



## ОГЛАВЛЕНИЕ

<b>ЦЕЛЬ ИССЛЕДОВАНИЯ.....</b>	<b>2</b>
<b>ОБЗОР МЕТОДИК РАСЧЕТА И АНАЛИЗА ЗАВИСИМОСТЕЙ.....</b>	<b>2</b>
<b>ФОРМИРОВАНИЕ ВЫБОРКИ ДАННЫХ.....</b>	<b>4</b>
<b>АНАЛИЗ ЗАВИСИМОСТЕЙ.....</b>	<b>5</b>
<b>МОДЕЛИРОВАНИЕ И ИНТЕРПРИТАЦИЯ РЕЗУЛЬТАТОВ.....</b>	<b>7</b>
<b>ЗАКЛЮЧЕНИЕ.....</b>	<b>9</b>
<b>ЛИТЕРАТУРА.....</b>	<b>9</b>



## Цель исследования

Целью данного исследования является анализ и выявление факторов, влияющих на срок экспозиции объектов недвижимости, а также оценка возможности прогнозирования сроков экспозиции на основе продвинутого алгоритма машинного обучения на примере офисной недвижимости в г. Москве.

## Список допущений

Поскольку в работе используются данные, полученные из объявлений о продаже объектов недвижимости на электронном ресурсе, исследование проводится при следующих допущениях:

- дата появления объявления является датой выставления объекта на продажу (дата начала продаж);
- дата снятия объявления является датой продажи объекта;
- дата завершения наблюдения является датой, на которую фиксируется статус объекта (продан или не продан). Это не означает, что прекращено наблюдение за объектом (при анализе и обработке данных информация о последующих событиях не используется).

## Обзор методик расчета и анализа зависимостей

Для более глубокого понимания предметной области исследования рассмотрим некоторые публикации, посвященные анализу зависимости срока экспозиции от цены предложения.

В статье [1] шведских специалистов исследуется взаимосвязь цены предложения, фактической цены продажи и срок экспозиции (*ТОМ*) в контексте рынка недвижимости. Срок экспозиции рассматривается как важный фактор, влияющий на разницу между фактической ценой продажи и предложенной ценой. В исследовании использовались данные о продаже жилой недвижимости в округе Стокгольма за период с 1991 по 2001 г., а в качестве методов исследования — регрессионный и сравнительный анализ. Авторы построили несколько регрессионных моделей, которые учитывали такие факторы, как площадь объекта, выраженная суммой жилой площади, размера участка и других внутренних площадей (в квадратных метрах), возраст недвижимости, качество интерьера, расстояние до центра города и месяц в качестве непрерывной временной переменной. Кроме того, для оценки нелинейной взаимосвязи между *ТОМ* и ценовыми показателями они добавили в модели нелинейные эффекты, такие как квадрат *ТОМ*.

Результаты показывают, что более длительный срок экспозиции часто приводит к снижению цены сделки. В частности, увеличение времени нахождения объекта на рынке ассоциируется с понижением цены, особенно в условиях, когда предложение превышает спрос. Однако на растущем рынке высокие цены предложения могут поддерживать более длительное нахождение на рынке, что в итоге приводит к более высокой цене сделки. Авторы также обнаружили, что связь между *ТОМ* и ценой сделки нелинейна: цена может расти до определенного момента времени на рынке, после чего начинает снижаться.

В другой работе [2] этого направления обсуждается взаимосвязь между ценой недвижимости и временем нахождения на рынке (*ТОМ*) в условиях различных рыночных ситуаций с акцентом на продажи объектов, находящихся в собственности банков или



государственных учреждений (*REO*). Авторы расширяют классические модели поиска и ценовых корректировок, чтобы учесть изменения рыночных условий, и проводят эмпирический анализ на основе большого массива данных *REO*-продаж. Результаты показывают, что взаимосвязь между ценой и *TOM* значительно варьируется в зависимости от состояния рынка. В условиях стабильного или растущего рынка более длительное время на рынке приводит к увеличению цены продажи, поскольку длительный поиск помогает найти покупателей, готовых предложить более высокую цену. Однако в условиях падающего рынка длительное нахождение на рынке может привести к снижению цены. Общий вывод работы заключается в том, что, несмотря на возможные колебания в крайне неблагоприятных рыночных условиях, положительная взаимосвязь между ценой и *TOM* преобладает. Это связано с выгодами, получаемыми от продолжительного периода поиска покупателей.

В работе [3] автор так же, как и в предыдущей статье, исследует ликвидность на рынке жилой недвижимости в Стокгольме при различных состояниях рынка, используя время на рынке (*TOM*) в качестве индикатора ликвидности. Исследование основано на количественном объяснительном анализе и регрессионном анализе. Данные были разделены на кварталы и рыночные фазы, что позволило оценить влияние рыночных условий на *TOM* и ценовые отношения. Результаты показали, что существует отрицательная взаимосвязь между *TOM* и ценами на жилье во всех рыночных условиях. Однако величина этого эффекта варьировалась в зависимости от фазы рынка. В условиях роста рынка продавцы стремятся воспользоваться положительными условиями, в то время как в условиях спада они предпочитают ждать более благоприятной ситуации. Результаты подчеркивают рациональный подход продавцов к ценообразованию и времени на рынке, что открывает возможности для дальнейших исследований в этой области.

Отметим несколько работ, в которых авторы предлагают использовать алгоритмы машинного обучения в качестве эффективных методов прогнозирования времени нахождения объекта недвижимости на рынке. Среди работ этого направления особо выделим диссертацию Али Рахбариманеша [4], в которой исследована эффективность применения современных алгоритмов машинного обучения, таких как XGBoost, CatBoost, LightGBM и Random Forest, для предсказания времени, в течение которого объекты недвижимости остаются на рынке в Нидерландах. В работе использован обширный набор данных из базы Funda-Sold, охватывающий продажи домов с 2020 по 2023 г. Основное внимание уделено методу рекурсивного исключения признаков с кросс-валидацией (RFECV) для выбора наиболее значимых факторов, влияющих на *TOM*. Результаты показывают, что алгоритмы машинного обучения, особенно CatBoost, значительно превосходят традиционный метод наименьших квадратов (OLS) по точности предсказаний. Это исследование подчеркивает потенциал машинного обучения в улучшении анализа рынка недвижимости и принятия решений для профессионалов в этой области.

В статье [5] предложен метод многозадачного прогнозирования для оценки ликвидности рынка недвижимости через показатель «дни на рынке» (*DOM*). Авторами работы были собраны и подготовлены данные о продаже недвижимости в Пекине, в которые входили характеристики домов, жилых комплексов, а также географические характеристики и временные факторы. Для прогнозирования *DOM* авторы применяли метод многозадачного обучения (multi-task learning), где каждая задача представляет собой построение прогноза *DOM* для конкретного района города. Кроме того, авторы разбили



выборку на три временных интервала и строили прогнозы для каждого интервала по отдельности. Эксперименты показали, что предложенный авторами подход демонстрирует достаточно высокую эффективность и хорошие результаты.

Приведенные работы позволяют увидеть основные тенденции в отношении развития методов анализа и прогнозирования сроков экспозиции. В целом работы в этой области фокусируются на выявлении взаимосвязей между временем нахождения объекта на рынке и его ценой, а также на применении различных методов для прогнозирования этих показателей. В некоторых работах прогнозирование осуществляется на основе традиционных статистических методов, тогда как других применяются современные алгоритмы машинного обучения для повышения точности прогноза и принятия решений.

## Формирование выборки данных

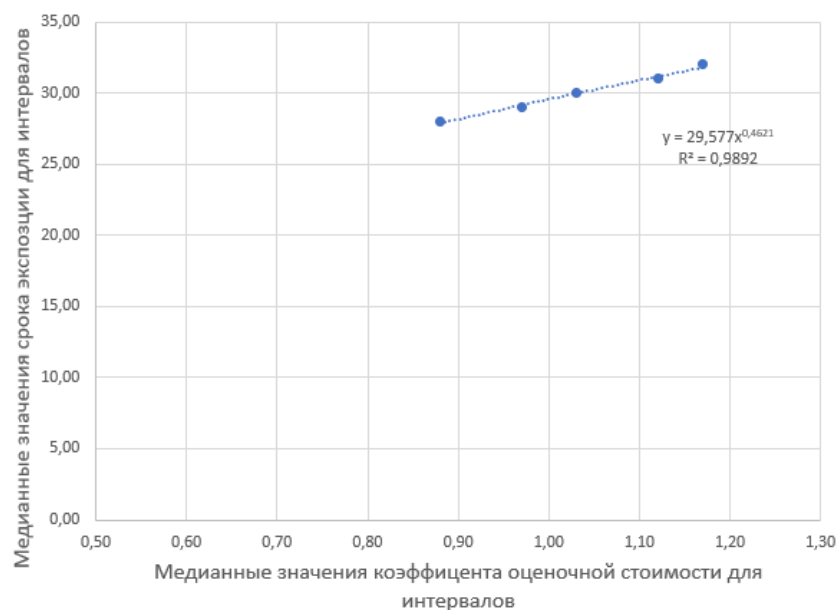
Все данные, использованные в исследовании, были получены с информационного портала в сфере недвижимости Move.ru. В выборку были включены исключительно офисные объекты в г. Москве, которые находились в продаже. Все объекты к моменту проведения исследования уже были сняты с продажи, что позволяет рассматривать их срок экспозиции как окончательный. Временной диапазон дат публикации, в котором находятся отобранные объявления, — с 1 января 2022 г. по 1 января 2024 г. Данные были предварительно предобработаны: была проведена очистка от выбросов по признакам «удельная цена», «количество просмотров» и «общая площадь». Все записи с пропущенными значениями были удалены, поскольку заполнение пропусков искусственным образом могло бы исказить исходные данные. Также данные были очищены от повторяющихся объявлений о продаже одних и тех же объектов.

В нашем исследовании каждый объект описывается следующими характеристиками:

- *общая площадь* — площадь объекта недвижимости в м<sup>2</sup>.
- *цена* — цена продажи объекта недвижимости;
- *коэффициент оценочной стоимости* — рассчитан как отношение удельной цены объекта, заявленного продавцом в объявлении на продажу, к расчетной удельной цене, которая отражает рыночную стоимость выставленного на продажу объекта. Смысл данного признака — получить числовую оценку недооцененности или переоцененности объекта недвижимости. Значения коэффициента ниже единицы означают, что объект был недооценен и выставлен на продажу по цене ниже реальной, тогда как значение выше единицы означает обратное;
- *среднее количество просмотров* в единицу времени (далее индикатор *спрос*) — количество просмотров, которое было зафиксировано за весь срок нахождения объекта на доске объявлений, деленный на этот срок в днях;
- *основная ценовая зона* — ценовая зона, в которой находится объект недвижимости, в соответствии с принятым зонированием территорий;
- *день недели* — день недели, в который было опубликовано объявление о продаже недвижимости, получен из даты создания;
- *месяц создания* — месяц, в который было опубликовано объявление о продаже недвижимости, получен из даты создания;
- *квартал* — квартал, в который было опубликовано объявление о продаже недвижимости, получен из даты создания;
- *дата создания* — дата, в которую объявление было выставлено на продажу;
- *срок экспозиции* — целевая переменная, количество дней от момента выставления объекта на продажу, до момента снятия объекта с продажи.

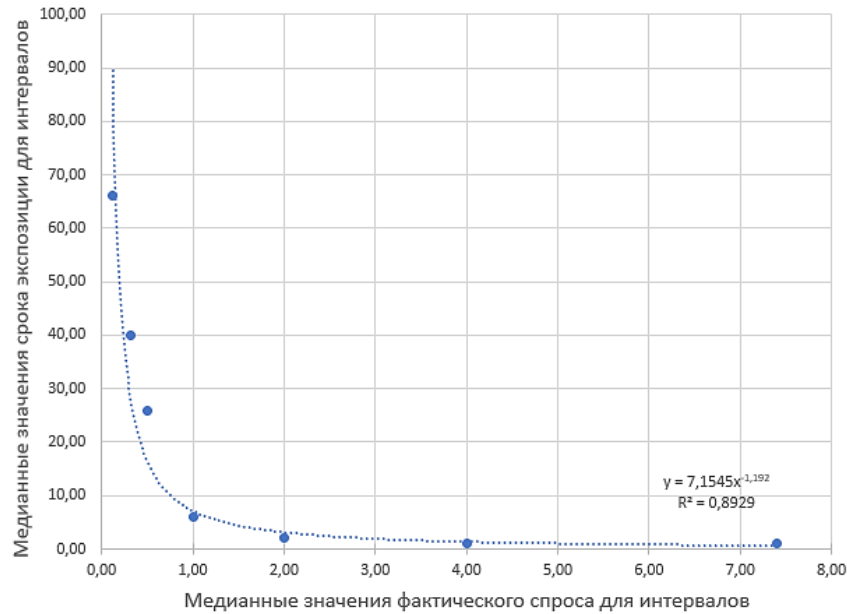
## Анализ зависимостей

В рамках анализа на примере данных по Москве были исследованы зависимости между ценообразующими факторами (признаками) объекта и его сроком экспозиции. Наиболее значимые признаки были включены в качестве предикторов для прогнозирования ликвидности. С целью наглядного представления на графиках непрерывные признаки были разбиты на интервалы. В каждом из интервалов было рассчитано медианное значение ценообразующего фактора и срока экспозиции и с помощью этих значений были построены диаграммы рассеяния для визуального отображения зависимостей, где на каждой из осей находятся медианные значения предиктора и срока экспозиции по данным из полученных интервалов. Такой подход способствует лучшему представлению зависимости между переменными, снижая влияние выбросов и нерегулярных колебаний в данных. Полученные зависимости между предикторами и сроком экспозиции представлены в графическом виде (рис. 1–4).



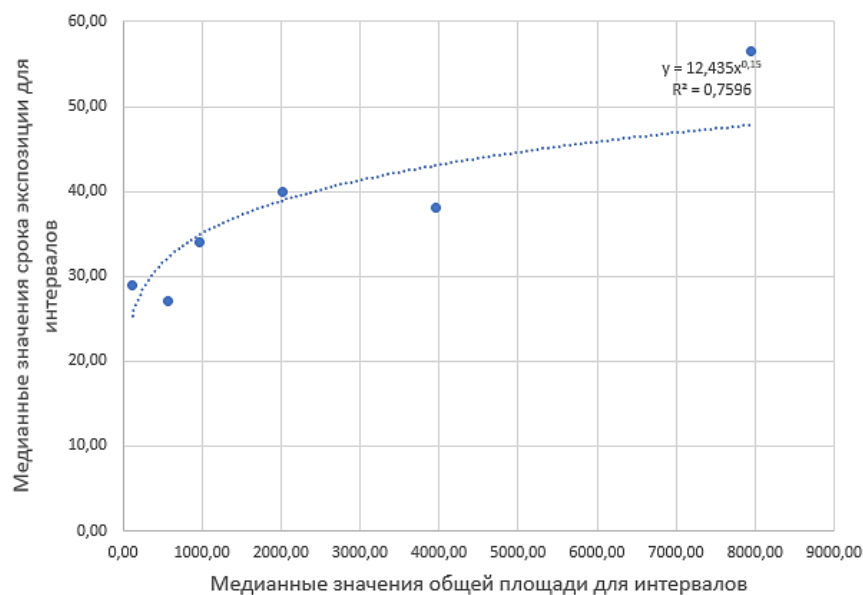
**Рис. 1.** Зависимость срока экспозиции от коэффициента оценочной стоимости, тренировочная выборка

Судя по полученной зависимости (см. рис. 1), переоцененные объекты часто продаются дольше, потому что цена, установленная продавцом, не соответствует реальной. Логично предположить, что покупатели менее склонны приобретать объекты, стоимость которых им кажется завышенной. Кроме того, в случае переоцененных объектов продавцы могут быть менее склонны снижать цену в процессе переговоров, ожидая, что рано или поздно найдется покупатель, который заплатит полную стоимость, что также может увеличивать срок экспозиции. Однако недооцененные объекты имеют более низкий срок экспозиции, поскольку покупка объекта, выставленного на продажу по цене ниже рыночной, является выгодной, из-за чего они привлекают больше внимания и соответственно продаются быстрее. Известно, что некоторые продавцы могут намеренно устанавливать такую цену для ускорения продажи объекта, если они хотят избежать лишних затрат, например, связанных с обслуживанием объекта продажи или если продажа объекта должна пройти в сжатые сроки.



**Рис. 2.** Зависимость срока экспозиции от фактического спроса, тренировочная выборка

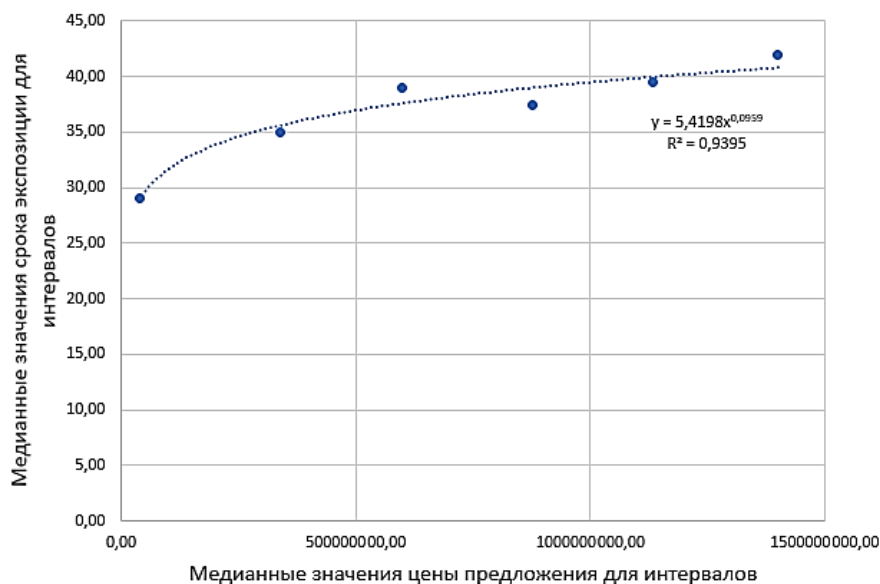
Зависимость срока экспозиции от фактического спроса (см. рис. 2) демонстрирует, что между этими показателями есть достаточно сильная взаимосвязь. Количество просмотров в единицу времени является показателем интереса к объекту со стороны потенциальных покупателей, поэтому логично, что объекты, привлекающие больший интерес, быстрее продаются. Высокий спрос может отражать адекватную цену продажи, хорошее местоположение объекта, его качество и другие характеристики, которые привлекают покупателей. Если объект не вызывает интереса, это может быть связано с завышенной ценой, недостаточной привлекательностью района или иными недостатками. В таком случае объект остается на рынке дольше, поскольку потенциальные покупатели могут отвергать его на этапе просмотра объявления.



**Рис. 3.** Зависимость срока экспозиции от общей площади

Если рассматривать зависимость срока экспозиции от общей площади (см. рис. 3), то можно увидеть, что объекты, имеющие большую площадь, продаются дольше. Можно

предположить, что крупные объекты недвижимости требуют большего времени для поиска подходящего покупателя, поскольку такие объекты часто ориентированы на более узкую аудиторию, особенно если речь идет о дорогих и элитных домах с большой площадью. Покупателей, желающих и имеющих возможность купить объектов, имеющих малую площадь, на рынке гораздо больше, поэтому медианный срок экспозиции небольших объектов обычно меньше, что и нашло отражение в повышательном характере зависимости сроков экспозиции от площади объектов, выставленных на продажу.



**Рис. 4.** Зависимость срока экспозиции от цены предложения

Схожая зависимость наблюдается и с ценой предложения — по мере увеличения цены объекта увеличивается и срок экспозиции (см. рис. 4). Логичным объяснением такой зависимости будет предположение о малом количестве покупателей, которые могут позволить себе более дорогие объекты недвижимости. Кроме того, при приобретении более дорогих объектов переговоры о покупке более дорогих объектов обычно ведутся дольше, поскольку последствия ошибочных решений более значимы.

## Моделирование и интерпретация результатов

Для предсказания срока экспозиции объекта недвижимости по его характеристикам использовали алгоритм CatBoost. В качестве исходных данных были выбраны объявления о продаже объектов недвижимости, взятые с известных электронных досок объявлений. Общее число объявлений с данными о продаже коммерческой недвижимости после очистки от повторяющихся объявлений составляло 19200 шт. Все объявления были разделены на три выборки: тренировочную (обучающую), валидационную и тестовую в соотношении 80 : 10 : 10.

В ходе обучения модели было принято решение целевой признак «срок экспозиции», представленный непрерывными значениями, перевести в разряд категориальных, что позволило свести поставленную задачу к задаче многоклассовой классификации. С этой целью шкала непрерывных значений срока экспозиции была разбита на следующие интервалы: от 0 до 60 дней, от 60 до 210 и больше 210. Интервалы выбирали эмпирическим путем, и такое разбиение показало наилучший результат. Для вычисления предиктора «коэффициент оценочной стоимости» отдельно решали задачу машинного обучения, в которой определяли удельную рыночную стоимость, необходимую для расчета данного коэффициента. Для этого также применяли модель CatBoost, при этом точность прогнозирования удельной стоимости, измеряемая с помощью метрики MdAPE для



тренировочной и тестовой выборок, была не хуже 20 %.

В процессе обучения и последующего тестирования получены результаты, представленные в виде матриц (рис. 5, 6).

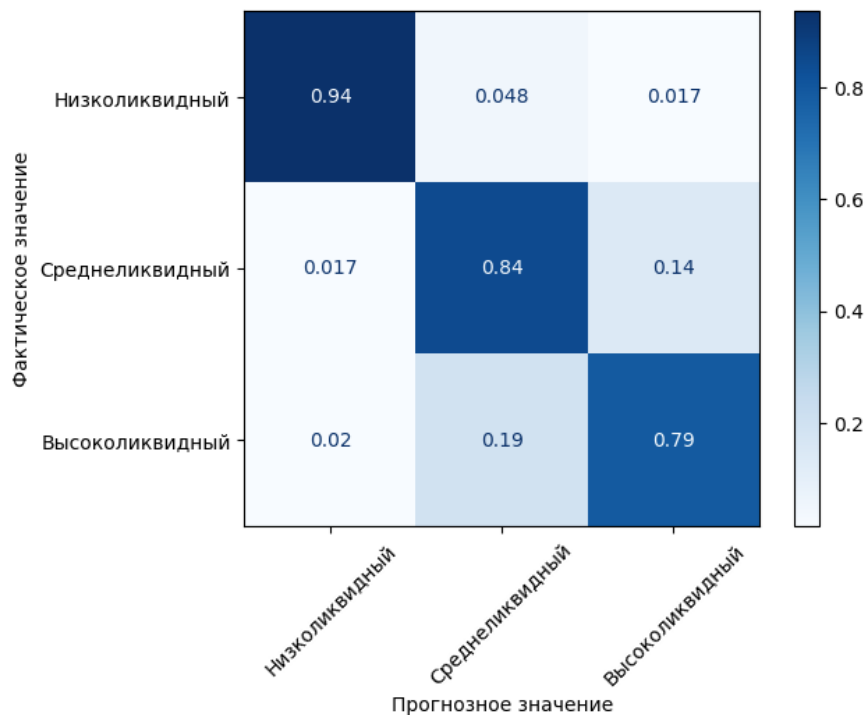


Рис. 5. Матрица ошибок, тренировочная выборка, данные по Москве

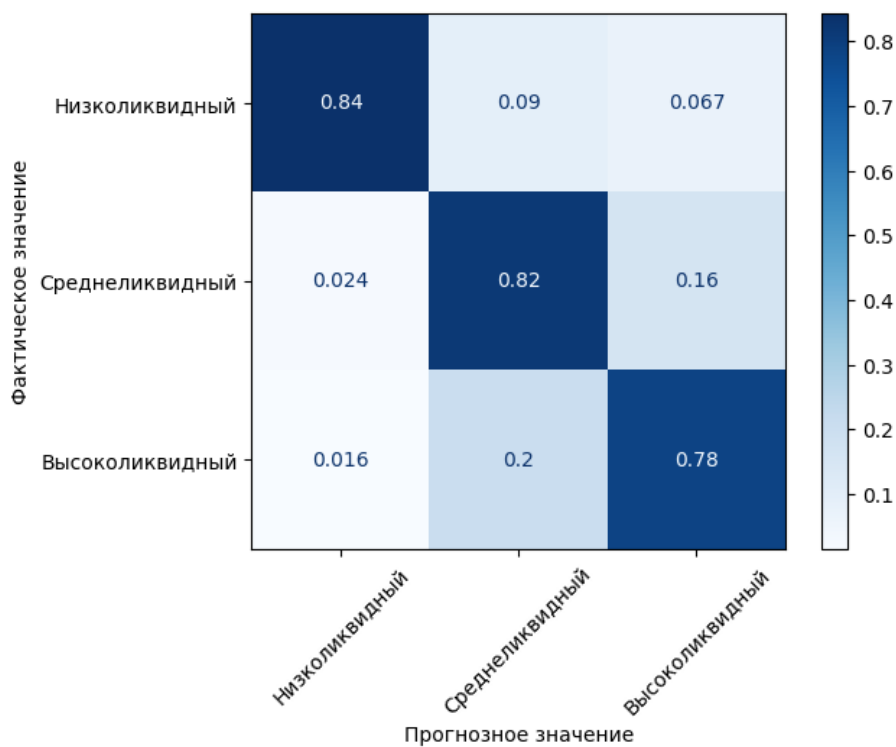


Рис. 6. Матрица ошибок, тестовая выборка, данные по Москве





Анализ результатов позволяет сделать следующие выводы.

1. Отсутствует переобучение, результаты на тренировочной и тестовой выборке схожи, а значит, модель сохраняет способность обобщать на новых данных.
  2. Модель достаточно хорошо различает объекты, относящиеся к классам, объединяющим низколиквидные и высоколиквидные объекты. Поэтому использование данной модели для прогнозирования сроков экспозиции следует признать достаточно эффективным.
  3. Однако выделение объектов, относящихся к промежуточному классу среднеликвидных объектов, не такое четкое. Ошибки в этом случае несколько выше, что естественно, поскольку связь действительного срока экспозиции с характеристиками объекта не такая сильная.
- Значение ROC-AUC составило 0,95 и 0,92 для тренировочной и тестовой выборки соответственно, что также подтверждает высокое качество модели и хорошую способность различать классы.

## Заключение

Подводя итоги, можно сделать вывод, что результаты оценки срока экспозиции на примере офисных объектов Москвы были удовлетворительными. Отдельно можно отметить следующую важную информацию.

1. Как было сказано ранее, модель демонстрирует высокую точность в определении всех классов, при этом достаточно хорошо распознает низколиквидные и высоколиквидные объекты, что очень важно при оценке рисков, связанных с недостаточной ликвидностью.
2. Судя по оценке эффективности модели с помощью матриц ошибок и метрики ROC AUC, можно сделать вывод, что использование алгоритма градиентного бустинга для прогнозирования срока экспозиции целесообразно и эффективно.
3. Важно отметить, что при оценке срока экспозиции коммерческой недвижимости стоит отдельно рассматривать данные Москвы и данные других городов, поскольку рынок в Москве заметно отличается от рынка в других городах. Кроме того, рынок может различаться и в зависимости от сегментов недвижимости, поэтому объекты из разных сегментов также стоит рассматривать отдельно.

## Литература

1. Björklund K., Dadzie J. A., Wilhelmsson M. Offer price, transaction price and time-on-market // Property Management. 2006. Vol. 24, no. 4. Pp. 415–426.
2. Zhiyong An Ping Cheng, Zhenguo Lin, Yingchun Liu. How do market conditions impact price-TOM relationship? Evidence from real estate owned (REO) sales // Journal of Housing Economics. 2013. Vol. 22. Pp. 250–263.
3. Luna Andonegui C. A. TOM: Measuring Liquidity in the Swedish Real Estate Market // KTH Royal Institute of technology. 2023.
4. Rahbarimanesh A. Comparative analysis of machine learning algorithms for time on market (TOM) // Tilburg University School of Humanities and Digital Sciences Department of Cognitive Science & Artificial Intelligence Tilburg, The Netherlands. 2024.
5. Hengshu Zhu, Hui Xiong, Fangshuang Tang, Qi Liu, Yong Ge, Enhong Chen, Yanjie Fu. Days on Market: Measuring Liquidity in Real Estate Markets // Proceedings of the 22nd ACM SIGKDD International Conference on KnowledgeDiscovery and Data Mining. 2016. Pp. 393–402.