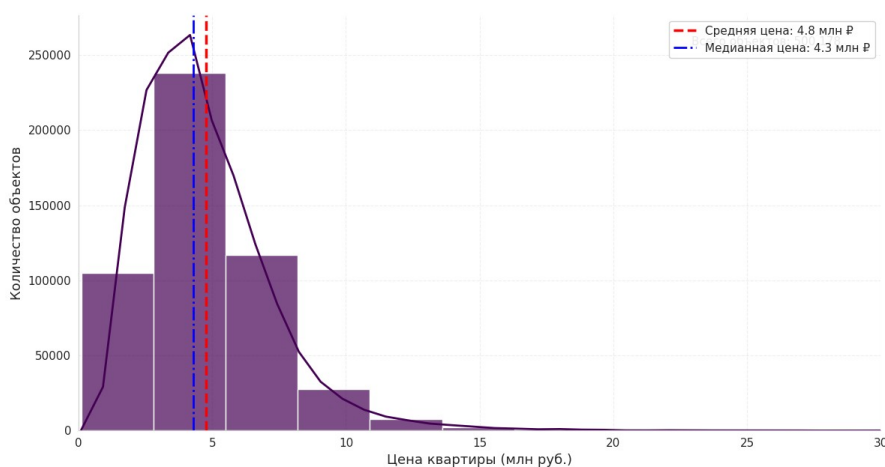


Рисунок 3 - Распределение цен на жилую недвижимость

DOI: <https://doi.org/10.60797/IRJ.2026.169.123.3>

Для более детального анализа пространственной структуры рынка были выявлены локальные кластеры стоимости недвижимости. В рамках пространственного анализа выделяются следующие типы кластеров, представленных на рисунке 4:

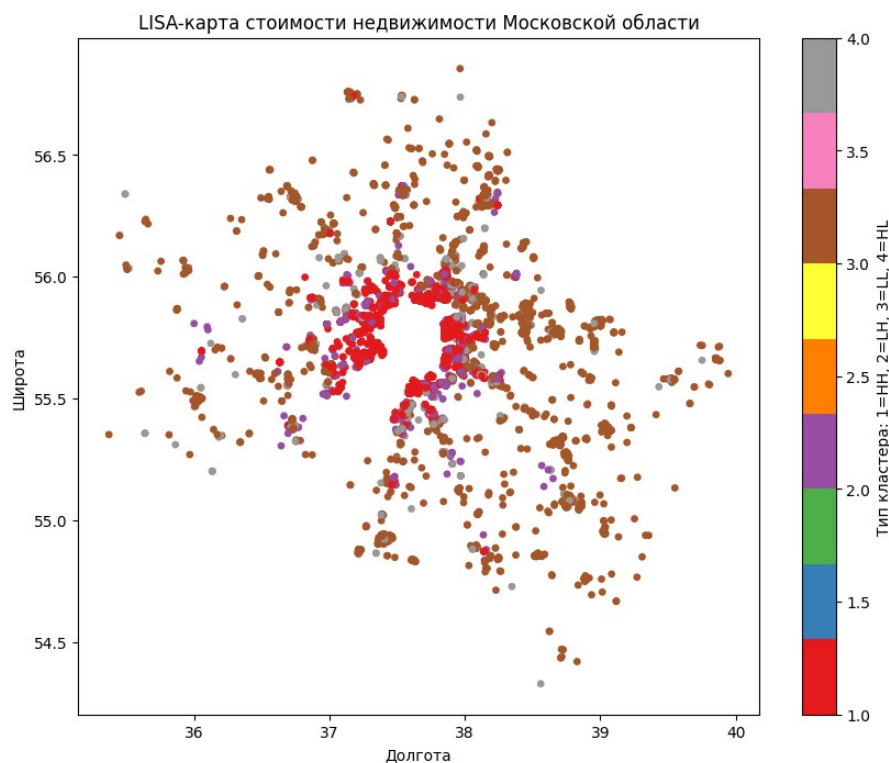


Рисунок 4 - Пространственные кластеры стоимости жилой недвижимости Московской области
DOI: <https://doi.org/10.60797/IRJ.2026.169.123.4>

1. High-High — области с высокой стоимостью жилья, окружённые объектами с аналогично высокой ценой.
2. Low-Low — зоны низких цен, формирующие кластеры доступного жилья.
3. High-Low — дорогие объекты в окружении более дешёвых.
4. Low-High — относительно дешёвые объекты, расположенные в районах с высокой стоимостью недвижимости.

Выявление подобных пространственных кластеров позволяет определить территории с различной ценовой структурой и выявить потенциально привлекательные зоны для инвестиций и развития инфраструктуры.

Методы исследования и моделирования стоимости недвижимости

Методологическая стратегия исследования строится на сочетании трёх взаимодополняющих аналитических подходов: системного анализа рынка как пространственно-распределённого объекта, инструментов пространственной эконометрики и ансамблевых методов машинного обучения. Пространственная связанность цен оценивалась с помощью индекса Морана, расчёт которого выполнялся на матрице пространственных весов, построенной по принципу k-ближайших соседей с $k=8$. Выбор $k=8$ обоснован результатами предварительного анализа чувствительности: при меньших значениях матрица становится разреженной и не отражает значимых пространственных связей, при больших происходит избыточное сглаживание, снижающее чувствительность индекса Морана.

Для прогнозирования стоимости жилой недвижимости были использованы методы машинного обучения и пространственной эконометрики. Применение данных подходов позволяет учитывать как характеристики объектов недвижимости, так и пространственные зависимости между ними. Основные модели, использованные в исследовании, представлены в таблице 1.

Таблица 1 - Методы моделирования стоимости недвижимости

DOI: <https://doi.org/10.60797/IRJ.2026.169.123.5>

Класс	Модель	Описание
Машинное обучение	Линейная регрессия	Базовая статистическая модель, описывающая линейную зависимость стоимости недвижимости от набора признаков. Используется как эталон для сравнения более сложных алгоритмов.
Машинное обучение	Случайный лес	Ансамблевый алгоритм на основе множества решающих

Класс	Модель	Описание
		деревьев, позволяющий моделировать сложные нелинейные зависимости между характеристиками объектов и их стоимостью [11], [12], [14], [18].
Машинное обучение	Градиентный бустинг (XGBoost, LightGBM, CatBoost)	Современные ансамблевые методы, основанные на последовательной минимизации ошибки модели и эффективно работающие на больших массивах данных.
Пространственные регрессии	Пространственная лаговая модель (Spatial Lag, ML_Lag)	Регрессионная модель, учитывающая влияние цен соседних объектов через матрицу пространственных весов.
Пространственная эконометрика	Географически взвешенная регрессия (GWR)	Метод локальной регрессии, позволяющий учитывать пространственную неоднородность факторов формирования цен на недвижимость.

Результаты исследования и их обсуждение

В рамках исследования была проведена комплексная оценка факторов, влияющих на стоимость жилой недвижимости Московской области, с использованием методов машинного обучения и пространственной эконометрики. Итоговая выборка включала 482 672 объекта недвижимости.

Корреляционный анализ показал значительную положительную зависимость цены от площади квартиры и количества комнат, умеренную — от площади кухни, а признаки этажности и этажа оказались менее значимыми. Пространственные признаки демонстрируют отрицательную корреляцию с расстоянием до центра города и подтверждают существование локальных кластеров высокой и низкой стоимости.

Пространственный анализ с использованием индекса Морана выявил выраженную положительную автокорреляцию цен, что подтверждает наличие локальных ценовых кластеров. Визуализация распределения стоимости объектов на карте Московской области позволила выявить зоны концентрации высоких и низких цен, а пространственная лаговая модель показала значимое влияние соседних объектов на формирование цены.

В ходе исследования была предпринята попытка оценки географически взвешенной регрессии для выявления локальной нестационарности влияния факторов ценообразования. Однако ввиду высокой мультиколлинеарности пространственных признаков (расстояния до центра и локальной плотности) и вычислительной сложности обращения матрицы пространственных весов, оценка GWR не позволила получить устойчивых результатов на используемой выборке.

Применение ансамблевых моделей машинного обучения позволило оценить точность прогноза стоимости жилья. Наименьшее значение средней абсолютной ошибки показала модель Случайный лес, при этом t-тест показал статистически значимое превосходство над XGBoost. SHAP-анализ подтвердил, что наибольший вклад в прогноз вносят площадь квартиры и количество комнат, значительное влияние оказывают пространственные признаки (расстояние до центра и локальная плотность объектов), а остальные признаки (этаж, этажность, площадь кухни) имеют умеренное воздействие. Результаты работы представлены в таблице 2.

Таблица 2 - Результаты исследования

DOI: <https://doi.org/10.60797/IRJ.2026.169.123.6>

Категория анализа	Показатель	Результат
Объем данных	Количество объектов	482 672 ед.
Статистика цены	Среднее	4 802 698 руб.
	Медиана	4 350 000 руб.
	25% квантиль	3 004 276 руб.
	75% квантиль	6 000 000 руб.
	Минимум / максимум	100 020 / 162 062 800
Пространственная	Индекс Морана	0,56 отн. ед.

Категория анализа	Показатель	Результат
автокорреляция	p-значение	0,001
Метрики моделей ML (MAE)	Линейная регрессия	0,228 отн. ед. (1 077 923 руб., 1,49%)
	Случайный лес	0,146 отн. ед. (660 051 руб., 0,95%)
	XGBoost	0,171 отн. ед. (786 867 руб., 1,12%)
	LightGBM	0,169 отн. ед. (776 881 руб., 1,11%)
	CatBoost	0,173 отн. ед. (793 870 руб., 1,13%)
	Spatial Lag	0,272 отн. ед. (1 313 634 руб., 1,78%)
Статистическая проверка	t-test XGB vs RF	t = 6,89, p = 0,002
Пространственная лаговая модель	Псевдо R ²	0,728
	Псевдо R ² с учетом пространственных весов	0,599
	Средневзвешенная зависимая переменная	0,601
Влияние признаков	Площадь квартиры	наибольший вклад
	Количество комнат	наибольший вклад
	Расстояние до центра	значительное влияние
	Локальная плотность объектов	значительное влияние
	Этаж, этажность, кухня	умеренное влияние

Ограничения исследования

Несмотря на полученные результаты и высокую точность прогнозов, исследование имеет ряд ограничений.

Во-первых, оно проводилось на основе данных о российской недвижимости в период 2018–2021 годов, при этом была определена выборка объектов Московской области для моделирования. В данном наборе данных отражены главным образом характеристики вторичного рынка жилья и не учтены характеристики рынка первичной недвижимости, которые могут оказывать значительное влияние на процессы ценообразования.

Во-вторых, в исследовании основное внимание уделялось структурным и пространственным характеристикам недвижимости без учета макроэкономических показателей, включая инфляцию, процентные ставки по ипотечным кредитам, изменения доходов и сезонные колебания спроса, которые могут оказывать значительное воздействие на динамику цен.

В-третьих, пространственные характеристики вытекают из географического местоположения объектов и плотности местных участков, но при этом не учитываются подробные особенности городов, такие как доступность транспорта, экологическая ситуация, качество социальной инфраструктуры и криминальная среда отдельных районов.

В-четвёртых, в работе предпринята попытка применения географически взвешенной регрессии (GWR) для выявления локальной нестационарности факторов. Однако ввиду высокой мультиколлинеарности пространственных признаков (расстояния до центра и локальной плотности, коэффициент корреляции между которыми составил 0,68) и вычислительной сложности обращения матрицы пространственных весов, устойчивая оценка GWR не была получена на используемой выборке. Данное ограничение может быть преодолено в будущих исследованиях с использованием регуляризованных версий GWR или методов машинного обучения с пространственными эффектами.

Кроме того, использование машинного обучения и пространственной эконометрики требует периодического обновления моделей и повторного обучения при получении новых данных, поскольку структура рынка недвижимости и влияние отдельных факторов меняются с течением времени; поэтому разработанный подход следует рассматривать в качестве основы для создания адаптивных систем интеллектуального прогнозирования, которые требуют дальнейшей доработки путем включения дополнительных пространственных, временных и социально-экономических факторов.

Заключение

Таким образом, проведённый анализ стоимости жилой недвижимости Московской области с использованием методов машинного обучения и пространственной эконометрики показал высокую точность прогнозирования и выявил ключевые детерминанты цен — площадь квартиры, количество комнат, расстояние до центра города и локальную плотность объектов. Полученные результаты демонстрируют, что интеграция ансамблевых моделей и пространственных регрессий позволяет учитывать как структурные характеристики объектов, так и их пространственные зависимости, что повышает практическую ценность прогноза.



Следует отметить, что применение SHAP-анализа и индекса Морана позволяет не только оценить влияние отдельных признаков на формирование стоимости, но и выявить локальные ценовые кластеры, что имеет важное значение для инвесторов, органов планирования и разработчиков геофинтех-платформ при принятии решений о вложениях и развитии инфраструктуры.

Перспективным направлением дальнейших исследований является разработка интеллектуальных геофинтех-систем. Реализация API позволит автоматически принимать сведения о новых объектах недвижимости, выполнять расчёт пространственных признаков, строить прогнозы стоимости и формировать интерпретируемые результаты в режиме, близком к реальному времени. Регулярное обновление моделей по мере поступления новых объявлений обеспечит адаптацию системы к изменению рыночной конъюнктуры, инфраструктурным преобразованиям территорий и влиянию внешних социально-экономических факторов. Разработанная технология может стать основой геофинтех-платформ, сервисов массовой оценки недвижимости, систем инвестиционной аналитики и интеллектуальных средств поддержки принятия решений органов государственного и муниципального управления.

Конфликт интересов

Не указан.

Рецензия

Шайхулова А.Ф., Уфимский университет науки и технологий, Уфа Российская Федерация

DOI: <https://doi.org/10.60797/IRJ.2026.169.123.7>

Богатырева Ю.И., Тульский государственный педагогический университет им. Л.Н. Толстого, Тула Российская Федерация

DOI: <https://doi.org/10.60797/IRJ.2026.169.123.8>

Conflict of Interest

None declared.

Review

Shaykhulova A.F., Ufa University of Science and Technology, Ufa Russian Federation

DOI: <https://doi.org/10.60797/IRJ.2026.169.123.7>

Bogatireva Y.I., Tula State Pedagogical University named after L.N. Tolstoy, Tula Russian Federation

DOI: <https://doi.org/10.60797/IRJ.2026.169.123.8>

Список литературы / References

1. Голубев Е.Ж. Анализ рынка недвижимости в Москве: применение алгоритмов машинного обучения для прогнозирования цен на квартиры / Е.Ж. Голубев // Экономика строительства. — 2024. — № 4. — С. 222–224.
2. Дмитриев С.Г. Исследование анизотропии экономического пространства России методами пространственной регрессии / С.Г. Дмитриев // Креативная экономика. — 2025. — Т. 19, № 2. — URL: <https://1economic.ru/lib/122566> (дата обращения: 10.03.26).
3. Катковская К.В. Прогнозирование стоимости недвижимости на основе модели линейной регрессии / К.В. Катковская, А.Д. Дудко // StudNet. — 2022. — № 6. — С. 5590–5603.
4. Павлова А.И. Применение машинных алгоритмов для прогнозирования стоимости недвижимости / А.И. Павлова, А.А. Корж // Сибирский журнал экономических и бизнес-исследований. — 2021. — № 4. — URL: <https://cyberleninka.ru/article/n/primenenie-mashinnyh-algoritmov-dlya-prognozirovaniya-stoimosti-nedvizhimosti> (дата обращения: 05.03.26).
5. Полетаева Н.Г. Классификация систем машинного обучения / Н.Г. Полетаева // Вестник Балтийского федерального университета им. И. Канта. Серия: физико-математические и технические науки. — 2020. — № 1. — URL: <https://cyberleninka.ru/article/n/klassifikatsiya-sistem-mashinnogo-obucheniya> (дата обращения: 05.03.26).
6. Родионов Д.Г. Методы машинного обучения в исследовании рынка жилой недвижимости / Д.Г. Родионов, Х.Х. Чан, Е.А. Конников и др. // Экономика и управление. — 2020. — № 7. — С. 458–464.
7. Седова Е.Н. Моделирование стоимости вторичного жилья на региональном рынке: пространственный подход / Е.Н. Седова // Экономика и управление. — 2014. — № 4(175). — С. 216–228.
8. Чесноков Е.А. Регрессионные модели оценки рыночной стоимости недвижимости в Санкт-Петербурге / Е.А. Чесноков // Московский экономический журнал. — 2025. — № 3. — URL: <https://ecience.ru/en/nauka/article/96103/view> (дата обращения: 10.03.26).
9. Шалагин А.А. Сравнительный анализ моделей прогнозирования рыночной стоимости жилой недвижимости и влияние динамики рынка на кадастровую стоимость / А.А. Шалагин, С.А. Григорьев // Известия высших учебных заведений. Геодезия и аэрофотосъёмка. — 2022. — № 4. — URL: <https://www.naukaru06.ru/ru/nauka/article/105506/view> (дата обращения: 10.03.26).
10. Anselin L. Spatial Econometrics: Methods and Models / L. Anselin. — Dordrecht : Kluwer Academic Publishers, 1988. — 284 p.
11. Breiman L. Random Forests / L. Breiman // Machine Learning. — 2001. — Vol. 45, № 1. — URL: <https://link.springer.com/article/10.1023/A:1010933404324> (accessed: 10.03.26).
12. Chen T. XGBoost: A Scalable Tree Boosting System / T. Chen, C. Guestrin // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. — 2016. — URL: <https://dl.acm.org/doi/10.1145/2939672.2939785> (accessed: 10.03.26).
13. Chen Y. Spatial autocorrelation equation based on Moran's index / Y. Chen // Scientific Reports. — 2023. — URL: <https://www.nature.com/articles/s41598-023-45947-x> (дата обращения: 10.03.26).
14. Ke G. LightGBM: A Highly Efficient Gradient Boosting Decision Tree / G. Ke, Q. Meng, T. Finley // Advances in Neural Information Processing Systems. — 2017. — URL: <https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree> (дата обращения: 10.03.26).



15. Li C. Review of Housing Price Forecasting Methods Based on Machine Learning and Deep Learning / C. Li // *Applied and Computational Engineering*. — 2025. — URL: <https://ace.ewapub.com/article/view/20931>. (дата обращения: 10.03.26)
16. Lundberg S.M. A Unified Approach to Interpreting Model Predictions / S.M. Lundberg, S.-I. Lee // *Advances in Neural Information Processing Systems*. — 2017. — URL: <https://arxiv.org/abs/1705.07874> (accessed: 10.03.26)
17. Neves F.T. The Impacts of Open Data and eXplainable AI on Real Estate Price Predictions in Smart Cities / F.T. Neves, M. Aparicio, M. de Castro Neto // *Applied Sciences*. — 2024. — № 5. — URL: <https://www.mdpi.com/2076-3417/14/5/2209> (accessed: 10.03.26)
18. Prokhorenkova L. CatBoost: Unbiased Boosting with Categorical Features / L. Prokhorenkova, G. Gusev, A. Vorobev et al. // *Advances in Neural Information Processing Systems*. — 2018. — URL: <https://arxiv.org/abs/1706.09516>. (дата обращения: 10.03.26)
19. Zhao H. Predicting Real Estate Price Using Stacking-Based Ensemble Learning / H. Zhao, K. Wang // *American Journal of Information Science and Technology*. — 2023. — № 2. — URL: <https://sciencepublishinggroup.com/article/10.11648/j.ajist.20230702.14> (accessed: 10.03.26)
20. Zhang H. Predicting Housing Prices Using Supervised Machine Learning Models / H. Zhang // *Advances in Economics, Management and Political Sciences*. — 2024. — URL: <https://aemps.ewapub.com/article/view/16015>. (дата обращения: 10.03.26)
21. Russia_Real_Estate_2018_2021 // Hugging Face. — URL: https://huggingface.co/datasets/daniilakk/Russia_Real_Estate_2018_2021 (accessed: 10.03.2026).
22. Molnar C. A Guide for Making Black Box Models Explainable / C. Molnar // *Interpretable Machine Learning*. — 2022. — URL: <https://christophm.github.io/interpretable-ml-book/>. (дата обращения: 22.06.26)
23. Anselin L. An Introduction to Spatial Data Science / L. Anselin // *GeoDa*. — 2020. — URL: <https://geodacenter.github.io/>. (дата обращения: 22.06.26)

Список литературы на английском языке / References in English

1. Golubev E.Zh. Analiz ry'nka nedvizhimosti v Moskve: primeneniye algoritmov mashinnogo obucheniya dlya prognozirovaniya cen na kvartiry' [Analysis of the real estate market in Moscow: application of machine learning algorithms for apartment price prediction] / E.Zh. Golubev // *Construction Economics*. — 2024. — № 4. — P. 222–224. [in Russian]
2. Dmitriev S.G. Issledovanie anizotropii ekonomicheskogo prostranstva Rossii metodami prostranstvennoi regressii [Study of anisotropy of the economic space of Russia using spatial regression methods] / S.G. Dmitriev // *Kreativnaya ekonomika [Creative Economy]*. — 2025. — Vol. 19, № 2. — URL: <https://1economic.ru/lib/122566> (accessed: 10.03.26). [in Russian]
3. Katkovskaya K.V. Prognozirovaniye stoimosti nedvizhimosti na osnove modeli lineynoy regressii [Forecasting real estate prices based on a linear regression model.] / K.V. Katkovskaya, A.D. Dudko // *StudNet*. — 2022. — № 6. — P. 5590–5603. [in Russian]
4. Pavlova A.I. Primeneniye mashinnikh algoritmov dlya prognozirovaniya stoimosti nedvizhimosti [Application of machine learning algorithms for real estate price prediction] / A.I. Pavlova, A.A. Korzh // *Sibirskii zhurnal ekonomicheskikh i biznes-issledovaniy [Siberian Journal of Economic and Business Research]*. — 2021. — № 4. — URL: <https://cyberleninka.ru/article/n/primeneniye-mashinnyh-algoritmov-dlya-prognozirovaniya-stoimosti-nedvizhimosti> (accessed: 05.03.26). [in Russian]
5. Poletaeva N.G. Klassifikatsiya sistem mashinnogo obucheniya [Classification of machine learning systems] / N.G. Poletaeva // *Vestnik Baltiiskogo federalnogo universiteta im. I. Kanta. Seriya: fiziko-matematicheskie i tekhnicheskie nauki [Bulletin of Immanuel Kant Baltic Federal University. Series: Physical and Mathematical and Technical Sciences]*. — 2020. — № 1. — URL: <https://cyberleninka.ru/article/n/klassifikatsiya-sistem-mashinnogo-obucheniya> (accessed: 05.03.26). [in Russian]
6. Rodionov D.G. Metody' mashinnogo obucheniya v issledovanii ry'nka zhiloy nedvizhimosti [Machine learning methods in residential real estate market research] / D.G. Rodionov, X.X. Chan, E.A. Konnikov et al. // *Economics and Management*. — 2020. — № 7. — P. 458–464. [in Russian]
7. Sedova E.N. Modelirovaniye stoimosti vtorichnogo zhil'ya na regional'nom ry'nke: prostranstvenny'j podkhod [Modeling the cost of secondary housing in the regional market: a spatial approach] / E.N. Sedova // *Economics and Management*. — 2014. — № 4(175). — P. 216–228. [in Russian]
8. Chesnokov Ye.A. Regressionnye modeli otsenki rinochnoi stoimosti nedvizhimosti v Sankt-Peterburge [Regression models for estimating the market value of real estate in St. Petersburg] / Ye.A. Chesnokov // *Moskovskii ekonomicheskii zhurnal [Moscow Economic Journal]*. — 2025. — № 3. — URL: <https://ecience.ru/en/nauka/article/96103/view> (accessed: 10.03.26). [in Russian]
9. Shalagin A.A. Sravnitel'nii analiz modelei prognozirovaniya rinochnoi stoimosti zhiloi nedvizhimosti i vliyaniye dinamiki rinka na kadastruvuyu stoimost [Comparative analysis of models for predicting the market value of residential real estate and the influence of market dynamics on cadastral value] / A.A. Shalagin, S.A. Grigorev // *Izvestiya visshikh uchebnikh zavedenii. Geodeziya i aerofotosyomka [Proceedings of Higher Educational Institutions. Geodesy and Aerial Surveying]*. — 2022. — № 4. — URL: <https://www.naukaru06.ru/ru/nauka/article/105506/view> (accessed: 10.03.26). [in Russian]
10. Anselin L. Spatial Econometrics: Methods and Models / L. Anselin. — Dordrecht : Kluwer Academic Publishers, 1988. — 284 p.
11. Breiman L. Random Forests / L. Breiman // *Machine Learning*. — 2001. — Vol. 45, № 1. — URL: <https://link.springer.com/article/10.1023/A:1010933404324> (accessed: 10.03.26).



12. Chen T. XGBoost: A Scalable Tree Boosting System / T. Chen, C. Guestrin // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. — 2016. — URL: <https://dl.acm.org/doi/10.1145/2939672.2939785> (accessed: 10.03.26)
13. Chen Y. Spatial autocorrelation equation based on Moran's index / Y. Chen // Scientific Reports. — 2023. — URL: <https://www.nature.com/articles/s41598-023-45947-x>. (accessed: 10.03.26)
14. Ke G. LightGBM: A Highly Efficient Gradient Boosting Decision Tree / G. Ke, Q. Meng, T. Finley // Advances in Neural Information Processing Systems. — 2017. — URL: <https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree>. (accessed: 10.03.26)
15. Li C. Review of Housing Price Forecasting Methods Based on Machine Learning and Deep Learning / C. Li // Applied and Computational Engineering. — 2025. — URL: <https://ace.ewapub.com/article/view/20931>. (accessed: 10.03.26)
16. Lundberg S.M. A Unified Approach to Interpreting Model Predictions / S.M. Lundberg, S.-I. Lee // Advances in Neural Information Processing Systems. — 2017. — URL: <https://arxiv.org/abs/1705.07874> (accessed: 10.03.26)
17. Neves F.T. The Impacts of Open Data and eXplainable AI on Real Estate Price Predictions in Smart Cities / F.T. Neves, M. Aparicio, M. de Castro Neto // Applied Sciences. — 2024. — № 5. — URL: <https://www.mdpi.com/2076-3417/14/5/2209> (accessed: 10.03.26)
18. Prokhorenkova L. CatBoost: Unbiased Boosting with Categorical Features / L. Prokhorenkova, G. Gusev, A. Vorobev et al. // Advances in Neural Information Processing Systems. — 2018. — URL: <https://arxiv.org/abs/1706.09516>. (accessed: 10.03.26)
19. Zhao H. Predicting Real Estate Price Using Stacking-Based Ensemble Learning / H. Zhao, K. Wang // American Journal of Information Science and Technology. — 2023. — № 2. — URL: <https://sciencepublishinggroup.com/article/10.11648/j.ajist.20230702.14> (accessed: 10.03.26)
20. Zhang H. Predicting Housing Prices Using Supervised Machine Learning Models / H. Zhang // Advances in Economics, Management and Political Sciences. — 2024. — URL: <https://aemps.ewapub.com/article/view/16015>. (accessed: 10.03.26)
21. Russia_Real_Estate_2018_2021 // Hugging Face. — URL: https://huggingface.co/datasets/daniilakk/Russia_Real_Estate_2018_2021 (accessed: 10.03.2026).
22. Molnar C. A Guide for Making Black Box Models Explainable / C. Molnar // Interpretable Machine Learning. — 2022. — URL: <https://christophm.github.io/interpretable-ml-book/>. (accessed: 22.06.26)
23. Anselin L. An Introduction to Spatial Data Science / L. Anselin // GeoDa. — 2020. — URL: <https://geodacenter.github.io/>. (accessed: 22.06.26)