

УДК 338
ББК 65.05

Оксана Сергеевна МУРАВЬЕВА

Кандидат экономических наук, доцент
Санкт-Петербургский Государственный Университет (Санкт-Петербург)
E-mail: o.muraveva@spbu.ru

Анна Алексеевна МАМАЕВА

Магистрант
Санкт-Петербургский Государственный Университет (Санкт-Петербург)
E-mail: st113327@student.spbu.ru

Oksana MURAVEVA

Ph.D. in Economics, associate professor
Saint Petersburg State University (Saint-Petersburg)
E-mail: o.muraveva@spbu.ru

Anna MAMAЕVA

Master Student
Saint Petersburg State University (Saint-Petersburg)
E-mail: st113327@student.spbu.ru

ПОТЕНЦИАЛ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В ЦЕНООБРАЗОВАНИИ НА РЫНКЕ ЖИЛОЙ НЕДВИЖИМОСТИ

Статья посвящена обзору особенностей спроса и предложения на рынке жилой недвижимости, формированию рыночных пузырей, а также выявлению и классификации факторов, влияющих на ценообразование в данной сфере. Охарактеризованы различные подходы к формированию цены, начиная от алгоритмов на основе условий, и заканчивая алгоритмами на основе методов машинного обучения в том числе нейронных сетей.

Ключевые слова: машинное обучение, нейронные сети, цифровые технологии, ценообразование, девелоперские компании.

THE POTENTIAL OF MACHINE LEARNING METHODS IN PRICING IN THE RESIDENTIAL REAL ESTATE MARKET

The article is devoted to an overview of the characteristics of supply and demand in the residential real estate market, the formation of market bubbles, as well as the identification and classification of factors affecting pricing in this area. Various approaches to price formation are characterized, ranging from algorithms based on conditions to algorithms based on machine learning methods, including neural networks.

Keywords: machine learning, neural networks, digital technologies, pricing, development companies.

Жилая недвижимость является одним из наиболее важных экономических секторов, где спрос на владение прямо коррелирует с социально-экономическим развитием. Рынок недвижимости обладает сложной структурой, включающей в себя множество элементов и подсистем. Недвижимость отличается от других товаров такими характеристиками как долговечность, неподвижность и уникальность, что делает рынок недвижимости особенно сложным для ценообразования и прогнозирования объема продаж. Также он достаточно дробно сегментирован, поскольку различные пользователи обладают разными потребностями, разным платежеспособным спросом.

Целью данной статьи является обобщение факторов, которые следует учитывать при ценообразовании на рынке жилой недвижимости, а также анализ подходов применения различных алгоритмов, в том числе искусственного интеллекта при формировании цены с учетом данных факторов.

Под спросом на рынке недвижимости понимается количество объектов и права на них, которые покупатели готовы приобрести по складывающимся ценам за определенный промежуток времени¹. В профессиональной среде разделяют потенциальный и эффективный спрос². Потенциальный спрос формируется потребностями населения без требуемого уровня дохода для совершения покупки. Эффективный спрос, в свою очередь, формируется реализованными потребностями, поддерживаемыми финансовыми возможностями. Спрос на жилье характеризуется высокой интенсивностью, слабой реакцией на изменение цен, значительными различиями между потенциальным и эффективным спросом, а также низкой эластичностью.

Предложение, в свою очередь, – это количество объектов недвижимости, которые собственники готовы продать по определенным ценам за определенный промежуток времени. Предложение медленно реагирует на краткосрочные изменения цен, поскольку оно подвержено влиянию процесса развития, деловых циклов, доступности ресурсов и потребительских расходов.

Рынок недвижимости находится под влиянием ряда внешних факторов (политических, экономических, демографических). Например, экономические колебания могут вызвать изменения в стоимости недвижимости, политическая неопределенность может уменьшить интерес к недвижимости в определенной местности, изменения в демографической структуре, такие как старение населения или урбанизация, могут влиять на спрос на определенные типы недвижимости.

¹ А. Н. Асаул, Д. А. Гордеев, Е. И. Ушакова, Развитие рынка жилой недвижимости как самоорганизующейся системы – СПбБАСУ, 2008, С 32

² Kucharska-Stasiak E, Globalization and its Influence on Real Estate Valuations: The Case of Poland // International Trade and Finance Association Working Papers, 2006. С.79

Анализ изменения этих факторов лежит в основе определения нормальной цены недвижимости, которая является базой для принятия инвестиционных и потребительских решений.

Многие исследователи пытались определить присущие рынку характеристики, опираясь на экономические теории, объясняющие поведение рынка жилой недвижимости. В частности, это находит отражение в работах Сорнетт Д. и Вудард Р.¹, Стивенсона С.² и др.

Боя З. использует модель множественного регрессионного анализа для изучения влияния различных факторов на ценообразование на жилую недвижимость, в том числе таких как: денежная масса, население, инвестиции в развитие недвижимости, предложение и спрос, а также располагаемый доход домохозяйств и их сбережения³. Все эти факторы оказывают значительное влияние на изменение цен на недвижимость. Исследование Ахуджа А. и Малликамас Т. показывает, что индекс цен на недвижимость коррелирует с основными макроэкономическими показателями: валовым внутренним продуктом (ВВП), прямыми иностранными инвестициями и безработицей, а также структурой финансовой системы государства⁴. В тех странах, где домохозяйства склонны накапливать средства в банковской системе, негативное воздействие экономического шока на сектор недвижимости будет сильнее, чем в странах, где домохозяйства накапливают средства на финансовых рынках.

Согласно модели оценки стоимости капитальных активов, спрос на недвижимость зависит от ожидаемых будущих показателей аренды недвижимости (соотношение риска и доходности) и процентной ставки. Если аренда остается постоянной, цена на недвижимость снижается при увеличении процентной ставки. При снижении процентной ставки спрос на жилье увеличивается, что приводит к повышению цен на жилье.

Социальные факторы, ожидания потребителей и другие субъективные показатели также оказывают значительное влияние на стабильность рынка недвижимости⁵. Ожидания отражают поведение участников рынка недвижимости, имеющих финансовые обязательства перед банковским сектором, и определяют их решения о продаже или непродать активов. Социальные факторы определяются привычками и поведением людей, которые формируются самим обществом. Галиниене Б. подчеркивает следующие социальные факторы, влияющие на стабильность рынка недвижимости: размер семей, возрастная структура населения, изменения в населении, образ и уровень жизни, экономическая политика, образование, отношение к собственности⁶.

Российские исследователи значительное внимание уделяют влиянию на стоимость недвижимости инфраструктурных факторов. В монографии Асаула А.Н. и пр. приведены расчеты значимости различных внешних эффектов⁷. Наибольшие веса получила социальная инфраструктура, в состав которой включены близость поликлиник, магазинов, торговых точек и учебных заведений. Также в исследовании, посвященном анализу московского рынка недвижимости, выявлено, что значимость различных объектов инфраструктуры может меняться в зависимости от удаленности жилой недвижимости от МКАД и исторического центра столицы.

На основе проведенного обзора литературы вышеперечисленные факторы могут быть разделены на две большие группы: внутренние и внешние (рис. 1). К внутренним относятся собственные закономерности развития рынка, определяющие его поведение при изменении условий внешней среды, например сезонные колебания спроса, изменение объема предложения, взаимное влияние первичного и вторичного рынков.



Рис. 1. Факторы ценообразования на рынке недвижимости

Рынок недвижимости остается крайне сложным в прогнозировании рыночной стоимости, спроса, формирования пузырей и крахов. Однако, не существует единого набора факторов, которые оказывают на него влияние: в каждом случае исследователь самостоятельно выбирает исследуемые характеристики, а значимость факторов может варьироваться в зависимости от ситуации.

Говоря о ценообразовании, девелоперы традиционно использовали Excel и проводили ценообразование в ручном режиме, при этом основными недостатками ручного подхода являются:

- низкая скорость реагирования на изменения;
- субъективная оценка изменений менеджером/аналитиком;
- невозможность учета всех параметров.

¹ Sornette D, R. Woodard and W.-X. Zhou, Oil Bubble: evidence of speculation and prediction, Physica A 388 p – 2009, С 1571-1576

² Stevenson S., Modeling Housing Market Fundamentals: Empirical Evidence of Extreme Market Conditions //Real estate Economics, Vol. 36, Is. 1 – 2008

³ Boya, Z., An Empirical Analysis of the Factors Affecting Real Estate Price under the Background of Loose Monetary Policy //Advances in Economics, Business and Management Research, 106 – 2009

⁴ Ahuja, A., Mallikamas, T., Asset Price Bubble and Monetary Policy: Identification and Policy Response under Inflation Targeting //Office of Central Banking Studies, Monetary Policy Group, Bank of Thailand – 2003

⁵ Kindleberger, C. and Aliber, R., Manias, Panics and Crashes: A History of Financial Crises. //Palgrave Macmillan, 2005, 309 С.

⁶ Galiniene, B., & Marciuskas, A. The impact of housing conditions on Lithuanian households' subjective well-being //International Journal of Strategic Property Management, 18(1) – 2014, С 84–98

⁷ Асаул А., Д. А. Гордеев, Е. И. Ушакова Развитие рынка жилой недвижимости как самоорганизующейся системы //СПБГАСУ – 2008, С. 32

Однако в результате развития технологий и цифровизации некоторые компании начали внедрять системы динамического ценообразования. Это изменение стоимости объекта в зависимости от изменения значений определенных факторов. Динамическое ценообразование активно применяется во многих отраслях от классического ритейла до пассажирских перевозок и разнообразных услуг. Толчком к его развитию стало внедрение средств работы с большими данными и методов машинного обучения. Так стало возможным оперативно анализировать большой поток данных и мгновенно проводить переоценку товара. В недвижимости применение данных методов ограничено малым количеством данных.

На текущий момент динамическое ценообразование в жилой недвижимости реализовано в нескольких вариантах у крупных застройщиков. Остальные пока проводят расчеты в ручном режиме.

Внедрение динамического ценообразования позволяет:

- оперативно реагировать на изменения;
- максимизировать выручку;
- обеспечить равномерную реализацию квартир.

Представленные на рынке решения можно разделить на несколько классов с точки зрения использованных методов¹:

1. Алгоритмы на основе условий. Ценообразование производится на основе простых условных операторов. Преимуществом данного подхода является простота и интерпретируемость результата. Недостатком – фактор субъективности, так как данный метод выполняет действия, заложенные человеком.

2. Алгоритмы на основе статистических моделей. Расчет цен производится на основе линейной регрессии. Преимуществом является простота и интерпретируемость результата, недостатком – высокая погрешность по сравнению с современными модифицированными методами типа градиентного спуска и леса деревьев.

3. Алгоритмы на основе методов машинного обучения в том числе нейронных сетей. Наиболее распространенный подход – расчет цен методами машинного обучения на основе градиентного спуска. Недостатком является низкая интерпретируемость модели. Преимуществом – объективная оценка, низкая погрешность по сравнению с другими вариантами (в случае качественных данных).

Существует множество методов машинного обучения, которые могут быть применены в недвижимости. Эти алгоритмы анализируют исторические данные о ценах на недвижимость, включая факторы, такие как местоположение, размер, количество комнат, возраст здания и многие другие, чтобы предсказать будущие цены.

Более того, применение машинного обучения в ценообразовании недвижимости имеет ряд преимуществ. Во-первых, оно может обеспечить более точные прогнозы цен по сравнению с традиционными методами. Во-вторых, машинное обучение может анализировать гораздо больше факторов, чем человек, что позволяет учесть более сложные взаимосвязи и зависимости. Наконец, машинное обучение может автоматизировать процесс прогнозирования цен.

Машинное обучение используют для предсказания цен на недвижимость с использованием различных методов, включая линейную регрессию, деревья решений, случайные леса и метод опорных векторов. Например, в исследовании, проведенном Чилгином С. и Гекченом Г., использованы различные методы машинного обучения для прогнозирования цен на недвижимость в Турции. Были включены данные о продажах недвижимости за июнь и июль 2021 года в провинции Анкара, и применены различные методы машинного обучения, включая линейную, Лассо и Ридж регрессию, градиентный бустинг (XGBoost) и искусственные нейронные сети (ANN). Наилучшие результаты были достигнуты при использовании XGBoost и ANN².

При этом наиболее широко распространенные методы, используемые для оценки стоимости недвижимости, это случайный лес и градиентный бустинг.

• Случайный лес – это алгоритм обучения с учителем, который использует метод ансамблевого обучения для классификации и регрессии. Он запускает n количество деревьев регрессии и объединяет их в одну модель, чтобы сделать более точный прогноз, чем одно отдельное дерево³.

• Градиентный бустинг – это техника машинного обучения, которая создает прогностическую модель в виде ансамбля слабых прогностических моделей, обычно деревьев решений. Концепция бустинга заключается в преобразовании признака, который не является сильным в прогнозировании результата Y (слабый ученик), в сильного ученика.

Глубокое обучение, являющееся подразделом машинного обучения, становится все более важным в сфере недвижимости. Это область искусственного интеллекта, которая использует сложные нейронные сети для анализа больших объемов данных, обучения на основе этих данных и прогнозирования будущих результатов или поведения. В контексте недвижимости глубокое обучение может быть использовано для прогнозирования цен на недвижимость, что является важным инструментом для покупателей, продавцов, инвесторов и агентов по недвижимости.

С середины 1990-х годов до настоящего времени была опубликована серия исследовательских работ, посвященных разработке и применению моделей нейронных сетей для массовой оценки объектов недвижимости, среди них исследования Мухлишина М.⁴, Ванга С.⁵ и др.

В исследовании, проведенном в 2022 году, была разработана модель глубокого обучения для прогнозирования цен на квартиры в Токио в реальном времени. Модель использовала 27 различных факторов, включая экономические показатели, такие как инфляция и процентные ставки банка, которые меняются каждые три месяца. Это позволило модели обеспечить высокую точность прогнозов и минимизировать ошибки цен⁶.

Глубокое обучение представляет собой мощный инструмент для прогнозирования цен на недвижимость. Оно может обеспечить более точные прогнозы, обрабатывать большое количество переменных и обрабатывать неструктурированные данные. Однако, несмотря на все его преимущества, глубокое обучение все еще является относительно новой областью в недвижимости, и требует дальнейшего исследования для полного понимания его потенциала и ограничений.

¹ Venkatesan, R. and Craddock, J. and Nagji, N., Automation of Marketing Models. Darden Case No. UVA-M-0965, Available at <http://dx.doi.org/10.2139/ssrn.3291191> (Дата обращения: 01.02.2024)

² Gilgin, C., and Gokcen, H., Machine Learning Methods for Prediction Real Estate Sales Prices in Turkey //Revista de la Construcion. Journal of Construction, 22(1) – 20023, С 163-177

³ Груднев А.В. Прогнозное моделирование в IBM SPSS Statistics, R и Python: метод деревьев решений и случайный лес. – М.: ДМК Пресс, 2018. – 642 с.

⁴ Mukhlislin, M. E., Saputra, R., & Wibowo, A., (2017) Predicting house sale price using fuzzy logic, Artificial Neural Network and K-Nearest Neighbor //1st International Conference on Informatics and Computational Sciences – 2017, С 171–176.

⁵ Wang, C. C., Wu, H., A new machine learning approach to house price estimation //New Trends in Mathematical Sciences, 6(4) – 2018, С 165–171

⁶ Peng, T.-C., Wang, C.-C., The application of machine learning approaches on real-time apartment prices in the Tokyo metropolitan area //Social Science Japan Journal, 25 – 2022, С 3-28

Для оценки качества моделей используется ряд показателей: средняя абсолютная ошибка (MAE), средняя процентная ошибка (MAPE), среднеквадратическая ошибка (MSE), и корень из среднеквадратической ошибки (RMSE). Последние два показателя (формулы 1-2) являются наиболее широко распространенными, благодаря чему могут использоваться для сравнения качества разных моделей¹.

$$MSE = \frac{1}{m} \sum_{i=1}^m \left(h(x^{(i)}) - y^{(i)} \right)^2 \quad (1)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m \left(h(x^{(i)}) - y^{(i)} \right)^2} \quad (2)$$

где $h(x)$ – прогнозируемое значение, y – фактическое значение. В данном случае, чем ниже ошибка, тем лучше модель описывает данные.

Также показательной метрикой качества является коэффициент детерминации R^2 – мера, которая показывает, насколько хорошо модель подходит для данных. Коэффициент может принимать значения от 0 до 1, где 1 означает идеальное соответствие.

В табл. 1 приведены усредненные показатели метрик, используемых для оценки качества моделей. Как видно из представленных данных, методы градиентного спуска и нейронных сетей демонстрируют сравнимые результаты.

Табл. 1. Результаты ошибок, усредненных по исследованиям в разрезе методов

Метод	MSE	RMSE	R2
Линейная регрессия	0,1655	0,4068	0,7
Лассо	0,1654	0,4066	0,77
Ридж	0,1652	0,4065	0,82
XGBoost	0,1214	0,3484	0,90
ANN	0,1216	0,3488	0,92

Составлено авторами

Недостатком моделей, предназначенных для массовой оценки, является то, что они не учитывают общерыночные изменения, макроэкономические параметры. В результате требуется регулярное обновление и дообучение модели.

Поэтому перспективным направлением исследования является разработка модели, которая будет одновременно учитывать как строительные и эксплуатационные характеристики объектов недвижимости, так и некоторые параметры, характеризующие меняющуюся экономическую ситуацию в регионе, стране и мире. Это позволит модели адаптироваться ко времени; то есть их прогностические возможности не будут ухудшаться на фоне меняющейся экономической ситуации со временем.

Список литературы

- Ahuja A., Mallikamas T. Asset Price Bubble and Monetary Policy: Identification and Policy Response under Inflation Targeting //Office of Central Banking Studies, Monetary Policy Group, Bank of Thailand – 2003
- Boya Z. An Empirical Analysis of the Factors Affecting Real Estate Price under the Background of Loose Monetary Policy //Advances in Economics, Business and Management Research, 106 – 2009
- Çiğın, C., and Gökçen, H., Machine Learning Methods for Prediction Real Estate Sales Prices in Turkey //Revista de la Construcción. Journal of Construction, 22(1) – 20023, С 163-177
- Galiniene B., & Marčinskas A. The impact of housing conditions on Lithuanian households' subjective well-being //International Journal of Strategic Property Management, 18(1) – 2014, С 84–98
- Kindleberger C. and Aliber R., Manias Panics and Crashes A History of Financial Crises. //Palgrave Macmillan, 2005
- Kucharska-Stasiak E. Globalization and its Influence on Real Estate Valuations: The Case of Poland // International Trade and Finance Association Working Papers, 2006
- Mukhlisshin, M. F., Saputra, R., & Wibowo, A., (2017) Predicting house sale price using fuzzy logic, Artificial Neural Network and K-Nearest Neighbor //1st International Conference on Informatics and Computational Sciences – 2017, С 171–176.
- Peng, T.-C., Wang, C.-C., The application of machine learning approaches on real-time apartment prices in the Tokyo metropolitan area //Social Science Japan Journal, 25 – 2022, С 3-28
- Sornette D. R. Woodard and W.-X. Zhou Oil Bubble: evidence of speculation and prediction, Physica A 388 p – 2009, С 1571-1576
- Stevenson S. Modeling Housing Market Fundamentals: Empirical Evidence of Extreme Market Conditions //Real estate Economics, Vol. 36, Is. 1 – 2008
- Venkatesan, R. and Craddock, J. and Nagji, N., Automation of Marketing Models. Darden Case No. UVA-M-0965, Available at <http://dx.doi.org/10.2139/ssrn.3291191>
- Wang, C. C., Wu, H., A new machine learning approach to house price estimation //New Trends in Mathematical Sciences, 6(4) – 2018, С 165–171
- Асаул А.Н., Гордеев Д.А., Ушакова Е.И. Развитие рынка жилой недвижимости как самоорганизующейся системы – СПбГАСУ, 2008. – 334 С.
- Груздев А.В. Прогнозное моделирование в IBM SPSS Statistics, R и Python: метод деревьев решений и случайный лес. – М.: ДМК Пресс, 2018. – 642 с.

¹ Левин Б. Наука о данных. Как проверить качество модели с помощью метрик <https://nplu1.ru/material/2020/03/27/course-data-science-chapter-6> (Дата обращения: 16.02.2024)