

УДК 519.862.6

UDC 519.862.6

5.2.2. Математические, статистические и инструментальные методы в экономике

5.2.2. Mathematical, statistical and instrumental methods in economics

МОДЕЛИРОВАНИЕ ЦЕНООБРАЗОВАНИЯ НА ВТОРИЧНОМ РЫНКЕ МОСКВЫ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ

MODELING PRICING IN THE MOSCOW SECONDARY MARKET UNDER UNCERTAINTY

Ханенков Павел Вячеславович
Аспирант 2 курс кафедры прикладной математики

Hanenkov Pavel Vyacheslavovich
2nd year postgraduate student of the Department of Applied Mathematics

SPIN-код: 4445-9320

RSCI SPIN-code: 4445-9320

phanenkov@mail.ru

phanenkov@mail.ru

Университет «Синергия», Москва, Россия

Synergy University, Moscow, Russia

В статье проводится сравнительный анализ эффективности различных методов машинного обучения для прогнозирования цен на однокомнатные квартиры на вторичном рынке Москвы в период с сентября 2022 по май 2025 года. Исследование осуществляется в условиях значительной макроэкономической неопределенности, характеризующейся высокой волатильностью ключевых показателей. Для моделирования используются как исторические данные о ценах, так и набор макроэкономических и гео-переменных. В работе тестируются модели линейной регрессии, ансамблевые методы (стэкинг), рекуррентные нейронные сети (LSTM) и модель прогнозирования временных рядов Prophet. Результаты показывают, что классические методы машинного обучения, в частности стэкинг регрессоров, демонстрируют наилучшую точность на данном временном интервале, превосходя более сложные нейросетевые архитектуры, что свидетельствует о специфике обрабатываемых данных и важности правильного подбора признаков

The article provides a comparative analysis of the effectiveness of various machine learning methods for predicting prices of one-room apartments on the secondary market in Moscow for the period from September 2022 to May 2025. The study is conducted under conditions of significant macroeconomic uncertainty, characterized by high volatility in key indicators. The modeling utilizes both historical price data and a set of macroeconomic and geographic variables. The study tests linear regression models, ensemble methods (stacking), recurrent neural networks (LSTM), and the Prophet time series forecasting model. The results show that classical machine learning methods, in particular regressor stacking, demonstrate the best accuracy on this time interval, outperforming more complex neural network architectures, which indicates the specificity of the processed data and the importance of proper feature selection

Ключевые слова: НЕДВИЖИМОСТЬ, ВТОРИЧНЫЙ РЫНОК МОСКВЫ, ЦЕНООБРАЗОВАНИЕ НЕДВИЖИМОСТИ, МАШИННОЕ ОБУЧЕНИЕ, LSTM, PROPHET, СТЭКИНГ

Keywords: REAL ESTATE, MOSCOW SECONDARY MARKET, REAL ESTATE PRICING, MACHINE LEARNING, LSTM, PROPHET, STACKING

<http://dx.doi.org/10.21515/1990-4665-214-049>

Введение

Российский рынок жилой недвижимости, будучи важнейшей составляющей экономики, напрямую влияет на социально-экономическое

<http://ej.kubagro.ru/2025/10/pdf/49.pdf>

развитие государства. Его состояние формируется под воздействием макроэкономической конъюнктуры, демографических изменений, мер государственного регулирования и инвестиционной привлекательности.

Актуальность исследования цен на жилую недвижимость в Москве, являющуюся крупнейшим рынком России, обусловлена следующими факторами:

1) Динамика цен отражает не только баланс спроса и предложения, но и чутко реагирует на изменения макроэкономической конъюнктуры [1]. Период с конца 2022 года характеризуется повышенной неопределенностью, вызванной действием санкционного давления, волатильностью на сырьевых рынках, колебаниями курса национальной валюты и изменением параметров денежно-кредитной политики. Эти факторы создают сложную среду для прогнозирования, где традиционные методы могут оказаться неэффективными.

2) Вторичный рынок однокомнатных квартир представляет особый интерес как наиболее ликвидный и массовый сегмент, что делает его удобным объектом для анализа. Традиционно для моделирования цен используются методы регрессионного анализа [2]. Однако в последние годы все более популярными становятся подходы, основанные на машинном обучении, включая рекуррентные нейронные сети (RNN), в частности сети с долгой краткосрочной памятью (LSTM), которые способны улавливать сложные временные зависимости [3], а также специализированные библиотеки для прогнозирования временных рядов, такие как Prophet [4].

Методы и материалы

Целью данного исследования является сравнительная оценка точности различных моделей машинного обучения для прогнозирования цен на однокомнатные квартиры в Москве в условиях макроэкономической

неопределенности. Для достижения цели были поставлены следующие задачи:

1. Сформировать набор данных, включающий транзакционные данные по сделкам купли-продажи и релевантные макроэкономические показатели.
2. Реализовать и обучить ряд моделей-кандидатов: от простой линейной регрессии до сложных ансамблевых и нейросетевых архитектур.
3. Провести сравнительный анализ точности моделей на основе метрики RMSE.
4. Выявить модель, демонстрирующую наилучшую прогностическую способность для рассматриваемой задачи, и проанализировать причины ее эффективности.

Объектом исследования в работе выбран рынок жилого недвижимого имущества Москвы.

Предметом исследования являются тенденции, характеристики и показатели рынка жилой недвижимости.

Для исследования будут использованы методы эмпирического исследования, методы анализа и синтеза материала, инструменты корреляционно-регрессионного анализа, инструменты нейросетевого моделирования и анализа временных рядов.

Научная новизна исследования заключается в проведении сравнительного анализа методов машинного обучения для прогнозирования цен на недвижимость в условиях макроэкономической нестабильности, характерной для России периода 2022-2025 гг. Впервые на реальных данных показано, что в условиях ограниченной выборки и высокой волатильности ансамблевые методы (стэкинг) и линейная регрессия превосходят по точности более сложные нейросетевые архитектуры (LSTM, Prophet), склонные к переобучению. Полученные

результаты ставят под вопрос универсальный подход к применению глубокого обучения для задач прогнозирования на нестабильных рынках.

Результаты исследования

Исходный набор данных содержал информацию о сделках по однокомнатным квартирам на вторичном рынке Москвы за период с 2022-09 по 2025-05. По каждой сделке фиксировалась цена объекта, а также его параметры: этаж, общая площадь, площадь кухни, количество этажей в доме и расстояние до центра города.

Для учета макроэкономического контекста в модель были включены следующие внешние признаки:

- Цена на нефть
- Уровень инфляции
- Курс доллара США к российскому рублю.
- Валовой внутренний продукт России (квартальные данные интерполировались для недельной частоты).
- Курс акций компании ПИК, как индикатор состояния сектора строительства и недвижимости.

Для устранения шума и приведения данных к единой временной частоте был применен метод *downsampling* с агрегацией по неделям. Это позволило сгладить временной ряд и получить более стабильную динамику для анализа.

В исследовании были использованы следующие модели:

Линейная регрессия (Linear Regression): Использовалась в качестве базовой модели. Модель предполагает линейную зависимость между целевой переменной (ценой) и признаками.

Модель Prophet: Специализированная аддитивная модель для прогнозирования временных рядов от компании Facebook, предназначенная для работы с данными, имеющими сильную сезонность и

эффект праздников [4]. Все доступные признаки были включены в модель в качестве регрессоров.

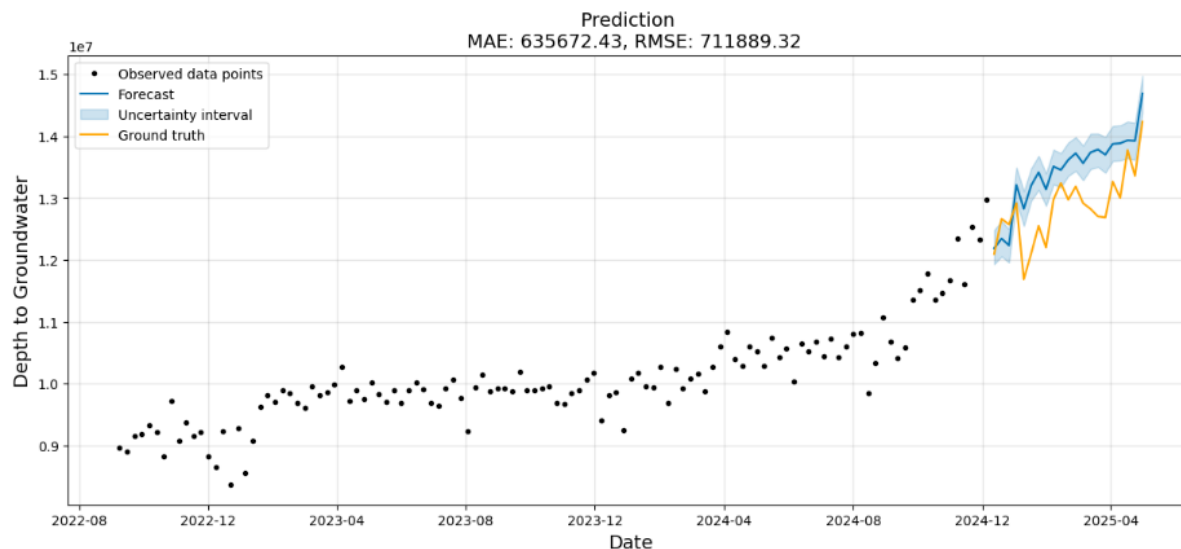


Рисунок 1 – Модель Prophet для цен на недвижимость (руб)

Многослойный перцептрон (MLP): Полносвязная нейронная сеть. Тестировались архитектуры разной сложности: простая (4 плотных слоя) и усложненная (с добавлением слоев Dropout и BatchNormalization для борьбы с переобучением).

Рекуррентная нейронная сеть (LSTM): Сеть с долгой краткосрочной памятью, предназначенная для работы с последовательностями. Было протестировано две конфигурации:

Простая LSTM: Использовала только историю цен за скользящее окно в 28 дней.

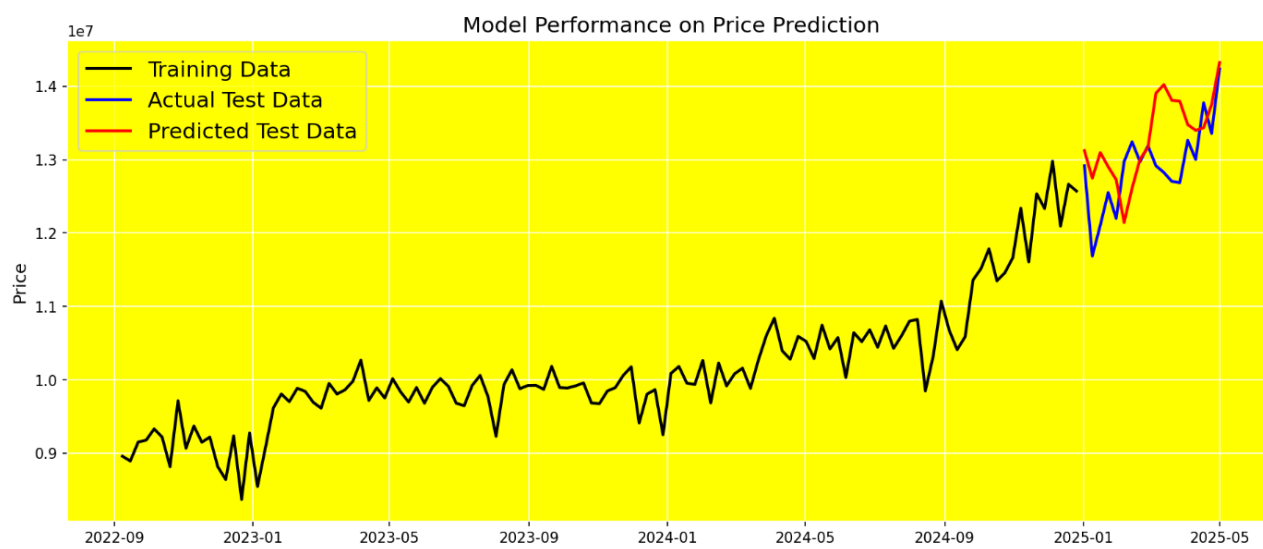


Рисунок 2 – Нейросетевая модель с LSTM для цен на недвижимость (руб)

Мультивариативная LSTM: Использовала то же скользящее окно, но включала в себя все макроэкономические признаки и дистанцию до центра города.

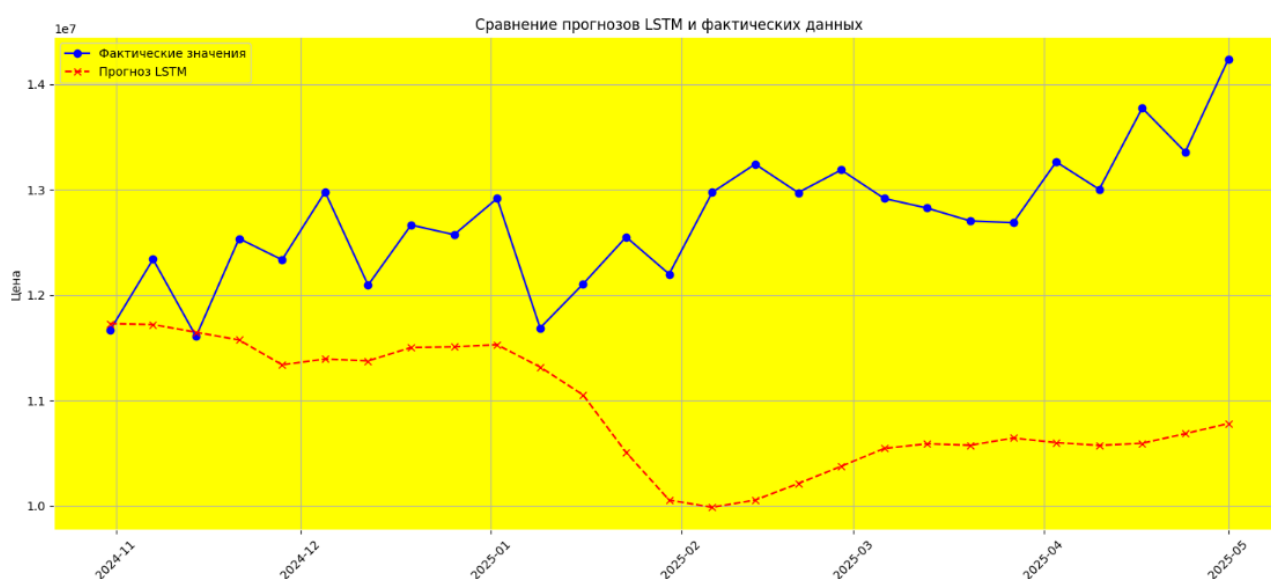


Рисунок 3 – Нейросетевая модель с LSTM со всеми макро-параметрами для цен на недвижимость (руб)

Стэкинг-регрессор (StackingCVRegressor): Ансамблевый метод, который комбинирует несколько базовых моделей (RandomForestRegressor, GradientBoostingRegressor, LinearRegression) с

помощью мета-обучателя (в данном случае - метод кросс-валидации) для получения более точного прогноза [5].

Оценка качества моделей проводилась на основе метрики RMSE.

Меньшее значение RMSE указывает на более высокую точность модели.

Результаты экспериментов по прогнозированию цен с использованием различных моделей представлены в Таблице 1.

Таблица 1. Сравнительные результаты моделей по метрике RMSE

Модель	Конфигурация	RMSE, руб
Стэкинг-регрессор	RandomForest, GradientBoosting, LinearRegression	323 000
Линейная регрессия	Все признаки	424 000
Prophet	Все признаки в качестве регрессоров	711 000
LSTM	Только цена, окно=28 дней	706 000
LSTM	Все макроэкономические признаки, окно=28 дней, усложненная архитектура	2 000 000
MLP	Все признаки, только 4 dense слоя	2 800 000
MLP	Все признаки, 4 dense слоя с	1 766 000

	Dropout и BatchNormalization	
--	------------------------------	--

Анализ результатов позволяет сделать несколько ключевых выводов:

1) Эффективность ансамблевых методов. Наилучший результат ($RMSE = 323\,000$ руб.) показал стэкинг-регрессор. Это согласуется с теорией машинного обучения, согласно которой ансамбли моделей часто позволяют компенсировать ошибки отдельных алгоритмов и давать более устойчивые прогнозы [5]. Комбинация "сильных" моделей, таких как градиентный бустинг, с простой линейной регрессией, позволила эффективно уловить как сложные нелинейные зависимости, так и общий тренд.

2) Устойчивость классических подходов. Линейная регрессия, несмотря на свою простоту, показала весьма конкурентоспособный результат ($RMSE = 424\,000$ руб.), существенно превзойдя более сложные нейросетевые архитектуры. Это может свидетельствовать о том, что в исследуемом периоде взаимосвязь между признаками и целевой переменной в значительной степени может быть аппроксимирована линейной моделью. Кроме того, линейные модели менее склонны к переобучению при ограниченном объеме данных.

3) Низкая эффективность нейросетевых моделей. Все тесты нейросетевой архитектуры (LSTM и MLP) показали наихудшие результаты. Высокий $RMSE$ мультивариативной LSTM и MLP (более 1.7 млн руб.) является явным признаком переобучения. Нейронные сети, особенно глубокие, требуют очень больших объемов данных для корректного обучения. Вероятно, имеющегося набора данных (около 2.5 лет недельных наблюдений) оказалось недостаточно для того, чтобы такие модели смогли выучить обобщающие закономерности, а не шум в обучающей выборке. Простая LSTM, использующая только цену, показала себя лучше, но все же

уступила линейным и ансамблевым методам, что указывает на недостаточность лишь исторических цен для точного прогноза в условиях неопределенности.

4) Ограничения моделей, учитывающих время. Модель Prophet, несмотря на свою популярность, показала результат, сопоставимый с простой LSTM. Это может быть связано с тем, что в данных отсутствуют ярко выраженные сезонные паттерны (например, годовые), которые Prophet призван улавливать, а влияние внешних регрессоров в рамках этой модели не позволило существенно повысить точность.

Таким образом, можно заключить, что для прогнозирования цен на недвижимость в условиях ограниченного объема данных и высокой макроэкономической неопределенности классические и ансамблевые методы машинного обучения оказываются предпочтительнее сложных нейросетевых архитектур.

Заключение

В ходе исследования была решена задача сравнительного анализа моделей машинного обучения для прогнозирования цен на однокомнатные квартиры на вторичном рынке Москвы. На основе данных за период с сентября 2022 по май 2025 года с привлечением макроэкономических показателей было протестировано семь различных подходов.

Наивысшую прогностическую точность продемонстрировал стэкинг-регрессор, объединяющий случайный лес, градиентный бустинг и линейную регрессию ($RMSE = 323\ 000$ руб.). Хорошие результаты также показала модель линейной регрессии ($RMSE = 424\ 000$ руб.), что подтверждает ее эффективность в качестве базового метода для подобных задач.

Ключевым выводом работы является демонстрация низкой эффективности глубоких нейронных сетей (LSTM и MLP) на данном конкретном наборе данных. Основной причиной их низкой

производительности является склонность к переобучению при недостаточном объеме данных.

Литература

1. Goodman, A. C., & Thibodeau, T. G. Housing market segmentation // Journal of Housing Economics. – 1998. – Vol. 7, no. 2. – P. 121-143.
2. Malpezzi, S. Hedonic pricing models: A selective and applied review // Housing economics and public policy. – Oxford: Blackwell Science, 2003. – P. 67-89.
3. Hochreiter, S., & Schmidhuber, J. Long short-term memory // Neural computation. – 1997. – Vol. 9, no. 8. – P. 1735-1780.
4. Taylor, S. J., & Letham, B. Forecasting at scale // The American Statistician. – 2018. – Vol. 72, no. 1. – P. 37-45.
5. Breiman, L. Stacked regressions // Machine learning. – 1996. – Vol. 24, no. 1. – P. 49-64.

References

1. Goodman, A. C., & Thibodeau, T. G. Housing market segmentation // Journal of Housing Economics. – 1998. – Vol. 7, no. 2. – P. 121-143.
2. Malpezzi, S. Hedonic pricing models: A selective and applied review // Housing economics and public policy. – Oxford : Blackwell Science, 2003. – P. 67-89.
3. Hochreiter, S., & Schmidhuber, J. Long short-term memory // Neural computation. – 1997. – Vol. 9, no. 8. – P. 1735-1780.
4. Taylor, S. J., & Letham, B. Forecasting at scale // The American Statistician. – 2018. – Vol. 72, no. 1. – P. 37-45.
5. Breiman, L. Stacked regressions // Machine learning. – 1996. – Vol. 24, no. 1. – P. 49-64.