

АНАЛИЗ, МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ ПРОЦЕССОВ УПРАВЛЕНИЯ В СОЦИАЛЬНЫХ ЭКОНОМИЧЕСКИХ СИСТЕМАХ

УДК 332.85

СТАТИСТИЧЕСКИЙ АНАЛИЗ ДИНАМИКИ СТОИМОСТИ КВАРТИР НА ВТОРИЧНОМ РЫНКЕ ЖИЛОЙ НЕДВИЖИМОСТИ ГОРОДА МОСКВЫ

Коды JEL: R 31, C 61, C 82

Вавилина А. В., кандидат экономических наук, доцент, доцент кафедры менеджмента, Российский университет дружбы народов имени Патриса Лумумбы (РУДН), г. Москва, Россия
E-mail: vavilina-av@rudn.ru; SPIN-код: 8939-4536

Ханенков П. В., магистр группы ДМЭ-101 им.цэ факультета экономики, Московский финансово-промышленный университет «Синергия», г. Москва, Россия
E-mail: phanenkou@mail.ru

Цыпин А. П., кандидат экономических наук, доцент, доцент департамента бизнес-аналитики, факультет налогов, аудита и бизнес-анализа, Финансовый университет при правительстве Российской Федерации, г. Москва, Россия
E-mail: aptsy-pin@fa.ru; SPIN-код: 1214-7508

Поступила в редакцию 15.06.2023. Принята к публикации 26.06.2023

Аннотация

Актуальность темы. Необходимость отслеживания изменений стоимости жилой недвижимости и определения наилучшего момента для приобретения вторичной недвижимости в г. Москва приобретает особую значимость в современных условиях.

Цель. Создание модели для описания влияния факторов на вариацию стоимости недвижимости на вторичном рынке жилья г. Москва.

Методология. Для создания модели использованы подходы машинного обучения, для оценки ее работы использована классическая метрика точности — квадратный корень из среднеквадратичной ошибки между тестовой выборкой, не участвующей в обучении модели, и предсказанием модели.

Результаты и выводы. Приобретать недвижимость целесообразно в июне, так как в июле начинается резкий рост после спада в предшествующем месяце. Каждый год цены неизменно растут, поэтому недвижимость в г. Москве — это хороший вариант для инвестиций.

Область применения. Сфера регулирования рынка жилья на общестрановом и локальном уровнях. Исследования в области моделирования временных рядов и сложных экономических систем.

Направления дальнейшего исследования: работы по повышению точности модели за счёт функционально-конструкторского подхода и использования дополнительных переменных, описывающих наполнение квартир, выставяемых на продажу. Также добавление в модель корректировки на локальное расположение объекта недвижимости в Москве.

Ключевые слова: недвижимость, вторичный рынок, цена жилья, факторы, моделирование.

STATISTICAL ANALYSIS OF THE DYNAMICS OF THE COST OF APARTMENTS IN THE SECONDARY RESIDENTIAL REAL ESTATE MARKET OF MOSCOW

JEL Codes: R 31, C 61, C 82

Vavilina A. V., Candidate of Economic Sciences, Associate Professor, Associate Professor, Patrice Lumumba Russian University of Peoples' Friendship (RUDN), Moscow, Russia

E-mail: vavilina-av@rudn.ru; SPIN code: 8939-4536

Khanenkov P. V., Master of the group DME-101imce, Faculty of Economics, Moscow Financial and Industrial University «Synergy», Moscow, Russia

E-mail: phanenkov@mail.ru

Tsy-pin A. P., Candidate of Economic Sciences, Associate Professor, Associate Professor, Business Intelligence Department. Faculty of Taxes, Audit and Business Analysis, Financial University under the Government of the Russian Federation, Moscow, Russia

E-mail: aptsypin@fa.ru; SPIN code: 1214-7508

Annotation

The relevance of the topic. *The need to track changes in the value of residential real estate and determine the best moment to purchase secondary real estate in Moscow is of particular importance in modern conditions.*

Goal. *Creation of a model to describe the influence of factors on the variation in the value of real estate in the secondary housing market in Moscow.*

Methodology. *Machine learning approaches were used to create the model, and the classical accuracy metric was used to evaluate its performance — the square root of the root-mean-square error between the test sample not participating in the training of the model and the prediction of the model.*

Results and conclusions. *It is advisable to purchase real estate in June, since a sharp increase begins in July after a decline in the previous month. Prices are steadily rising every year, so real estate in Moscow is a good option for investment.*

Scope of application. *The scope of regulation of the housing market at the national and local levels. Research in the field of modeling time series and complex economic systems.*

Directions for further research: *work to improve the accuracy of the model through a functional design approach and the use of additional variables describing the filling of apartments for sale. Also adding adjustments to the model for the local location of the property in Moscow.*

Keywords: *real estate, secondary market, housing price, factors, modeling.*

DOI: 10.22394/1997-4469-2023-60-1-165-172

Введение

Рынок жилой недвижимости Москвы очень динамичен и его изучение позволяет определять наиболее подходящее время для приобретения объекта жилья, что может оказать существенную экономю средств для рядового покупателя.

Поскольку Москва является столицей Российской Федерации, то спрос на недвижимость в ней постоянно растёт, но тем не менее есть определенные периоды внутри года, когда стоимость снижается.

Стоит отметить, что разработка вопроса статистического изучения (моделирования) рынка жилья, уже проводилась нами ранее. Так в ряде научных статей [8, 9], для анализа рынка жилья, нами были использованы такие методы как табличный и графический, корреляци-

онный анализ, а также регрессионный анализ с учетом местоположения объекта недвижимости (географически взвешенная регрессия).

В качестве недостатка примененного подхода можно назвать тот факт, что использованное регрессионное уравнение является классической «сильной» моделью, но у неё отсутствует градиентный спуск, который позволяет минимизировать размер ошибки, поэтому в настоящем исследовании, предпочтение отдадим моделям с этой функцией.

Также при подготовке материалов статьи, были проанализированы работы таких российских ученых как: Богданова Т. К. [6], Грахов В. П. [1], Мамедли М. О. [4], Осадчий Н. К. [5], Пяткина Д. А. [7]. Мнения и рекомендации указанных исследователей были учтены в проводимом исследовании.

Материалы и методы

Для проведения исследования были взяты данные из открытого источника — платформы для соревнования дата-саентистов (Data Scientist) Kaggle.com [2].

В используемом наборе данных содержится 5,4 млн записей, собранных с сайтов по продаже недвижимости с 2018 г. по 2021 г.

Предварительный (разведочный) анализ этого набора данных привел нас к пониманию, что использовать его для построения модели не представляется возможным, по причине наличия «выбросов», поэтому была осуществлена очистка, выделены показатели (минимальная и максимальная площадь квартиры и кухни, а также цена недвижимости) и жилые объекты, относящиеся к г. Москва. При этом тип рынка был выбран как — «вторичная» недвижимость.

Площадь квартиры $\in [20; 120]$ кв. метров, площадь кухни $\in [6; 30]$ кв. метров, цена квартиры $\in [1,5; 50,0]$ млн руб.

Также были введены дополнительные агрегаты-переменные для повышения точности модели: даты разделены на месяц и года; определено отношение этажа квартиры ко всей этажности дома; определена площадь, в среднем выделяемая под каждую комнату.

Весь программный код, по написанию модели и работе с данными можно найти по ссылке [3], он был реализован на языке программирования Python.

Стоит заметить, что в первом приближении к динамике средних цен по Москве, было очевидно, что цены «нереалистично» низкие (3—5 млн рублей) и присутствует высокая колеблемость, что не характерно для поведения цен на московскую недвижимость. При более подробном изучение набора данных, было установлено, что в рассматриваемую совокупность были включены города, находящиеся за пределами Московской области.

Соответственно, обучать модель на подобном наборе данных не имеет смысла, так как они не отражают реальное состояние и динамику рынка жилья [4]. Для решения этой проблемы была проведена группировка, так вокруг Москвы был описан квадрат по долготе и широте. Координаты, для формирования условий отбора объектов, были определены следующим образом: Московский Кремль был взят за центр, зафиксирована самая восточная и самая северная точки города, и за счёт разниц широты и долготы боковых точек с центральной, был получен квадрат. При этом объекты жилой недвижимости, находящиеся в «новой» Москве, не вошли в анализ.

Указанные манипуляции с исходным набором данных необходимы для улучшения обучающей способности модели, так как цель модели — найти закономерность в данных и использовать ее для построения прогнозов. Если данные, на которых модель обучается, будут слишком разнородными, то она не сможет «уловить» фактическую закономерность, что существенно повысит уровень ошибки при тестировании модели на тестовых данных.

Стоит заметить, что в качестве тестовых данных была использована подвыборка из всей совокупности данных, на которых модель не обучалась, в отличие от обучающей выборки. Размер тестовой выборки составил 20 % объектов от исходной совокупности.

Таким образом в итоговый набор данных вошли 16 133 объектов жилой недвижимости, тестовая выборка составила 3227 квартир г. Москва.

Результаты исследования

В результате манипуляции с исходной совокупностью данных, была получена динамика средних цен на вторичном рынке жилья, представленная на рисунке 1.

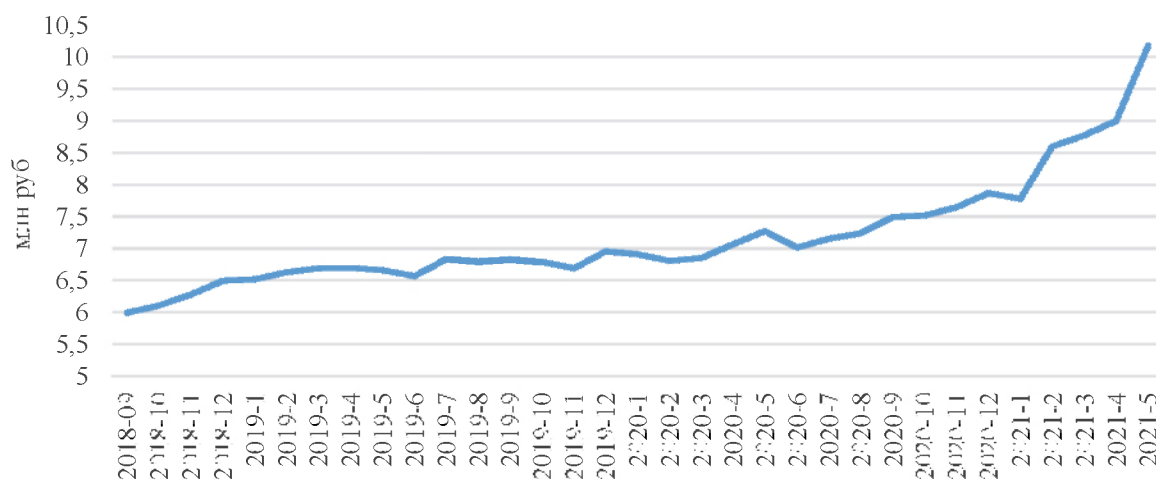


Рис. 1. Динамика средних цен на вторичную недвижимость в Москве за период сентябрь 2018 — май 2021, млн руб. (Источник: построено на основе данных портала Kaggle.com)

Анализируя поведение цен на московском вторичном рынке жилья (рисунок 1), можно сделать ряд выводов:

1) цены на недвижимость в Москве растут непрерывно, каждый год, это означает, что чем на более раннем сроке покупатель приобретёт недвижимость, тем дешевле она для него будет стоить (даже с учётом выплаты процентов банку, потому что они могут быть меньше 1 млн в год);

2) вторая половина года всегда дороже первой;

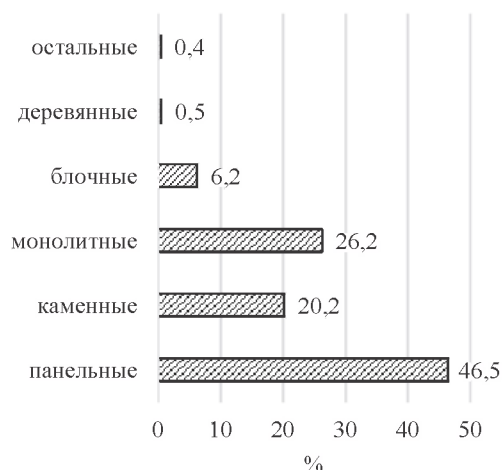
3) оптимальное время для покупки недвижимости — это июнь, так как цены в этом месяце «проседают».

Помимо динамики, полученный набор данных, позволяет провести анализ структуры по

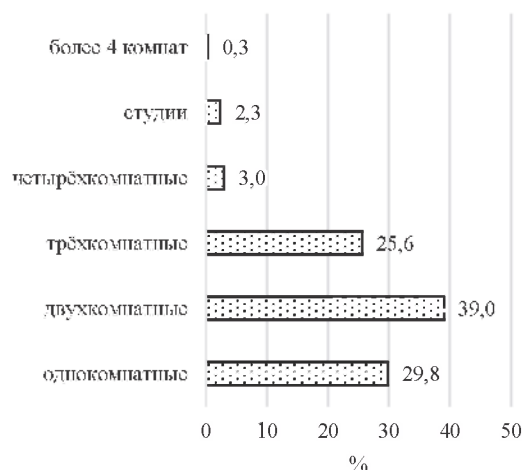
таким признакам как: тип строения; количество комнат; распределение площади квартиры и площади кухни; этажность.

Для построения диаграмм структуры вторичного рынка жилья (рисунок 2) нами была использована библиотека `matplotlib.pyplot` языка программирования Python, в частности команда `matplotlib.pyplot.pie()`

Согласно данным приведенным на рисунке 2, преобладают панельные дома, затем каменные, потом монолитные, блочные и деревянные. По убыванию количество недвижимости по параметру количества комнат: две комнаты, одна комната, 3 комнаты, 4 комнаты, студии и более 4 комнат.



А) Распределение по типам строения дома



Б) Распределение по количеству комнат

Рис. 2. Структуры московского рынка жилья, % (Источник: построено на основе данных портала Kaggle.com)

Далее обратимся к библиотеке `matplotlib.pyplot` и команде `matplotlib.pyplot.hist()` для по-

строения диаграммы распределения объектов недвижимости по площади (рисунок 3).

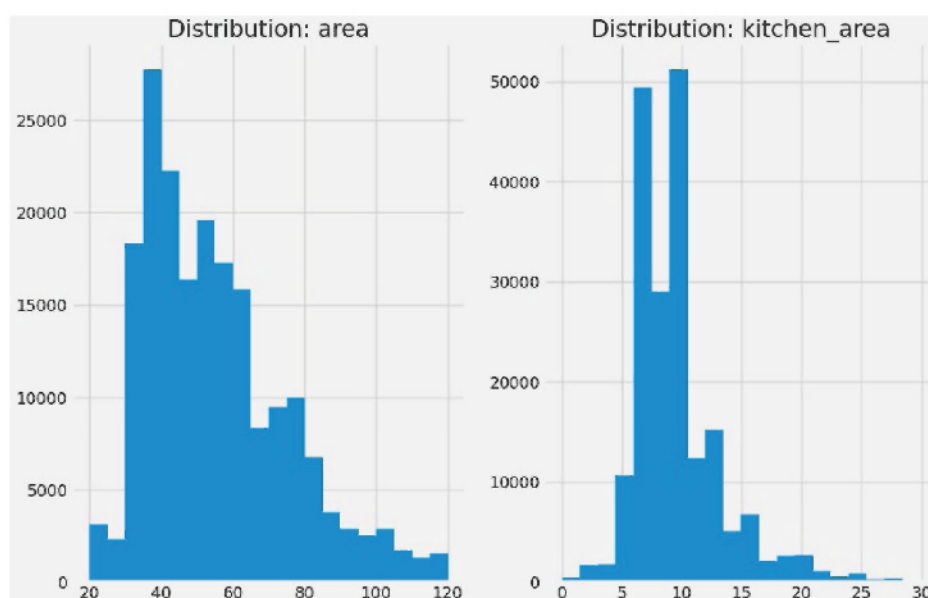


Рис. 3. Распределение по площади недвижимости и площади кухни, кв. м (Источник: построено на основе данных портала Kaggle.com)

Как видно из представленного распределения данных на рисунке 3, левый и правый повторяют друг друга, кроме недвижимости с площадью 50—60 кв. метров. Это значит, что свойственная ей площадь кухни в 10 кв. метров находится и в других строениях.

Аналогичным образом были построены диаграммы распределения по количеству этажей

домов в которых находятся объекты недвижимости и номеру этажа (рисунок 4).

Как следует из представленного распределения данных (рисунок 4), на левом графике, большинство людей предпочитает покупать квартиру до 10 этажа, это значит, что и её стоимость при продаже будет выше. Из правого графика можно понять, что множество домов 4, 10, 12, 14, 18-этажны.

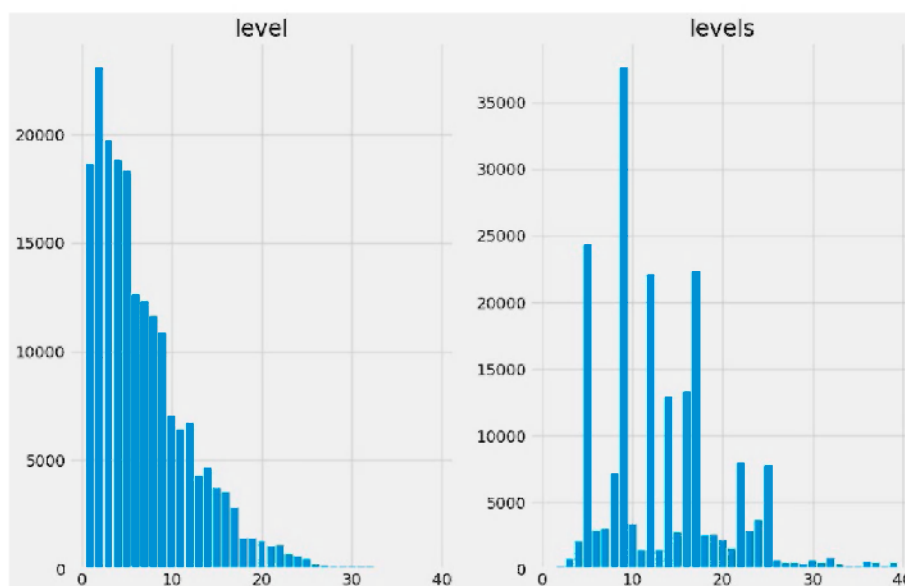


Рис. 4. Распределение по номеру этажа квартиры и этажности домов (Источник: построено на основе данных портала Kaggle.com)

Для построения парных линейных коэффициентов корреляции, между признаками объектов жилой недвижимости, была исполь-

зована библиотека seaborn и команда `seaborn.heatmap(correlation)`. Результаты работы алгоритма приведены на рисунке 5.

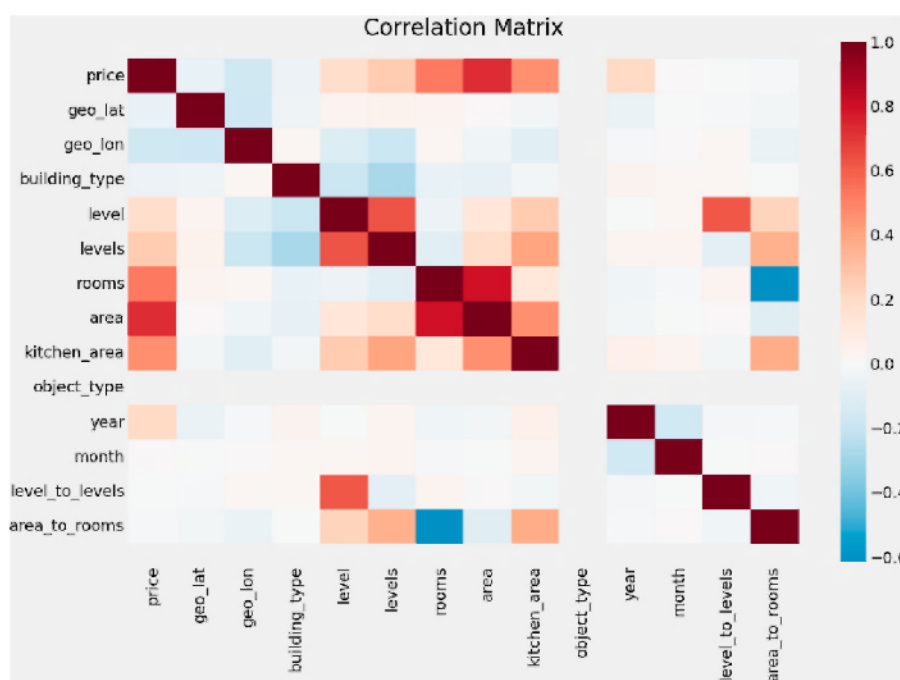


Рис. 5. Корреляционная матрица признаков, характеризующих объекты недвижимости на вторичном рынке Москвы (Источник: построено на основе данных портала Kaggle.com)

Согласно значений матрицы парных коэффициентов корреляции (рисунок 5), не было обнаружено каких-то необычных линейных корреляций между переменными, ключевые параметры «завязаны» на цене (price) и площади недвижимости (area).

Далее были сформированы две модели машинного обучения: XGBoost и LGBM, являющиеся моделями градиентного спуска, который минимизирует функцию ошибки за счёт передвижения по градиенту. На текущий момент времени (середина 2023 года) эти модели являются лидерами в области машинного об-

учения. Линейная, полиномиальная регрессии, деревья решений и другие модели не обладают такой точностью, как модели, использующие градиентный спуск в своём обучении [5].

Для построения моделей были использованы библиотеки «from xgboost import XGBRegressor» и «from lightgbm import LGBMRegressor».

Для того, чтобы оценить их качество, воспользуемся метрикой RMSE (root mean squared error — квадратный корень из среднеквадратичной ошибки). Результаты расчета рассматриваемой метрики, приведены на рисунке 6.

XGBoost average RMSE: 1173733.8482429956	LGBM average RMSE: 1213719.1148944688
Model 0 RMSE: 1239332.1827156602	Model 0 RMSE: 1206399.56522601
Model 1 RMSE: 1123007.04466807	Model 1 RMSE: 1217505.5163773333
Model 2 RMSE: 1197674.0786738254	Model 2 RMSE: 1195524.1380294133
Model 3 RMSE: 1145570.089026903	Model 3 RMSE: 1238461.3643609895
Model 4 RMSE: 1163085.8461305194	Model 4 RMSE: 1210704.990478598

Рис. 6. Результаты тестирования моделей по метрике RMSE
(Источник: построено на основе данных портала Kaggle.com)

В модели используется кросс-валидация, которая позволяет перемешивать выборки, каждый раз меняя обучающую и валидационную так, чтобы каждый запуск валидационные данные не пересекались. Как мы видим, из приведенных на рисунке 6 результатов расчета, в среднем модель XGBoost лучше справилась с обучением, чем LGBM, поэтому остановим выбор именно на ней.

Стоит заметить, что несмотря на хорошие результаты подгонки, размер ошибки составил 1,1 млн рублей, что довольно много. Т. е. 1,1 млн от средней стоимости недвижимости в 10 млн — это 10 % ошибки от стоимости квартиры.

Также была оценена важность критериев, которыми пользовалась модель, для этого использовалась библиотека XGBoost и команда feature_importances_ (рисунок 7).

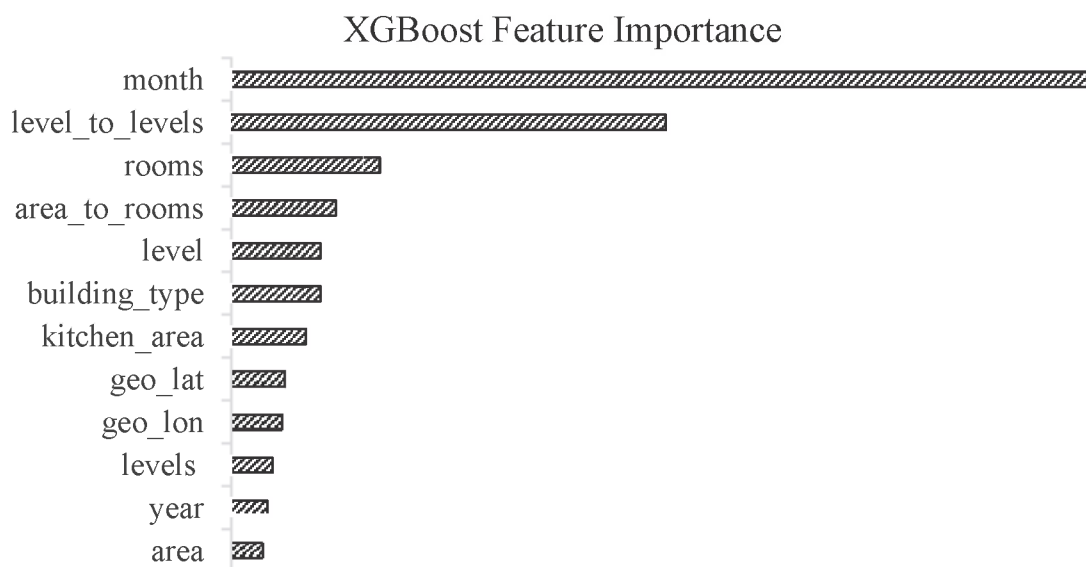


Рис. 7. Важность параметров для модели XGBoost
(Источник: построено на основе данных портала Kaggle.com)

Из диаграммы (рисунок 7) видим, что площадь (area), год продажи (year) и этажность (levels) дома сильно воздействуют на стоимость.

Итак, видно, что модель выдает не идеальные результаты. Это объясняется тем, что не хватает важных параметров при оценке недви-

жимости: мебель; ремонт; расстояние до метро; наличие инфраструктуры в шаговой доступности (школы, точки розничной продажи, торговые центры, медицинские учреждения). Также у модели отсутствует корректировка на то, в какой части Москвы продаётся квартира, что сильно меняет её стоимость и вес критерия.

В дальнейшем исследовании цен на рынке жилой недвижимости Москвы, стоит сосредоточиться именно на этих параметрах, что по нашему мнению увеличит прогнозную точность модели.

Выводы

Проведенный анализ московского рынка жилой недвижимости, позволяет нам сформулировать ряд выводов:

1) не смысла откладывать покупку недвижимости в Москве, так как стоимость жилья непрерывно растет, в некоторых периодах в большей степени чем доходы населения;

2) анализ месячной динамики показывает, что покупать недвижимость на московском рынке предпочтительнее, в первой половине года, а именно в июне, во второй половине года стоимость имеет повышающий тренд.

3) большинство людей предпочитает покупать квартиру до 10 этажа, поэтому инвестировать стоит именно в квартиры, находящиеся ниже этого этажа;

4) модели градиентного спуска позволили установить, что значительное влияние на стоимость квартир в Москве оказывают такие количественные факторы как площадь объекта недвижимости и этажность дома.

Информация о конфликте интересов

Мы, авторы данной статьи, со всей ответственностью заявляем о частичном и полном отсутствии фактического или потенциального конфликта интересов с какой бы то ни было третьей стороной, который может возникнуть вследствие публикации данной статьи.

ЛИТЕРАТУРА

1. *Грахов В. П.* Моделирование структурных сдвигов на рынке жилищного строительства РФ / В. П. Грахов, Л. Д. Гумерова, Д. А. Поздеев // Экономика и предпринимательство. — 2022. — № 3 (140). — С. 1135—1140. — DOI 10.34925/EIP.2022.140.03.216.

2. Датасет для анализа [Электронный ресурс]. — URL: <https://www.kaggle.com/datasets/mrdaniilak/russia-real-estate-20182021> (дата обращения: 10.06.2023).

3. Код по созданию модели и аналитике [Электронный ресурс]. — URL: <https://www.kaggle.com/code/phanenkov/real-estate> (дата обращения: 10.06.2023).

4. *Мамедли М. О.* Оценка стоимости недвижимости на основе больших данных / М. О. Мамедли, А. В. Умнов // Вопросы экономики. — 2022. — № 12. — С. 118—136. — DOI 10.32609/0042-8736-2022-12-118-136.

5. *Осадчий Н. К.* Оценка перспективного состояния рынка недвижимости города Москвы на основе количественных методов исследования / Н. К. Осадчий, И. С. Сызonenko // Научные записки молодых исследователей. — 2020. — Т. 8, № 1. — С. 5—20.

6. Проблемы моделирования оценки стоимости жилой недвижимости / Т. К. Богданова, А. Р. Камалова, Т. К. Кравченко, А. И. Полторак // Бизнес-информатика. — 2020. — Т. 14, № 3. — С. 7—23. — DOI 10.17323/2587-814X.2020.3.7.23.

7. *Пяткина Д. А.* Методология эконометрического моделирования на рынке недвижимости Москвы / Д. А. Пяткина // Экономика и предпринимательство. — 2021. — № 6 (131). — С. 410—416. — DOI 10.34925/EIP.2021.131.6.080.

8. *Цыпин А. П.* Модели оценки стоимости жилья с учетом пространственной вариации данных (на примере городов ПФО) / А. П. Цыпин, О. И. Стебунова, А. К. Салиева // Экономика и предпринимательство. — 2015. — № 11-2 (64). — С. 369—373.

9. *Цыпин А. П.* Статистический анализ закономерностей рынка арендного жилья г. Казань / А. П. Цыпин, О. В. Леднева // Вестник еurasийской науки. — 2020. — Т. 12, № 6. — С. 21.

LITERATURE

1. *Grakhov V. P.* Modeling structural shifts in the housing construction market of the Russian Federation / V. P. Grakhov, L. D. Gumerova, D. A. Pozdeev // Economics and entrepreneurship. — 2022. — № 3 (140). — Pp. 1135—1140. — DOI 10.34925/EIP.2022.140.03.216.

2. Analysis Date Set [Electronic Resource]. — URL: <https://www.kaggle.com/datasets/mrdaniilak/russia-real-estate-20182021> (Accessed Date: 10.06.2023).

3. Model Creation and Analytics Code [Electronic Resource]. — URL: <https://www.kaggle.com/code/phanenkov/real-estate> (Accessed Date: 10.06.2023).

4. *Mammadli M. O.* Estimating the value of real estate based on big data / M. O. Mammadli, A. V. Umnov // Economic issues. — 2022. — № 12. — Pp. 118—136. — DOI 10.32609/0042-8736-2022-12-118-136.

5. *Osadchiy N. K.* Assessing the promising state of the real estate market in the city of Moscow on the basis of quantitative research methods / N. K. Osadchiy, I. S. Syzonenko // Scientific notes of young researchers. — 2020. — Т. 8, № 1. — Pp. 5—20.

6. Problems of modeling the valuation of residential real estate / Т. К. Bogdanova, A. R. Kamalova, Т. К. Kravchenko, A. I. Poltorak // Business informatics. — 2020. — Т. 14. № 3. — Pp. 7—23. — DOI 10.17323/2587-814X.2020.3.7.23.

7. Pyatkina D. A. Methodology of econometric modeling in the real estate market in Moscow / D. A. Pyatkina // Economics and entrepreneurship. — 2021. — № 6 (131). — Pp. 410—416. — DOI 10.34925/EIP.2021.131.6.080.

8. Tsypin A. P. Models for assessing the cost of housing, taking into account the spatial variation of data (on the example of the cities of the Volga Federal District / A. P. Tsypin, O. I. Stebunova, A. K. Salieva // Economics and entrepreneurship. — 2015. — № 11-2 (64). — Pp. 369—373.

9. Tsypin A. P. Statistical analysis of the patterns of the rental housing market in Kazan / A. P. Tsypin, O. V. Ledneva // Bulletin of Eurasian Science. — 2020. — Т. 12, № 6. — Pp. 21.

УДК: 338.001.36

ПРОГНОЗИРОВАНИЕ И ПЛАНИРОВАНИЕ СПРОСА: КЕЙС ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ УПРАВЛЕНИИ ЦЕПОЧКАМИ ПОСТАВОК

Коды Jel: R41

*Рогулин Р. С., кандидат экономических наук, Владивостокский государственный университет, кафедра математики и моделирования, г. Владивосток, Россия
E-mail: rafassiaofusa@mail.ru; SPIN-код: 4125-0948*

Поступила в редакцию 16.06.2023. Принята к публикации 26.06.2023

Аннотация

Актуальность темы. *Необходимость рассмотрения степени влияния методов анализа данных и машинного обучения на прогнозирование спроса и планирование в управлении цепочками поставок.*

Цель. *Оценка степени влияния методов анализа данных и машинного обучения на прогнозирование спроса и планирование в управлении цепочками поставок.*

Методология. *Обзор тематических исследований и исследовательских работ для изучения потенциальных преимуществ этих методов в повышении эффективности цепочки поставок, управлении запасами, сокращении дефицита и повышении удовлетворенности клиентов.*

Результаты и выводы. *Выявлено несколько проблем и ограничений, включая проблемы с качеством данных и потребность в квалифицированном персонале. Даны предложения по преодолению этих проблем, включая повышение качества данных и инвестирование в обучение и развитие персонала. В исследовании также освещаются будущие направления исследований в области интеграции данных в реальном времени и прогнозной аналитики. Результаты этого исследования имеют важные последствия для менеджеров цепей поставок и исследователей с точки зрения преимуществ, проблем и будущих направлений исследований, использования методов анализа данных и машинного обучения в прогнозировании и планировании спроса.*

Область применения. *Сфера управления цепями поставок на мезоуровне.*

Ключевые слова: *управление цепочками поставок, анализ данных, машинное обучение, прогнозирование спроса, управление запасами, удовлетворенность клиентов, дефицит, данные в реальном времени, прогнозная аналитика.*

UDC: 338.001.36

FORECASTING AND FORECASTING DEMAND: THE CASE OF ARTIFICIAL INTELLIGENCE IN SUPPLY CHAIN MANAGEMENT

JEL Codes: R41

*Rogulin R. S., Candidate of Economic Sciences, Vladivostok State University, Department of Mathematics and Modeling, Almaty, st. Vladivostok, Russia
E-mail: rafassiaofusa@mail.ru; SPIN code: 4125-0948*