

Научная статья
УДК 332.63:330.43:004.8
<https://doi.org/10.17586/2713-1874-2026-1-72-88>

СРАВНИТЕЛЬНЫЙ АНАЛИЗ НЕЙРОСЕТЕВЫХ И АНСАМБЛЕВЫХ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗНОЙ КАДАСТРОВОЙ ОЦЕНКИ ЗЕМЛИ В УСЛОВИЯХ НЕПОЛНЫХ ПРОСТРАНСТВЕННЫХ ДАННЫХ

Максим Викторович Натыкин^{1✉}, Сергей Александрович Митягин²

^{1,2}Университет ИТМО, Санкт-Петербург, Россия

¹mvin@itmo.ru ✉, <https://orcid.org/0009-0003-8195-0007>

²mityagin@itmo.ru, <https://orcid.org/0000-0001-9877-1687>

Язык статьи – русский

Аннотация: Цель исследования заключается в повышении оперативности и достоверности прогнозной кадастровой оценки земельных участков при неполных и разнородных пространственных данных за счёт применения и сопоставления современных методов искусственного интеллекта, использующих параметры застройки и характеристики территориального окружения. В качестве материалов использованы официальные данные государственной кадастровой оценки, сведения Единой цифровой платформы «Национальная система пространственных данных», а также пространственные, социально-экономические и инфраструктурные показатели, полученные из ГИС-источников и открытых данных.

Методология включает формирование квартальной пространственно-сетевой модели, расчёт морфометрических индексов, показателей доступности и проверку пространственной автокорреляции, обучение моделей CatBoost/XGBoost и нейросети с оценкой по R^2 , MAE и RMSE и интерпретацией факторов методом SHAP. Показано, что MLP достигает $R^2=0,76$, а CatBoost обеспечивает близкую точность ($R^2=0,73$) при большей объяснимости; ключевыми факторами выступают площадь, плотность застройки, функциональное назначение и транспортная доступность.

Сделаны выводы о воспроизводимости, масштабируемости и применимости метода в других регионах России, обеспечение унификации и сопоставимость подходов к оценке неучтённых земель и влияние на цифровизацию процессов пространственного планирования. Дальнейшие исследования будут направлены на расширение набора данных за счёт инженерной инфраструктуры, внедрения больших языковых моделей, а также разработку программного комплекса для практического использования модели.

Ключевые слова: автоматизированная оценка недвижимости, земельные ресурсы, информационные технологии в управлении, искусственный интеллект, кадастровая оценка, машинное обучение, цифровая трансформация, CatBoost

Ссылка для цитирования: Натыкин М. В., Митягин С. А. Сравнительный анализ нейросетевых и ансамблевых моделей машинного обучения для прогнозной кадастровой оценки земли в условиях неполных пространственных данных // Экономика. Право. Инновации. – 2026. – Т. 14. – № 1. – С. 72–88. – <https://doi.org/10.17586/2713-1874-2026-1-72-88>

COMPARATIVE ANALYSIS OF NEURAL NETWORKS AND ENSEMBLE MACHINE LEARNING MODELS FOR PREDICTIVE CADASTRAL LAND VALUATION UNDER INCOMPLETE SPATIAL DATA

Maksim V. Natykin^{1✉}, Sergey A. Mityagin²

^{1,2}ITMO University, Saint Petersburg, Russia

¹mvin@itmo.ru ✉, <https://orcid.org/0009-0003-8195-0007>

²mityagin@itmo.ru, <https://orcid.org/0000-0001-9877-1687>

Article in Russian

Abstract: The purpose of the study is to increase the efficiency and reliability of the forecasted cadastral valuation of land plots with incomplete and heterogeneous spatial data through the application and comparison of modern artificial intelligence methods using building parameters and characteristics of the territorial environment. The materials used are official data from the state cadastral assessment, information from the Unified Digital Platform «National Spatial Data System», as well as spatial, socio-economic and infrastructural indicators obtained from GIS sources and open data.

The methodology includes the formation of a quarterly spatial network model, the calculation of morphometric indices, accessibility indicators and spatial autocorrelation verification, training of CatBoost/XGBoost and neural network

models with R2, MAE and RMSE estimates and interpretation of factors using the SHAP method. It is shown that MLP reaches R2=0.76, and CatBoost provides close accuracy (R2=0.73) with greater explainability; the key factors are area, building density, functional purpose and transport accessibility.

Conclusions are drawn about the reproducibility, scalability and applicability of the method in other regions of Russia, ensuring the unification and comparability of approaches to assessing unaccounted for lands and the impact of spatial planning processes on digitalization. The results of further research will be aimed at expanding the data set through engineering infrastructure, the introduction of large language models, as well as the development of a software package for the practical use of the model by government authorities.

Keywords: automated property valuation, land resources, information technologies in management, artificial intelligence, cadastral valuation, machine learning, digital transformation, CatBoost

For citation: Natykin M. V., Mityagin S. A. Comparative Analysis of Neural Networks and Ensemble Machine Learning Models for Predictive Cadastral Land Valuation under Incomplete Spatial Data. *Ekonomika. Pravo. Innovacii*. 2026. Vol. 14. No. 1. pp. 72–88. (In Russ.). <https://doi.org/10.17586/2713-1874-2026-1-72-88>

Введение. Современная система градостроительного планирования сохраняет ряд пробелов, необходимость устранения которых отражена в перечне поручений Президента РФ от 6 сентября 2023 г. № Пр-1770 [1]. В методических рекомендациях Минцифры также подчёркивается важность автоматизации оценочных процедур как элемента цифровой трансформации градостроительства [2].

Кадастровая стоимость земли формирует налоговую базу и используется в управленческих решениях, однако данные ЕГРН нередко неполны или неактуальны, что осложняет получение обоснованных оценок и снижает качество фискального и территориального планирования. В связи с этим требуется разработка методов массовой оценки, устойчивых к дефициту пространственных сведений и пригодных для аналитических расчётов. В мировой практике для этих целей широко применяются модели машинного обучения, учитывающие характеристики застройки, инфраструктуры и окружения объектов.

Задача особенно актуальна для проектов планировки территории и комплексного развития территорий (ППТ/КРТ), когда необходимо оценивать потенциальную кадастровую стоимость при отсутствии фактической застройки и опоре преимущественно на проектные параметры [3–6]. Предлагаемый подход ориентирован на сценарное прогнозирование изменений стоимости до очередного цикла государственной кадастровой оценки, выполняемой по ФЗ № 237-ФЗ и методикам Росреестра [7, 8]. Инструмент может использоваться органами власти для оценки последствий различных градостроительных сценариев и предварительного анализа налоговых

эффектов на стадии ТЭО [5, 6], а также застройщиками для экспресс-оценки инвестиционной привлекательности земельных участков в логике ППТ/КРТ [4, 5]. При этом модель не заменяет официальную ГКО, а применяется как аналитическое средство поддержки решений в рамках механизмов развития территорий [7, 8].

Постановка задачи исследования. Целью исследования является повышение скорости и точности прогнозной кадастровой оценки земельных участков в условиях неполноты и неоднородности пространственных данных путём применения и сравнительного анализа современных методов машинного обучения на основе параметров застройки и территориального окружения.

Объектом исследования являются земельные участки урбанизированных территорий, рассматриваемые как элементы территориальной системы.

Предметом исследования являются методы и алгоритмы предсказательного моделирования кадастровой стоимости земельных участков с использованием пространственных, планировочных и инфраструктурных признаков.

Для достижения поставленной цели в работе решаются следующие задачи.

1) Выполнить аналитический обзор существующих методов кадастровой оценки и предсказательного моделирования стоимости земли, применяемых в условиях массовой оценки и неполноты исходных данных.

2) Выявить системные особенности и ограничения исходного набора данных (неоднородность, пространственная зависимость, пропуски значений), определяющие постановку задачи прогнозирования кадастровой стоимости.

3) Сформировать воспроизводимый конвейер подготовки данных, включающий агрегацию пространственных и планировочных характеристик земельных участков и их территориального окружения.

4) Реализовать и обучить несколько классов моделей машинного обучения, включая ансамблевые и нейросетевые архитектуры, для прогнозирования кадастровой стоимости земельных участков и провести их сравнительный анализ по показателям точности, устойчивости и чувствительности к ограниченности входных данных.

5) Использовать нейросетевые модели в качестве инструмента исследования верхней границы достижимого качества прогнозирования при заданном минимальном наборе признаков и сопоставить полученные результаты с более простыми и устойчивыми моделями.

6) Оценить адекватность полученных моделей с использованием статистических метрик качества и анализа ошибок, а также выявить ограничения применимости различных подходов в задачах кадастровой оценки.

7) Проанализировать результаты экспериментов, сформулировать выводы и разработать рекомендации по использованию предсказательных моделей в задачах кадастровой оценки и территориального планирования.

Научная гипотеза исследования состоит в том, что даже при использовании минимального набора исходных данных, включающего агрегированные параметры планировочной структуры и застройки территории: площадь квартала, показатели плотности застройки FSI и GSI, функциональное назначение, а также базовые показатели транспортной доступности, возможно получение прогнозной кадастровой оценки земельных участков с приемлемым уровнем точности, сопоставимым с результатами моделей массовой оценки, применяемых в практике кадастровых расчётов. При этом применение более сложных моделей, включая нейросетевые архитектуры, позволяет оценить предельный потенциал качества прогнозирования и обосновать выбор методов, наиболее целесообразных для практических задач территориального управления.

Литературный обзор. Массовая оценка недвижимости используется для налогообложения и территориального планирования и в РФ выполняется уполномоченными учреждениями по федеральным нормам и методическим указаниям [7, 8]. Вместе с тем, традиционные подходы (сравнительный и доходный анализ, включая многомерные регрессии/hedonic pricing models), испытывают ограничения при учёте нелинейностей, взаимодействий факторов и неполноты рыночных данных; предпосылки линейности и независимости признаков нередко снижают точность [9]. В современной литературе показано, что методы машинного обучения повышают качество AVM за счёт выявления скрытых закономерностей: ансамбли (случайные леса, градиентный бустинг XGBoost/LightGBM/CatBoost) лучше работают с нелинейными эффектами и взаимодействиями, а CatBoost дополнительно эффективен при наличии категориальных признаков.

Отмечается рост интереса к глубокому обучению: в работе [10] динамическая нейросетевая модель для массовой оценки жилья в Китае показала высокую точность. На валидационном наборе тестирования модель обеспечила $R^2 \approx 0,93$, $MAPE \approx 8,2\%$ и $RMSE \approx 97$, превышая по всем основным метрикам традиционные реализации множественной регрессии и BP-нейросети.

В то же время архитектуры типа KAN предлагают более выразительное представление нелинейных зависимостей [11]. При этом для табличных структурированных данных ансамблевые методы градиентного бустинга, основанные на последовательном обучении ансамбля решающих деревьев с минимизацией функции потерь, часто сохраняет стабильность и интерпретируемость, тогда как нейросети выигрывают при усложнении данных и при необходимости учитывать пространственную/мультимодальную информацию [10]. Доказано, что изображения фасадов, панорамы улиц, спутниковые снимки совместно с табличными признаками повышает точность прогнозирования цен за счёт мультимодального анализа [12].

Отдельное направление связано с прогнозированием стоимости земли по параметрам проектируемой застройки, востребованным в ППТ и КРТ: FSI/FAR, GSI, этажность, функ-

циональное назначение и др. Зарубежные результаты подтверждают информативность таких показателей: например, в Нью-Йорке (США) выявлена прямая связь цены земли с разрешённой интенсивностью застройки (FAR) [13]. Подчёркивается значимость корректного отбора и дифференциации признаков для объективности оценки [14]. В целом, интеграция ГИС-данных и методов ML рассматривается как основа гибридных систем массовой оценки, повышающих обоснованность и точность, что согласуется с курсом на цифровизацию кадастровой оценки [15].

Обработка и анализ данных. В качестве источника данных для формирования обучающей выборки использовались Платформа цифровой урбанистики Университета ИТМО [20] и открытые данные OpenStreetMap (OSM) [21]. Апробируемыми территориями выбран Санкт-Петербург и Ленинградская область. Подготовка набора данных проходила в три этапа:

- 1) Построение пространственного слоя городских кварталов.
- 2) Получение параметров городских зданий.
- 3) Определение параметров функционального зонирования.

Для создания пространственного слоя городских кварталов была использована python-библиотека открытого кода block-snet [16]. Пространственные данные загружались из OSM при помощи OSMnx. Были использованы три вспомогательных слоя:

- 1) Водные объекты: теги riverbank, reservoir, basin, dock, canal, pond, natural=water/bay, waterway=river/canal/ditch, landuse=basin, water=lake.
- 2) Автомобильные дороги: теги highway (например, motorway, primary, living_street, construction), а также service=living_street/emergency_access.
- 3) Железные дороги: тег railway=rail.

Для генерации слоя все полученные вспомогательные слои были приведены к единой проекционной системе координат EPSG:32636. Результатом генерации пространственного слоя городских кварталов является GeoDataFrame из 9 988 строк, содержащий следующие параметры:

Geometry: Polygon – геометрия здания;
 Site area: float >= 0 – площадь квартала (м²);

Length: float >= 0 – периметр квартала (м).

В качестве источника пространственных данных городских зданий была выбрана Платформа цифровой урбанистики Университета ИТМО. Полученный пространственный слой состоял из 163049 зданий и был приведен к следующему формату:

1) Geometry : Polygon | MultiPolygon | Point – геометрия здания.

2) Population : int >= 0 – количество населения, проживающего в выбранном доме.

3) Footprint_area : float >= 0 – площадь пятна застройки здания.

4) Build_floor_area : float >= 0 – поэтажная площадь здания:

– Living_area : float >= 0 – поэтажная жилая площадь здания.

– Non_living_area : float >= 0 – поэтажная нежилая площадь здания.

– Number_of_floors : int >= 1 – количество этажей здания.

– Is_living : bool – является ли здание жилым.

Чтобы агрегировать геометрические параметры зданий, их исходные контуры сначала приводились к центроидам и перепроецировались в систему координат EPSG:32636. Полученный точечный слой сопоставлялся с полигональным городских кварталов, внутри которых агрегировались показатели всех зданий, расположенных в квартале. Индексы FSI, GSI и MXI для кварталов были вычислены по представленным ниже формулам. Визуализация полученных параметров представлена на рисунке 1.

$$FSI = \frac{build_floor_area}{site_area}, FSI \in [0, +\infty] \quad (1)$$

$$GSI = \frac{footprint_area}{site_area}, GSI \in [0, 1] \quad (2)$$

$$MXI = \frac{living_area}{build_floor_area}, MXI \in [0, 1] \quad (3)$$

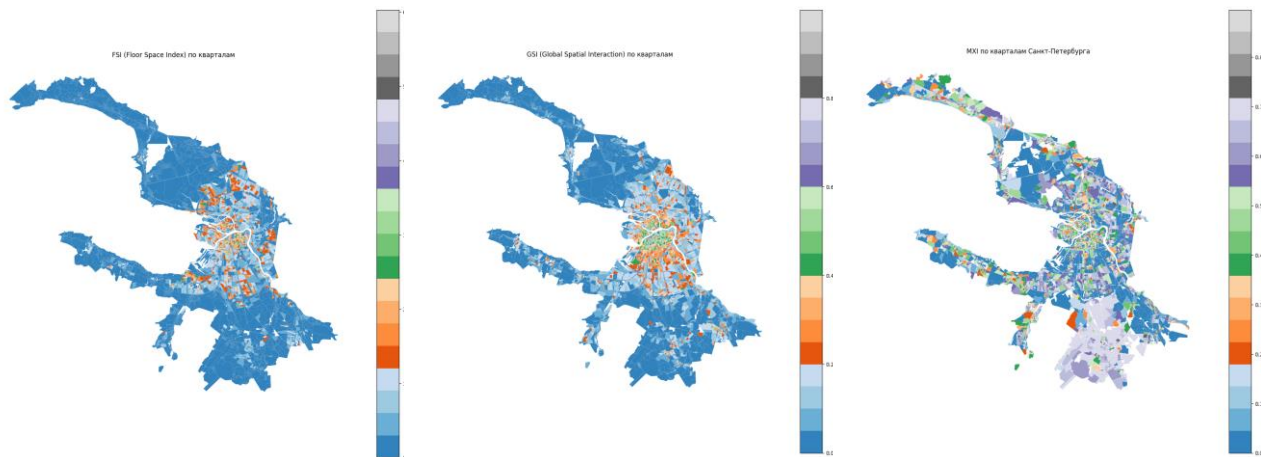


Рисунок 1 – Визуализация параметров застройки кварталов

Источник: составлен авторами

Определяется определенный паттерн предсказываемых значений: значения параметров FSI и GSI увеличиваются к центру. При этом определить визуально паттерн параметров MXI не представляется возможным.

Таким образом, результатом на данном этапе является пространственный слой городских кварталов по городу Санкт-Петербург, представленный в системе координат EPSG:32636 со следующими параметрами:

- Fsi : float ≥ 0 – floor space index;
- Gsi : float $\geq 0, \leq 1$ – ground space index;
- Mxi : float $\geq 0, \leq 1$ – mixed use index.

Для получения параметров функционального зонирования был использован слой территориального зонирования правил землепользования и застройки за 2023 год. Территориальные зоны были приведены к следующим функциональным зонам:

- RESIDENTIAL – жилое функциональное назначение.
- RECREATION – рекреационное функциональное назначение.
- BUSINESS – деловое функциональное назначение.
- INDUSTRIAL – складское и промышленное функциональное назначение.
- TRANSPORT – транспортное функциональное назначение.
- SPECIAL – специальное функциональное назначение.
- AGRICULTURE – сельскохозяйственное функциональное назначение.

Полученный пространственный слой был соотнесен со слоем городских кварталов. Таким образом, были получены соотношения функционального зонирования каждого типа функционального зонирования в рамках квартала по следующей формуле:

$$share_i = \frac{\sum intersection_{ij}}{site_area} \quad (4)$$

где:

- $share_i$ – доля функционального зонирования i -го типа в данном квартале.
- $intersection_{ij}$ – площадь пересечения j -й функциональной зоны i -го типа.
- $site_area$ – площадь квартала.

Полученные данные представляют собой геопространственную модель морфологии города. Визуальный анализ показал закономерное радиальное повышение FSI и GSI к историческому центру, тогда как MXI не демонстрирует выраженного пространственного тренда, что подтверждает разнородность функциональной структуры периферийных зон. Дополнительно по интермодальному графу улично-дорожной сети вычисляется матрица связанности всех кварталов. Это позволяет получить полный и статистически разнообразный набор признаков для последующего получения важных топологических признаков центральности, связанности кварталов, а также регрессионного моделирования стоимости земли. Визуализация функционального зонирования представлена на рисунке 2.

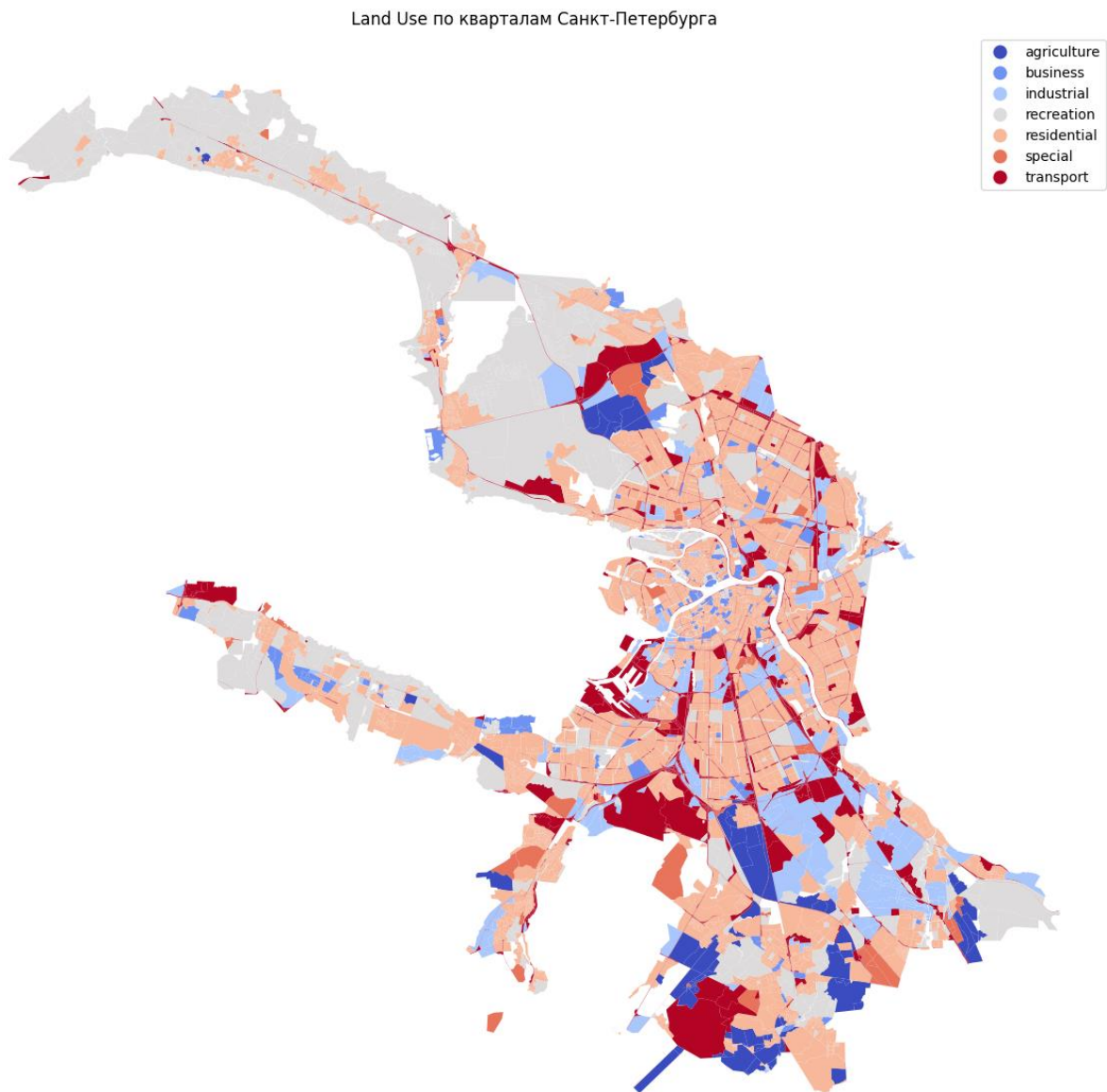


Рисунок 2 – Визуализация функционального зонирования кварталов

Источник: составлен авторами

Экспериментальной базой исследования послужили данные по земельным участкам городских территорий, собранные по Санкт-Петербургу. Данные кадастровой стоимости были получены из открытых данных Росреестра и городских кадастровых ведомств и с платформы Национальная система пространственных данных (НСПД) [16]. Набор данных по кадастровым участкам представлен на рисунке 3. Перенос кадастровой стоимости с уровня отдельных участков на квартальную модель позволяет увязать разнородные источники в единую структуру анализа. Квартал служит минимальной единицей territori-

ального планирования, в которой консолидируются площади участков, их кадастровая стоимость, параметры застройки (FSI, GSI, MXI), а также данные о населении и функциональном зонировании. Такая агрегация устраняет геометрические несоответствия между кадастровыми полигонами, заполняет пробелы в атрибутике ЕГРН и создаёт сопоставимую количественную базу для последующей регрессии, одновременно привязывая каждый квартал к вершинам транспортного графа для расчёта доступности, связанности и центральности кварталов для учета их топологических свойств.

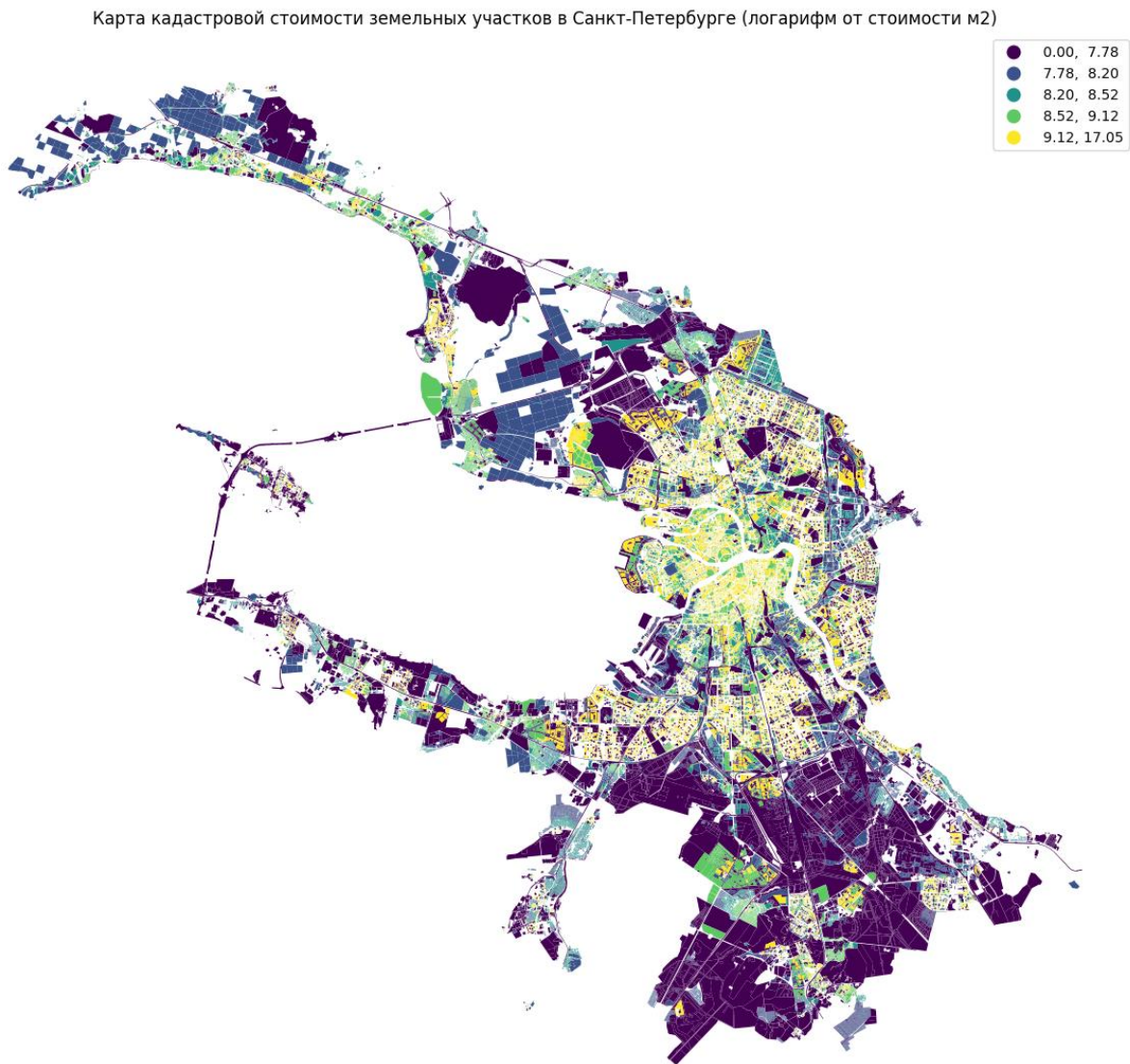


Рисунок 3 – Кадастровые участки по Санкт-Петербургу

Источник: составлен авторами

Кроме согласованности данных, квартал-сетевая модель даёт существенный выигрыш в производительности. Сокращая число объектных единиц на порядок по сравнению с участками, она ускоряет построение матриц доступности и массовые переоценки: время вычислений снижается

статистически значимо ($p \approx 3 \cdot 10^{-8}$) без заметной потери точности итоговых оценок ($p \approx 0,9$). Итоговый набор предобработанных данных и сводная статистика приведена в таблице 1. На рисунке 4 показана гистограмма распределения логарифма цены за квадратный метр.

Таблица 1

Набор исходных данных для обучения

Источник: составлена авторами

Признак	Count	Min	Max	Mean	Median
buid_floor_area	9 988	0.00	1,659,527.76	38,306.64	878.68
living_demand	4548	0.00	561.19	43.76	32.66
living_area	9 988	0.00	658,793.14	16,413.36	0.00

Признак	Count	Min	Max	Mean	Median
share_living	6013	0.00	21.02	1.22	0.70
business_area	9 988	0.00	1,659,527.76	21,893.28	446.53
share_business	6013	0.11	65.22	1.54	1.01
site_area	9 988	1.07	16,147,809.90	129,144.01	21,660.70
population	9 988	0.00	31,367.00	524.37	0.00
footprint_area	9 988	0.00	757,715.64	9,991.32	755.65
Fsi	9 988	0.00	6.03	0.30	0.02
gsi	9 988	0.00	0.98	0.09	0.02
l	6013	0.38	65.22	2.77	1.61
osr	6013	0.02	71,794.79	182.88	5.62
mxi	6013	0.00	0.90	0.38	0.46
land_value	9 386	127.20	31,200,003,381	420,336,608	54,844,085
area_sqm	9 988	1.07	16,147,809.90	129,144.01	21,660.70
accessibility	9 988	78.19	255.16	112.78	104.77

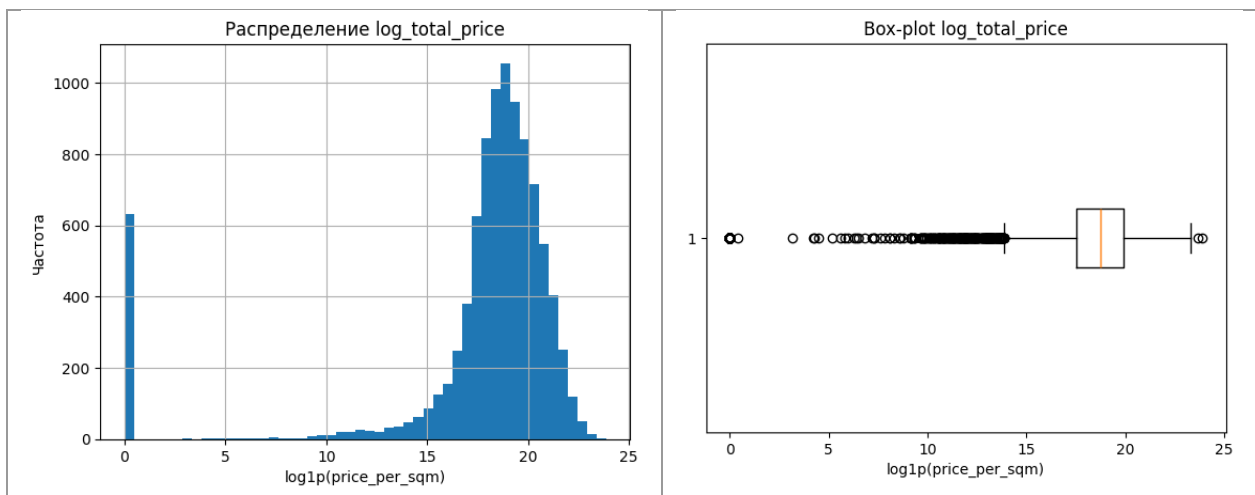


Рисунок 4 – Гистограмма распределения логарифма цены за м²

Источник: составлен авторами

Из графика видно, что примерно 620 кварталов имеют нулевую или отсутствующую оценку стоимости. Эти наблюдения были исключены, в результате чего итоговый рабочий датасет составил 9 386 кварталов. Такого объёма достаточно для обучения и вали-

дации нейросетевой модели, при этом дисбаланс классов по плотности и функциональному назначению может быть дополнительно скорректирован методами стратифицированного разбиения. Для проводимых экспериментов набор показателей достаточен.

Для детального анализа структуры пространственной зависимости была выполнена LISA-кластеризация (Local Indicators of Spatial Association – локальные индикаторы пространственной ассоциации), позволяющая выявить локальные кластеры и пространственные аномалии кадастровой стоимости (рисунок 5). Расчёт выполнялся на основе матрицы пространственных весов, отражаю-

щей смежность территориальных единиц. Глобальный индекс Морана позволил количественно подтвердить наличие положительной пространственной автокорреляции: значения кадастровой стоимости статистически значимо кластеризуются в пространстве. Это означает, что стоимость земельного участка в значительной степени зависит от характеристик соседних территорий.

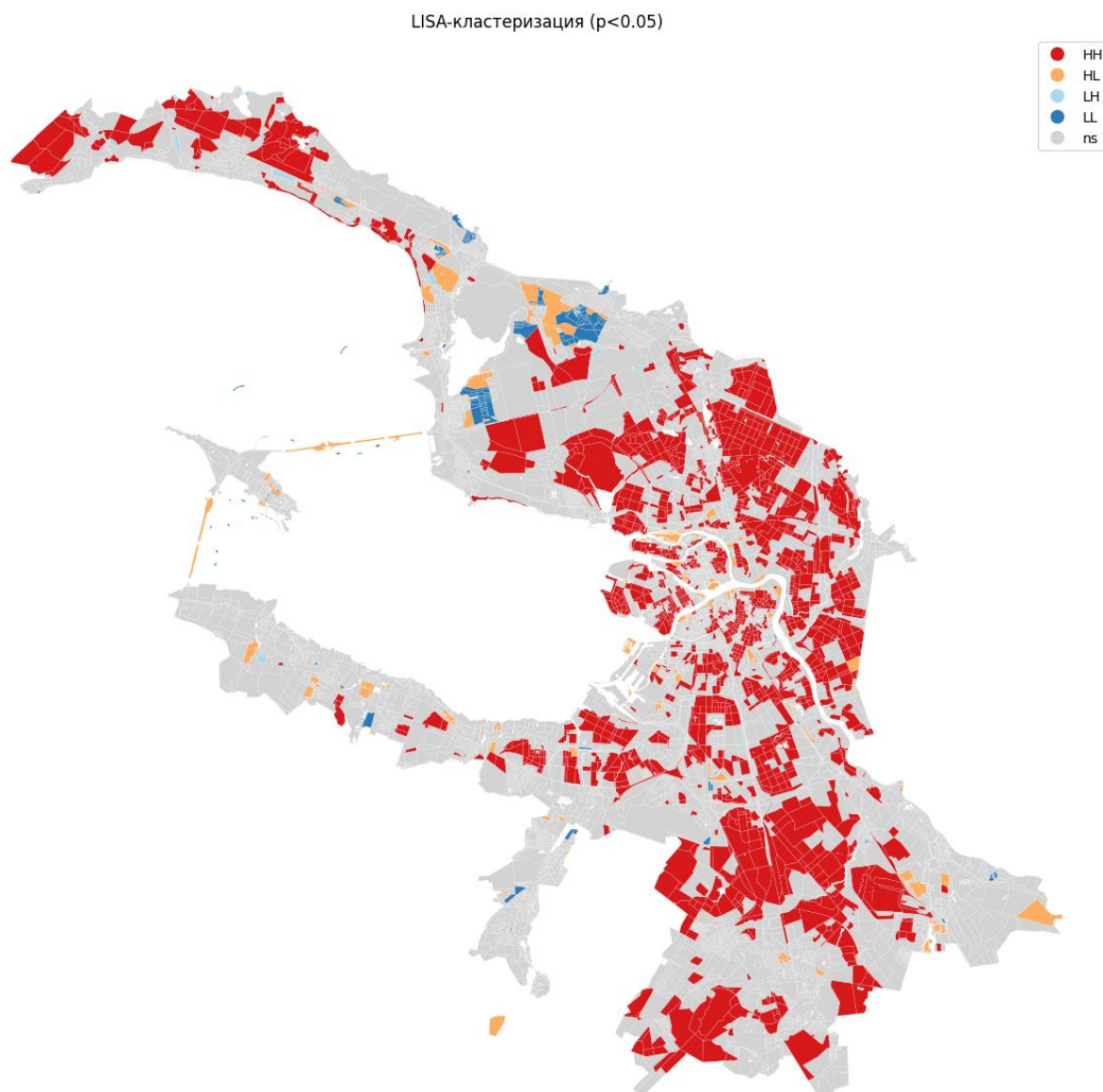


Рисунок 5 – LISA кластеризация кварталов

Источник: составлен авторами

Полученные результаты показывают преобладание кластеров типа High–High (HH), формирующих пространственно-связные зоны с высокой кадастровой стоимостью, а также локальные кластеры Low–Low (LL), соответствующие территориям с устойчиво низкими значениями показателя. Наличие

кластеров High–Low (HL) и Low–High (LH) отражает зоны пространственных разрывов и переходных территорий, где стоимость участка контрастирует с окружением. Выявленная структура пространственной автокорреляции подтверждает необходимость учёта пространственных лагов и контекстных

характеристик окружения при построении предсказательных моделей кадастровой стоимости. Игнорирование данных эффектов может приводить к смещению оценок и снижению устойчивости моделей, тогда как их включение позволяет более корректно описывать территориальные закономерности и повышать качество прогнозирования.

Описание метода и задействованных инструментов. С учётом постановки задачи прогнозирования числового значения кадастровой стоимости по набору табличных признаков, в качестве нейросетевой модели был выбран многослойный перцептрон (MLP) как базовая архитектура для регрессионного анализа структурированных данных. Выбор данной архитектуры обусловлен её способностью аппроксимировать нелинейные зависимости между признаками при умеренной вычислительной сложности.

Конфигурация сети подбиралась экспериментально с целью достижения баланса между точностью прогнозирования и устойчивостью к переобучению. Входной слой сети соответствует числу используемых признаков (17, таблица 1), далее следуют несколько скрытых полно-связных слоёв, выполняющих нелинейное преобразование

признакового пространства. Конфигурация сети (глубина и ширина слоёв) подбиралась экспериментально с учётом баланса между сложностью и обобщающей способностью. В скрытых слоях применялась активация ReLU, а для снижения переобучения использовались Dropout и L2-регуляризация весов, что обеспечивает более устойчивую аппроксимацию зависимостей на шумных и неоднородных данных. Схема архитектуры приведена на рисунке 6.

Программная реализация нейросети выполнена на Python 3.10 с использованием PyTorch [17]. Для сопоставления качества и устойчивости на тех же признаках были обучены два современных алгоритма градиентного бустинга – CatBoost [18] и XGBoost – как сильные табличные бенчмарки. Для подготовки данных, разбиения выборки, кросс-валидации и расчёта метрик использовался scikit-learn [19]; для интерпретации вкладов признаков – SHAP; для геопривязки и расчёта расстояний – GeoPandas [20]. Методологический дизайн эксперимента выстроен в логике воспроизводимого прикладного исследования: признаки формировались единообразно, код документировался, а результаты проверялись статистически.

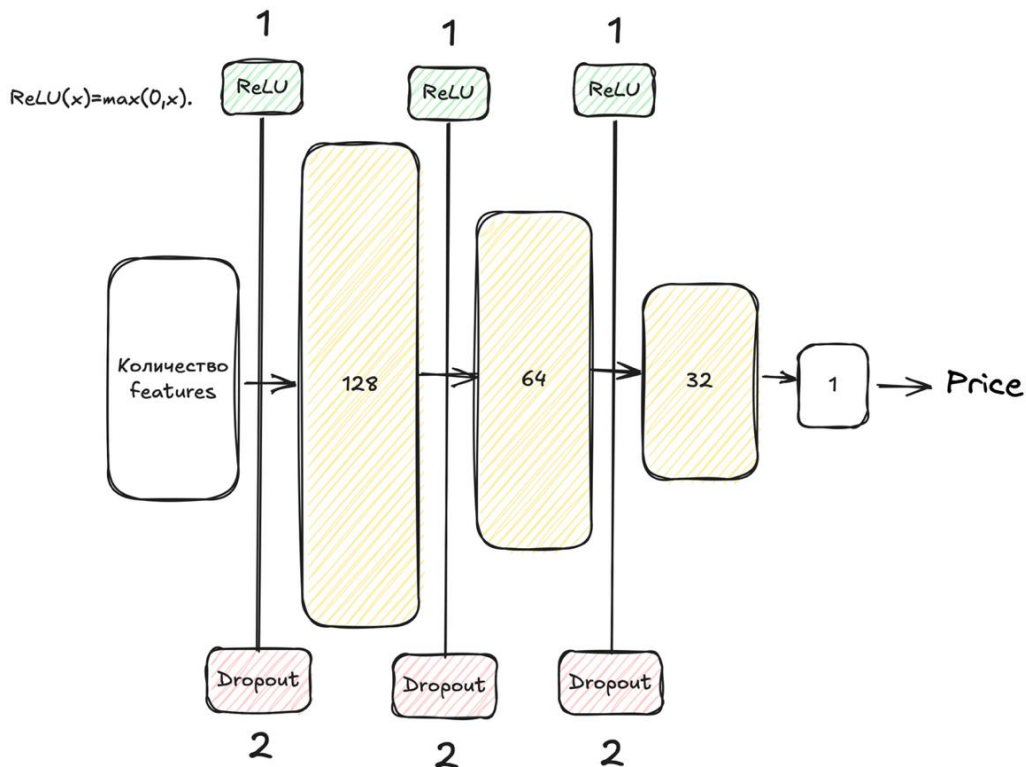


Рисунок 6 – Архитектура нейронной сети

Источник: составлен авторами

Обучение MLP проводилось итеративно по эпохам с контролем качества на валидационной части. В качестве функции потерь применялась MSE, оптимизация выполнялась алгоритмом Adam; дополнительно использовалась стратегия ранней остановки по метрикам R^2 , MAE и RMSE, что предотвращало избыточное обучение после достижения плато. Гиперпараметры нейросети (число слоёв, количество нейронов, Dropout, скорость обучения) подбирались с помощью Grid Search; итоговая конфигурация включала четыре скрытых слоя и Dropout = 0.3, при этом дальнейшее усложнение архитектуры не давало устойчивого прироста точности.

Для оценки качества прогнозирования использовались три ключевые метрики регрессии: коэффициент детерминации R^2 , средняя абсолютная ошибка

(MAE) и корень средней квадратичной ошибки (RMSE).

Для CatBoost и XGBoost гиперпараметры настраивались кросс-валидацией на обучающей выборке. CatBoost учитывает категориальные переменные встроенными механизмами кодирования, однако для сопоставимости всем моделям подавался единый набор подготовленных признаков. После обучения для каждой модели фиксировались предсказания на тестовом наборе и рассчитывались R^2 , MAE и RMSE на обучающей и тестовой выборках, что обеспечивало корректное сравнение нейросетевого подхода с ансамблями и одиночной древовидной моделью. Концептуальная схема разработанного метода с возможным вариантом дополнительной оценки инвестиционной привлекательности земли представлена на рисунке 7.

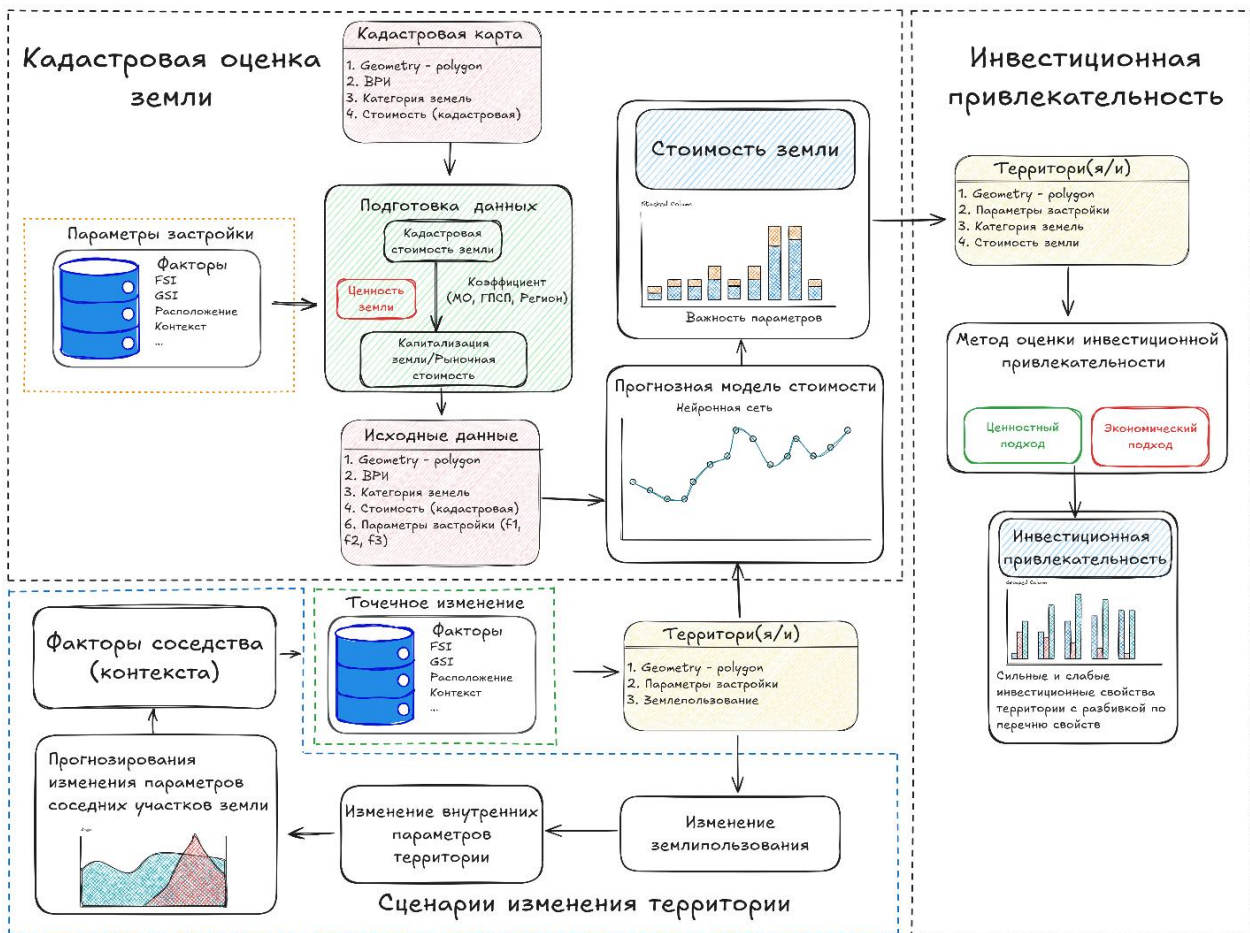


Рисунок 7 – Концептуальная схема работы метода автоматизированной оценки земли
 Источник: составлен авторами

Помимо оценки точности, важно понять, за счет каких факторов модель формирует предсказание стоимости – это повышает доверие к модели со стороны экспертов и

позволяет сделать выводы о ценообразующих факторах. В исследовании применен метод SHAP [22] для интерпретации результатов модели CatBoost. Для каждого отдельного

участка можно разложить прогнозируемую стоимость на сумму вкладов всех признаков, что показывает, какие характеристики повысили или понизили оценку относительно среднего уровня [9]. Были рассчитаны SHAP-значения для всех признаков на тестовой выборке, а также суммарные важности признаков, чтобы определить наиболее значимые факторы.

Результаты. В качестве базовых (baseline) моделей для оценки эффективности

предложенного нейросетевого подхода были использованы широко применяемые в задачах массовой оценки алгоритмы градиентного бустинга CatBoost и XGBoost, демонстрирующие устойчивые результаты на табличных пространственно-экономических данных и часто рассматриваемые как сильные бенчмарки в прикладных исследованиях. Итоговые значения метрик качества на тестовой выборке приведены в таблице 2.

Таблица 2

Сравнение качества моделей по основным метрикам

Источник: составлен авторами

Модель	Коэффициент детерминации R^2	Средняя абсолютная ошибка MAE	Корень средней квадратичной ошибки RMSE
Нейронная сеть (MLP)	0.76	3.85	5.20
CatBoost	0.73	3.61	4.96
XGBoost	0.71	3.75	5.10

Анализ результатов показывает, что различия в точности между моделями носят умеренный характер. Нейросетевая модель демонстрирует наибольшее значение коэффициента детерминации ($R^2 = 0,76$), однако преимущество по сравнению с CatBoost ($R^2 = 0,73$) и XGBoost ($R^2 = 0,71$) является не критичным с практической точки зрения. Значения MAE и RMSE для CatBoost сопоставимы, а в отдельных случаях даже ниже, чем у нейросетевой модели, что указывает на близкое качество прогнозов при существенно различной сложности моделей.

Несмотря на то, что нейросетевая модель показала немного более высокие значения R^2 , её применение в задачах кадастровой оценки сталкивается с существенным ограничением: низкой интерпретируемостью внутренних механизмов формирования прогноза. Даже при использовании методов пост-хок интерпретации нейросетевые архитектуры остаются трудными для формального объяснения в контексте регуляторных и управленческих процедур.

В отличие от этого, модель CatBoost обеспечивает баланс между высокой предсказательной способностью и интерпретируемостью. Использование SHAP-анализа для

CatBoost позволяет не только ранжировать факторы по степени влияния, но и проверить экономическую и градостроительную обоснованность прогнозов. Суммарные SHAP-значения признаков, представленные на рисунке 8, показывают, что наибольший вклад в формирование кадастровой стоимости вносят площадь территории, параметры плотности застройки, функциональное назначение и транспортная доступность. Эти результаты согласуются с устоявшимися представлениями в теории оценки земли и практике территориального планирования. Таким образом, CatBoost выбран в качестве основной модели для дальнейшей апробации и масштабирования, поскольку он обеспечивает достаточную точность при существенно более высокой прозрачности и объяснимости результатов по сравнению с нейросетевыми архитектурами.

Следующим этапом исследования стала проверка работы выбранной модели на новых данных, ранее не участвовавших в обучении, в частности на земельных участках Санкт-Петербурга с пропущенными исходными характеристиками. Результаты восстановления и оценки кадастровой стоимости для таких территорий представлены на рисунке 9.

Всего было получено прогнозных оценок для 694 участков, исключённых из обучающей выборки из-за неполноты данных. Анализ распределения прогнозных значений показал их реалистичность: модель не формирует аномально высоких или низких оценок, а диапазон результатов соответствует общей ценовой картине города. Это подтверждает способность модели обобщать выявленные закономерности и корректно работать в условиях неполной информации, что является типичной ситуацией для кадастровых данных. Пропущенные данные относятся к виду разрешенного использования – эксплуатация улично-дорожной сети и отсутствуют в данных НСПД по Санкт-Петербургу. Включение

пропущенных значений позволит повысить полноту охвата городских территорий данными по кадастровой оценке земельных участков.

В завершающей части исследования была проверена применимость разработанного подхода за пределами массива данных, на котором модель обучалась. Модель CatBoost, обученная на данных Санкт-Петербурга, была использована для прогнозирования кадастровой стоимости земельных участков на территории Ленинградской области. Пространственная дифференциация рассчитанных значений для Тихвинского и Сланцевского муниципальных районов представлена на рисунке 10.

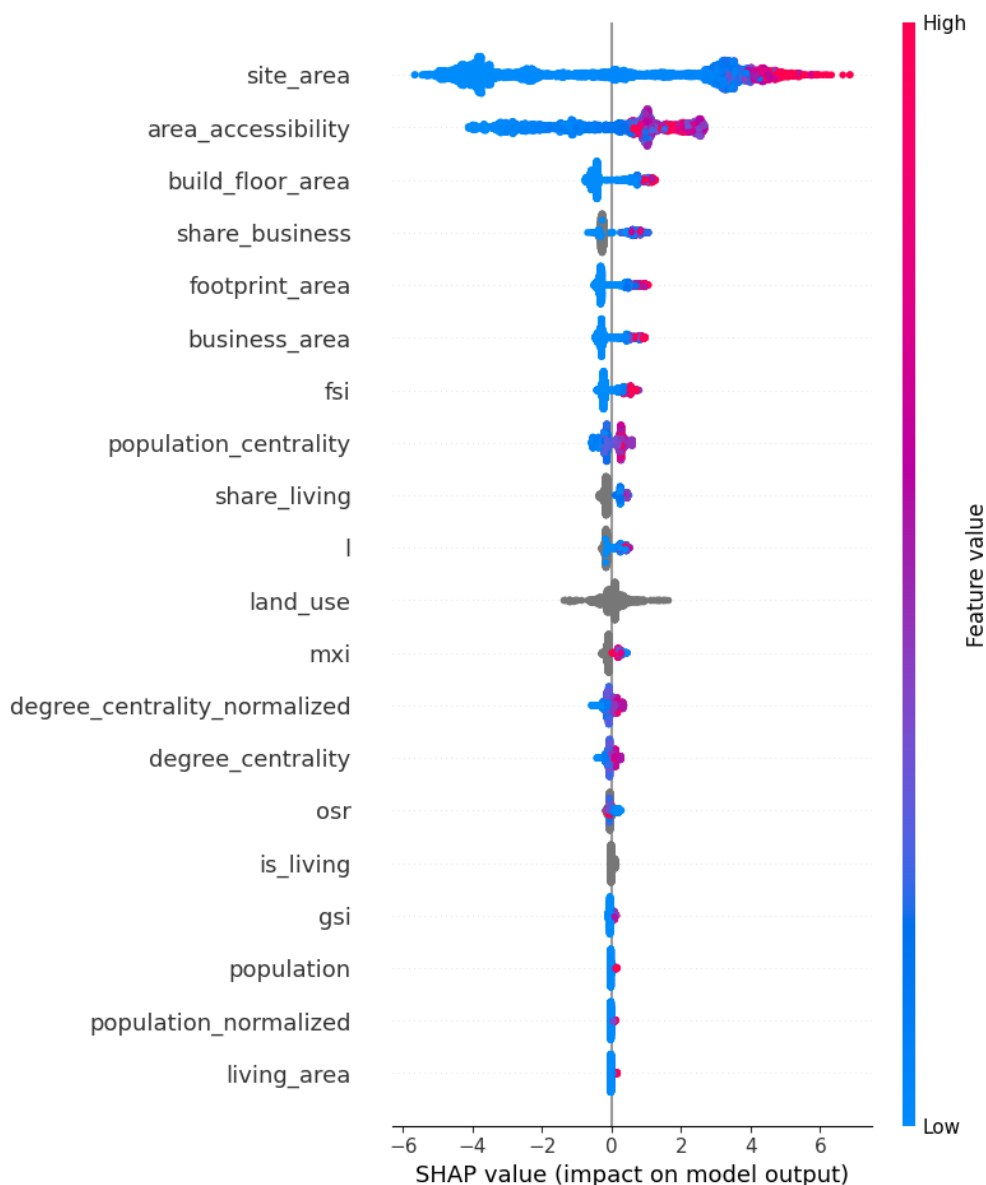


Рисунок 8 – SHAP анализ важности признаков

Источник: составлен авторами

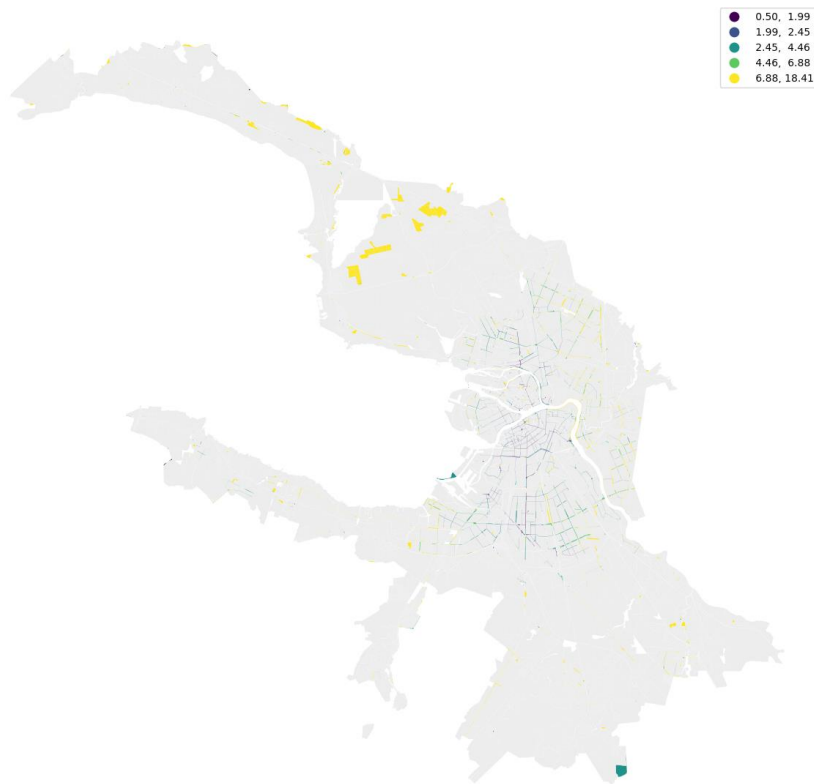


Рисунок 9 – Апробация модели на пропущенных данных
Источник: составлен авторами

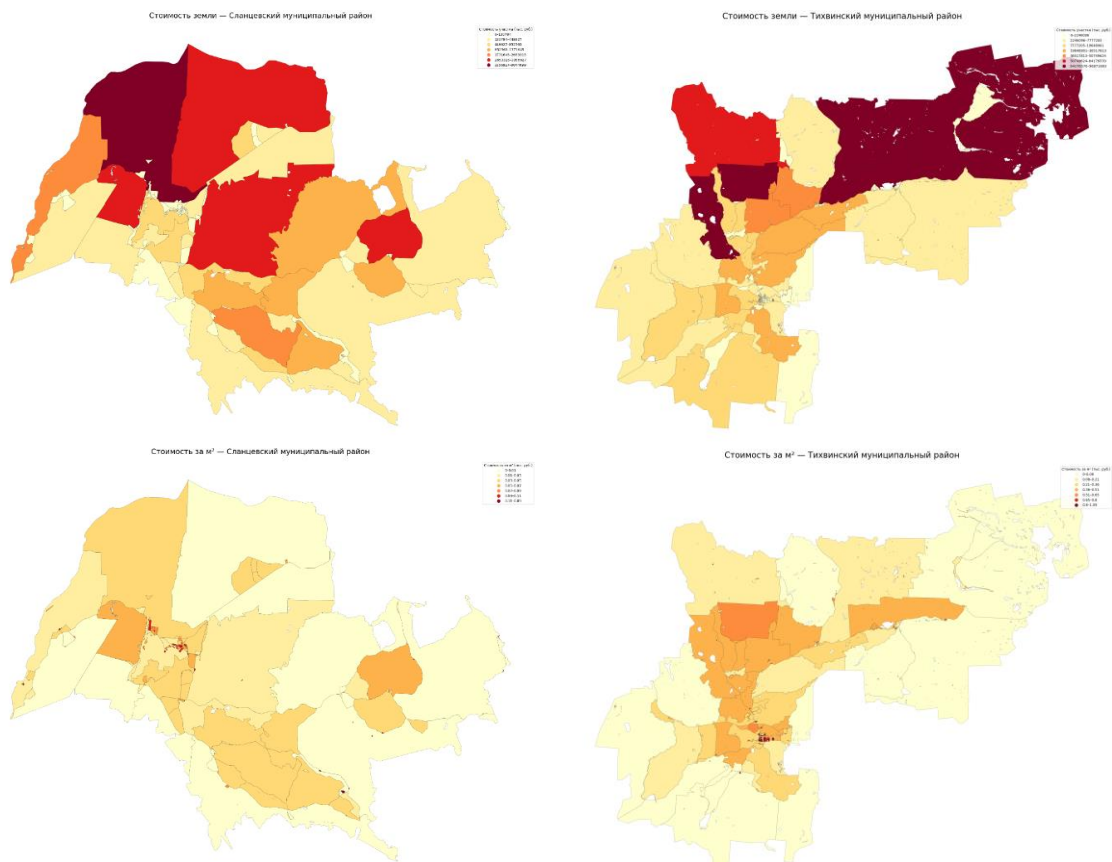


Рисунок 10 – Пространственная дифференциация кадастровой стоимости земельных участков в Тихвинском и Сланцевском муниципальном районе
Источник: составлен авторами

Границы кадастровых участков по Ленинградской области были получены с помощью библиотеки BlocksNet [9] методом нарезки исходной геометрии Ленинградской области линейными объектами улично-дорожной сети, железнодорожной сети и водными объектами. Полученные результаты отражают ожидаемые территориальные закономерности: более высокие прогнозные значения соответствуют участкам, расположенным вблизи городских поселений, транспортных узлов и промышленных центров, тогда как удалённые от инфраструктуры территории характеризуются более низкой стоимостью. Такая пространственная структура соответствует экономической логике и косвенно подтверждает корректность работы модели вне обучающей выборки.

Дополнительным преимуществом является масштабируемость решения: оценка большого числа объектов на региональном уровне выполняется с сопоставимой вычислительной эффективностью, что делает метод пригодным для задач регионального планирования и управления земельными ресурсами, где требуется оперативная и объяснимая оценка больших территорий.

Выводы и рекомендации. Исследование показало, что параметры планировочной структуры и застройки территории содержат достаточный объём информации для прогнозирования кадастровой стоимости земельных участков и даже при ограниченном наборе исходных признаков модели машинного обучения обеспечивают устойчивые и воспроизводимые результаты, пригодные для поддержки управленческих решений в сфере территориального планирования.

Сравнительный анализ различных классов моделей продемонстрировал, что нейросетевые архитектуры могут обеспечивать незначительно более высокие показатели точности, однако это преимущество носит ограниченный характер и сопровождается существенным снижением интерпретируемости. В условиях кадастровой оценки, выступающей элементом системы управления и регуляторной практики, приоритетное значение имеют

прозрачность, объяснимость и воспроизводимость результатов. С этой точки зрения ансамблевая модель CatBoost обеспечивает оптимальный баланс между точностью прогнозирования и возможностью интерпретации факторов влияния.

Применение методов Explainable AI для модели CatBoost показало, что структура прогнозов соответствует экономической и градостроительной логике: ключевыми ценообразующими факторами выступают площадь территории, параметры плотности застройки, функциональное назначение и транспортная доступность. Это подтверждает адекватность модели и повышает доверие к её использованию в практических задачах.

Практическая апробация метода продемонстрировала его применимость для массовой оценки и сценарного анализа. Модель успешно использована для восстановления кадастровых оценок на неполных данных в пределах Санкт-Петербурга и масштабирована на территорию Ленинградской области без необходимости повторного обучения. Это подтверждает устойчивость и масштабируемость предложенного подхода при работе с большими территориями.

Полученные результаты позволяют рекомендовать разработанный метод для:

- поддержки принятия решений органами муниципального и регионального управления при разработке проектов планировки территории и комплексного развития территорий;
- экспресс-оценки инвестиционной привлекательности земельных участков на ранних стадиях проектирования;
- интеграции в цифровые ГИС-платформы и аналитические сервисы в качестве интерпретируемого модуля оценки стоимости земли.

Перспективным направлением дальнейших исследований является использование больших языковых моделей для повышения интерпретируемости результатов прогнозирования кадастровой стоимости. Таким образом, полученные результаты позволяют сделать вывод о достижении поставленной цели исследования.

Список источников

1. Перечень поручений по итогам совещания с членами Правительства Российской Федерации (Москва, 6 сент. 2023 г.) № Пр-1770 [Электронный ресурс]. – URL: <https://www.kremlin.ru/acts/assignments/orders/72211> (дата обращения: 12.06.2025). – Текст: электронный.
2. Министерство цифрового развития, связи и массовых коммуникаций РФ. Методические рекомендации по цифровой трансформации государственных корпораций и компаний с государственным участием. – М., 2023. – 140 с. – URL: https://digital.gov.ru/uploaded/files/140020231228obnovlenniemetodicheskierekomendatsii_v12sokraschennyie-1.pdf (дата обращения: 12.06.2025). – Текст: электронный.
3. Федеральный закон Российской Федерации от 30 декабря 2020 г. № 494-ФЗ «О внесении изменений в Градостроительный кодекс Российской Федерации и отдельные законодательные акты Российской Федерации в целях обеспечения комплексного развития территорий // СПС «КонсультантПлюс» [Электронный ресурс]. – URL: https://www.consultant.ru/document/cons_doc_LAW_372677/ (дата обращения: 12.06.2025). – Текст: электронный.
4. Градостроительный кодекс Российской Федерации от 29 декабря 2004 г. № 190-ФЗ. Гл. 10 «Комплексное развитие территории» (ред. от 26 декабря 2024 г.) // СПС «КонсультантПлюс» [Электронный ресурс]. – URL: https://www.consultant.ru/document/cons_doc_LAW_51040/381915d21931faa78bc4ef1508a48dfe972fc537/ (дата обращения: 12.06.2025). – Текст: электронный.
5. Методические рекомендации по комплексному развитию территорий жилой застройки [Электронный ресурс]. – URL: <https://www.minstroyrf.ru/docs/225071/> (дата обращения: 12.06.2025). – Текст: электронный.
6. Постановление Правительства Москвы от 24 октября 2023 г. № 2035-ПП «Об утверждении Положения об особенностях состава, порядка подготовки и согласования проектов планировки территории в целях комплексного развития территории» (с изм. на 19 марта 2025 г.) [Электронный ресурс]. – URL: <https://docs.cntd.ru/document/1303675242> (дата обращения: 12.06.2025). – Текст: электронный.
7. Федеральный закон Российской Федерации от 03 июля 2016 г. № 237-ФЗ «О государственной кадастровой оценке» (ред. от 23 ноября 2024 г.) // СПС «КонсультантПлюс» [Электронный ресурс]. – URL: https://www.consultant.ru/document/cons_doc_LAW_200504/ (дата обращения: 12.06.2025). – Текст: электронный.
8. Методические указания о государственной кадастровой оценке: утв. приказом Росреестра от 04 августа 2021 г. № П/0336 (с изм. на 11 сентября 2024 г.) [Электронный ресурс]. – URL: <https://docs.cntd.ru/document/726730589> (дата обращения: 12.06.2025). – Текст: электронный.
9. Gnat S. Determining the Influence of Real Estate Features on Prices with Partial Dependence Plots: A Case Study in Szczecin, Poland // *Real Estate Management and Valuation*. – 2024. – Vol. 32. – No. 4. – pp. 105–116. – DOI: 10.2478/remav-2024-0039. (In Eng.).

References

1. New Decree of the Government of the Russian Federation (Moscow, September 6, 2023) No. Pr-1770. [Electronic resource]. URL: <https://www.kremlin.ru/acts/assignments/orders/72211> (Accessed: 12.06.2025) (In Russ.).
2. Ministry of Digital Development, Communications and Mass Media of the Russian Federation. Methodological Recommendations on the Digital Transformation of State Corporations and Companies with State Participation. *Moscow, 2023*. 140 p. URL: https://digital.gov.ru/uploaded/files/140020231228_obnovlenniemetodicheskierekomendatsiiv12sokraschennyie-1.pdf (Accessed: 12.06.2025) (In Russ.).
3. Federal Law of the Russian Federation No. 494-FZ dated December 30, 2020 «On Amendments to the Urban Planning Code of the Russian Federation and Certain Legislative Acts of the Russian Federation in order to ensure the Integrated Development of Territories». *SPS «KonsultantPlyus»*. URL: https://www.consultant.ru/document/cons_doc_LAW_372677/ (Accessed: 12.06.2025) (In Russ.).
4. Urban Planning Code of the Russian Federation No. 190-FZ dated December 29, 2004, Chapter 10 «Integrated Development of the Territory» (as amended on December 26, 2024). *SPS «KonsultantPlyus»*. URL: https://www.consultant.ru/document/cons_doc_LAW_51040/381915d21931faa78bc4ef1508a48dfe972fc537/ (Accessed: 12.06.2025) (In Russ.).
5. Methodological Recommendations for the Integrated Development of Residential Areas [Electronic resource]. URL: <https://www.minstroyrf.ru/docs/225071/> (Accessed: 12.06.2025) (In Russ.).
6. Decree of the Government of Moscow dated October 24, 2023 No. 2035-PP «On Approval of the Regulations on the Specifics of the Composition, Procedure for the Preparation and Approval of Territorial Planning Projects for the Integrated Development of the Territory» (as amended on March 19, 2025) [Electronic resource]. URL: <https://docs.cntd.ru/document/1303675242> (Accessed: 12.06.2025) (In Russ.).
7. Federal Law of the Russian Federation No. 237-FZ dated July 03, 2016 «On State Cadastral Valuation» (as amended on November 23, 2024). *SPS «KonsultantPlyus»*. URL: https://www.consultant.ru/document/cons_doc_LAW_200504 (Accessed: 12.06.2025) (In Russ.).
8. Methodological Guidelines on the State Cadastral Assessment: Approved by Rosreestr Order No. P/0336 dated August 04, 2021 (as amended on September 11, 2024) [Electronic resource]. URL: <https://docs.cntd.ru/document/726730589> (Accessed: 12.06.2025) (In Russ.).
9. Gnat S. Determining the Influence of Real Estate Features on Prices with Partial Dependence Plots: A Case Study in Szczecin, Poland. *Real Estate Management and Valuation*. 2024. Vol. 32. No. 4. pp. 105–116. DOI: 10.2478/remav-2024-0039.

10. Chen C., Ma X., Zhang X. Empirical Study on Real Estate Mass Appraisal Based on Dynamic Neural Networks // *Buildings*. – 2024. – Vol. 14. – No. 7. – Art. 2199. – DOI: 10.3390/buildings14072199. – Текст: электронный. (In Eng.).
11. Viktoratos I., Tsadiras A. Advancing Real-Estate Forecasting: A Novel Approach Using Kolmogorov–Arnold Networks // *Algorithms*. – 2025. – Vol. 18. – № 2. – Art. 93. – DOI: 10.3390/a18020093. – Текст: электронный. (In Eng.).
12. Lin Deng. Real Estate Valuation with Multi-Source Image Fusion and Enhanced Machine Learning Pipeline // *Plos.one*. – 2024. – DOI: 10.1371/journal.pone.0321951. – Текст: электронный. (In Eng.).
13. Бессмертный И. А. Основы научных исследований в области информационных систем и технологий [Электронный ресурс]. – М.: Юрайт, 2020. – URL: <https://urait.ru/book/osnovy-nauchnyh-issledovaniy-v-oblasti-informacionnyh-sistem-i-tehnologiy-580150> (дата обращения: 12.06.2025). – Текст: электронный.
14. Лосева Е. Н., Митрофанова Н. В. О применении искусственных нейронных сетей при государственной кадастровой оценке земельных участков // *Вестник СГУГиТ*. – 2021. – № 5. – С. 180–190. – DOI: 10.33764/2411-1759-2021-26-5-180-190. – Текст: электронный.
15. Mete M. O., Yomralioglu T. A Hybrid Approach for Mass Valuation of Residential Properties Through Geographic Information Systems and Machine Learning Integration // *Geographical Analysis*. – 2022. – Vol. 55. – No. 4. – P. 535–559. – DOI: 10.1111/gean.12350. – Текст: электронный. (In Eng.).
16. Единая цифровая платформа «Национальная система пространственных данных» (ЕЦП НСПД) [Электронный ресурс]. – URL: <https://nspd.gov.ru/> (дата обращения: 12.06.2025).
17. Imambi S., Prakash K. B., Kanagachidambaresan G. R. PyTorch // *Programming with TensorFlow: solution for edge computing applications*. – 2021. – P. 87–104. – DOI: 10.1007/978-3-030-57077-4_10. – Текст: электронный. (In Eng.).
18. Prokhorenkova L., Gusev G., Vorobev A., Dorogush A.V., Gulin A. CatBoost: Unbiased Boosting with Categorical Features // *Advances in Neural Information Processing Systems*. – 2018. – Vol. 31. – P. 117–127. (In Eng.).
19. Kramer O. Scikit-Learn // *Studies in Big Data*. Cham: Springer. – 2016. – Vol. 20. – P. 45–53. (In Eng.).
20. Jordal K., et al. Geopandas/geopandas: v0.5.0 // *Zenodo* [Электронный ресурс]. – URL: <https://zenodo.org/records/2705946> (дата обращения: 27.04.2025). – Текст: электронный. (In Eng.).
21. OpenStreetMap [Электронный ресурс]. – URL: <https://www.openstreetmap.org/> (дата обращения: 12.06.2025).
22. aimclub/blocksnet: Open Library with Tools for Generation the City Model and Optimal Requirements for Future Development with Specified Target Parameters // *GitHub* [Электронный ресурс]. – 2023. – URL: <https://github.com/aimclub/blocksnet/> (дата обращения: 28.05.2025). (In Eng.).
10. Chen C., Ma X., Zhang X. Empirical Study on Real Estate Mass Appraisal Based on Dynamic Neural Networks. *Buildings*. 2024. Vol. 14. No. 7. Art. 2199. DOI: 10.3390/buildings14072199.
11. Viktoratos I., Tsadiras A. Advancing Real-Estate Forecasting: A Novel Approach Using Kolmogorov–Arnold Networks. *Algorithms*. 2025. Vol. 18. № 2. Art. 93. DOI: 10.3390/a18020093.
12. Lin Deng. Real Estate Valuation with Multi-Source Image Fusion and Enhanced Machine Learning Pipeline. *Plos.one*. 2024. DOI: 10.1371/journal.pone.0321951.
13. Bessmertny I. A. Fundamentals of Scientific Research in the Field of Information Systems and Technologies [Electronic resource]. *Moscow: Yurait*, 2020. URL: <https://urait.ru/book/osnovy-nauchnyh-issledovaniy-v-oblasti-informacionnyh-sistem-i-tehnologiy-580150> (Accessed: 12.06.2025) (In Russ.).
14. Loseva E. N., Mitrofanova N. V. On the Use of Artificial Neural Networks in the State Cadastral Valuation of Land Plots. *Vestnik SGUGiT*. 2021. No. 5. pp. 180–190. DOI: 10.33764/2411-1759-2021-26-5-180-190. (In Russ.).
15. Mete M. O., Yomralioglu T. A Hybrid Approach for Mass Valuation of Residential Properties Through Geographic Information Systems and Machine Learning Integration. *Geographical Analysis*. 2022. Vol. 55. No. 4. pp. 535–559. DOI: 10.1111/gean.12350.
16. Unified Digital Platform «National Spatial Data System» [Electronic resource]. URL: <https://nspd.gov.ru/> (Accessed: 12.06.2025) (In Russ.).
17. Imambi S., Prakash K. B., Kanagachidambaresan G. R. PyTorch. *Programming with TensorFlow: solution for edge computing applications*. 2021. pp. 87–104. DOI: 10.1007/978-3-030-57077-4_10.
18. Prokhorenkova L., Gusev G., Vorobev A., Dorogush A.V., Gulin A. CatBoost: Unbiased Boosting with Categorical Features. *Advances in Neural Information Processing Systems*. 2018. Vol. 31. P. 117–127.
19. Kramer O. Scikit-Learn. *Studies in Big Data*. Cham: Springer. 2016. Vol. 20. P. 45–53.
20. Jordal K., et al. Geopandas/geopandas: v0.5.0. *Zenodo* [Electronic resource]. URL: <https://zenodo.org/records/2705946> (Accessed: 27.04.2025).
21. OpenStreetMap [Electronic resource]. URL: <https://www.openstreetmap.org/> (Accessed: 12.06.2025) (In Russ.).
22. aimclub/blocksnet: Open Library with Tools for Generation the City Model and Optimal Requirements for Future Development with Specified Target Parameters. *GitHub* [Electronic resource]. 2023. URL: <https://github.com/aimclub/blocksnet/> (Accessed: 28.05.2025)