

УДК 33      DOI: 10.14451/1.241.354

# Инструменты интеллектуального анализа данных для оценки инвестиций в недвижимость

© 2024 Шлычков Дмитрий Сергеевич

Кандидат экономических наук, доцент, доцент Кафедры аудита и корпоративной отчетности. Финансовый университет при Правительстве РФ, Россия, Москва.  
E-mail: DSSHlychkov@fa.ru

© 2024 Савченкова Елена Валерьевна

Кандидат экономических наук, доцент кафедры информатики ИКБ. МИРЭА – Российский технологический университет, Россия, Москва.  
E-mail: DSSHlychkov@fa.ru

© 2024 Терёшин Кирилл Александрович

Студент, Институт кибербезопасности и цифровых технологий, направление Фундаментальная информатика и информационные технологии, профиль Искусственный интеллект и машинное обучение. МИРЭА – Российский технологический университет, Россия, Москва.  
E-mail: DSSHlychkov@fa.ru

**Ключевые слова:** инвестиционный анализ, анализ данных, интеллектуальный анализ, инструменты анализа данных, линейная регрессия, логистическая регрессия, случайный лес, градиентный бустинг.

В современных условиях существует необходимость в дальнейшем развитии методов инвестиционного анализа с использованием инструментов интеллектуального анализа данных. В настоящее время исследователи все большее внимание уделяют разработке и развитию аналитических методов и методов оценки, основанных на искусственном интеллекте. В статье рассмотрены как традиционные методы инвестиционного анализа, так и предлагается использовать модели интеллектуального анализа данных.

Несмотря на значительные достижения в области оценки недвижимости, существующие методы зачастую ограничены в своей точности и применимости. Традиционные подходы, основанные на экспертных оценках или статистических моделях, могут не учитывать множество факторов, влияющих на цену недвижимости, а также не

обладать достаточной гибкостью для адаптации к изменяющимся рыночным условиям. Также стоит учитывать, что чаще всего модели стоимости жилой недвижимости создаются для определенного региона и не могут быть использованы без внесения изменений в других. В связи с этим возникает необходимость в дальнейшем раз-

вители методов инвестиционного анализа с использованием инструментов интеллектуального анализа данных, что, в сущности, и обуславливает актуальность вопросов оценки инвестиций в недвижимость.

Традиционно для стоимостной оценки инвестиций эксперты используют следующие методы: сравнительный, затратный и доходный [2].

Основная идея сравнительного метода заключается в том, чтобы оценить стоимость объекта недвижимости, сравнив его с ценами, полученными при продаже или аренде сходных объектов в том же регионе и в то же время. Этот метод предполагает, что рыночная стоимость объекта определяется рыночными ценами подобных объектов. Метод сравнительного анализа является ключевым в Международных стандартах оценки стоимости (IVS) и Европейских стандартах оценки стоимости (EVS) [11].

Затратный метод в оценке стоимости недвижимости представляет собой совокупность подходов, основанных на определении затрат, необходимых для восстановления или замещения объекта с учетом его износа. Он основан на принципе замещения, согласно которому предполагается, что покупатель не будет платить за недвижимость больше, чем стоимость постройки аналогичного объекта, учитывая его полезность [5].

Инвестиционный метод оценки стоимости жилой недвижимости и доходный метод представляют собой различные подходы к определению стоимости недвижимости, основанные на доходах, которые могут быть получены от объекта недвижимости. Основное различие между этими методами заключается в фокусе исследования.

Инвестиционный метод уделяет основное внимание прогнозированию будущих денежных потоков, которые могут быть получены от объекта недвижимости. Этот метод включает анализ возможных доходов, таких как арендные платежи или выручка от продажи, а также расходов на обслуживание и управление недвижимостью.

Важной частью инвестиционного метода является оценка резидуальной стоимости объекта недвижимости в конце анализируемого периода.

Кроме того, доходный метод сконцентрирован на анализе текущих или прошлых доходов, получаемых от объекта недвижимости. Этот метод обычно используется для определения текущей рыночной стоимости объекта на основе его доходности. Он также может включать в себя капитализацию ежегодных доходов для оценки стоимости объекта. По сравнению с другими подходами, доходный подход наиболее точно отражает ситуацию, когда инвестор рассматривает недвижимость как источник доход [1].

К недостаткам данных подходов к определению стоимости относят то, что они, как правило, дают различающиеся оценки. Помимо этого, часто учитывается ограниченный набор факторов, и конечная цена объекта обычно формируется экспертом, опирающимся на свой опыт, текущую обстановку на рынке недвижимости и имеющуюся информацию о воздействии макроэкономических и демографических факторов [10].

В настоящее время исследователи все большее внимание уделяют разработке и развитию аналитических методов и методов оценки, основанных на искусственном интеллекте.

Выбираемый метод во многом зависит от целей исследования и имеющихся данных. Так, например, в оценке стоимости жилой недвижимости на территории Москвы хорошо показала себя порядковая логистическая регрессия [10]. Целью исследования было оценить, насколько стоимость жилой недвижимости зависит от внутренних характеристик квартиры, таких как общая площадь, количество комнат, высота потолков, состояние ремонта, и от характеристик дома, которые создают так называемую комфортность объекта жилой недвижимости. К характеристикам дома, которые значительно влияют на комфортность проживания в квартире, авторы, в частности, отнесли транспортную доступность, возраст дома, этажность и тип дома. Средняя

ошибка прогноза оказалась равна 10%, при этом грубых ошибок модель не допустила [7].

В исследовании 2023 года авторы сравнивали эффективность трех различных методов интеллектуального анализа на данных для города Эймс, американский штат Айова [4]. Методами, которые они использовали, были линейная регрессия, случайный лес и градиентный бустинг. Были отмечены некоторые сильные и слабые стороны данных методов для оценки объектов недвижимости. Было отмечено, что линейная регрессия проста в интерпретации, но требует предположений о структуре данных и использует неизменяемую линейную функцию. Случайный лес характеризуется высокой точностью, но набор данных формируется случайным образом. Что касается градиентного бустинга, то он тоже обладает высокой точностью, но слишком чувствителен к выбросам. Несмотря на указанный недостаток, среди использованных моделей в исследовании градиентный бустинг показал наилучший результат.

В целом в работах зарубежных и российских исследователей существуют два основных вида параметров при прогнозировании стоимости недвижимости [6]. Во-первых, это анализ стоимости в зависимости от собственных характеристик объекта. К таким характеристикам, например, относятся площадь, этаж, на котором расположен объект и общее количество этажей в здании, материал стен, конструктивные элементы [6], наличие ремонта, парковки, удаленность от центра, транспортная доступность и имеющаяся инфраструктура [10]. Во-вторых, также распространен анализ стоимости в зависимости от разнообразных внешних факторов. В первую очередь к ним относятся изменение ключевой ставки, условия ипотечного кредитования [10], стоимость строительства, доходы населения, норма амортизации жилых зданий, налог на недвижимость, темп роста ВВП, уровень инфляции [6].

Факторы ценообразования на рынке жилой недвижимости:

Внутренние:

- Площадь жилья.
- Количество комнат.
- Этаж объекта и этажность.
- Материал стен.
- Наличие ремонта.
- Удаленность от центра.
- Транспортная доступность.
- Имеющаяся инфраструктура.

Внешние

- Ключевая ставка.
- Условия ипотеки.
- Стоимость строительства.
- Доходы населения.
- Норма амортизации зданий.
- Налог на недвижимость.
- Темп роста ВВП.
- Уровень инфляции.

Стоимость жилых объектов недвижимости является одним из ключевых факторов, которые влияют на решение потенциальных покупателей об инвестировании в недвижимость. Она также является важным показателем для инвесторов, которые рассматривают возможность вложения средств в недвижимость. Поэтому понимание того, какие факторы влияют на стоимость жилых объектов недвижимости, является важным для всех заинтересованных сторон.

Стоимость жилых объектов недвижимости – это многогранный показатель, зависящий от множества разнообразных факторов, которые взаимодействуют между собой и оказывают существенное влияние на формирование конечной цены объекта. Эти факторы могут быть классифицированы в соответствии с их характером и сферой влияния и, следовательно, могут быть подразделены на три основные категории [3]:

- 1) факторы общего характера;
- 2) локальные факторы;
- 3) характеристики жилого объекта.

Основные группы факторов, влияющие на стоимость объекта инвестиций:

- Факторы общего характера.
- Локальные факторы.
- Характеристики жилого объекта.

Факторы общего характера напрямую не связаны с оцениваемым объектом, но они представляют собой важный аспект в анализе цен на недвижимость, поскольку они формируют контекст и общую динамику рынка. Эти факторы также называют экзогенными и разделяют на факторы мирового, странового и регионального уровней [1].

К экзогенным факторам мирового уровня российские исследователи относят в первую очередь такие факторы как валютный курс и цена нефти [10]. Укрепление или ослабление национальной валюты оказывает воздействие на цены строительных материалов, инвестиционные потоки и покупательную способность населения, что отражается на стоимости объектов жилой недвижимости. Цена нефти также имеет существенное значение для экономики многих стран, включая Россию, как ключевой фактор доходов от экспорта и внешней торговли. Изменения в цене нефти могут сказаться на уровне экономического развития, занятости и доходов населения, что влияет на спрос и предложение на рынке недвижимости, а следовательно, и на цены на жилье.

Необходимо отметить, что все эти факторы не связаны с конкретным объектом недвижимости, поэтому напрямую оценить их влияние в стоимостном выражении достаточно трудно. Кроме того, в научных работах отмечается, что значительное влияние на российский рынок недвижимости оказывали такие события, как пандемия коронавируса и ужесточение мировых санкций в отношении России [1].

К факторам общего характера странового уровня относят реальную процентную ставку, реальный ВВП [10], уровень инфляции, существующие льготы на покупку жилья [1], политику в области налогообложения недвижимости, варианты финансирования строительства [3].

К региональным экзогенным факторам общего

относятся численность проживающего в регионе населения, уровень их доходов, климатические условия, региональные программы в области жилого фонда в аварийном состоянии [1], обеспеченность жильем, спрос и предложение [3].

Локальные факторы, влияющие на стоимость недвижимости, в отличие от факторов общего характера, непосредственно относятся к определенному объекту жилой недвижимости и характеризуют окружающие его условия. К таким факторам относят транспортную доступность [4], развитость инфраструктуры, наличие и состояние коммуникаций [3], экологическую обстановку [6], наличие таких удобств, как паркинг, охраняемая придомовая территория, консьерж [1].

К характеристикам объекта жилой недвижимости относятся его личные характеристики и характеристики здания, в котором он находится. Среди наиболее значимых характеристик здания следует отметить его год постройки, тип материала стен, общую этажность [1], наличие лифта, общее состояние здания [4].

В рамках данной статьи для прогнозирования стоимости инвестиций предлагается использовать следующие модели интеллектуального анализа данных: линейная регрессия, логистическая регрессия, случайный лес и градиентный бустинг. Ниже рассмотрены принципы и основы построения и анализа данных моделей прогнозирования.

Логистическая регрессия является одним из самых популярных методов статистического моделирования и анализа данных для задач классификации. Она используется для оценки вероятности принадлежности объекта к одной из двух или более категорий на основе независимых переменных.

Логистическая регрессия основана на предположении, что вероятность принадлежности объекта к определенной категории может быть выражена с помощью логистической функции,

которая является нелинейной функцией независимых переменных. Математически, логистическая функция может быть представлена в виде уравнения:

$$P = \frac{1}{1 + e^{-y}},$$

где  $P$  – вероятность наступления события, то есть значение, принадлежащее отрезку  $[0, 1]$ ,  $e$  – основание натурального логарифма,  $y$  – стандартное уравнение регрессии.

Зависимость вероятности наступления события от величины  $y$  изображена на графике (рис. 1).

Цель логистической регрессии состоит в том, чтобы оценить коэффициенты регрессии, которые максимизируют правдоподобие наблюдаемых данных, то есть вероятность получения наблюдаемых данных при заданных параметрах модели. Это делается с помощью метода максимального правдоподобия, который является стандартным методом оценки параметров логистической регрессии.

Логистическая регрессия может быть разделена на два основных типа: бинарная логистическая регрессия и многоклассовая логистическая регрессия. Бинарная логистическая регрессия используется для моделирования отношения между одной независимой переменной и одной зависимой переменной, которая имеет только две категории (например, 0 и 1). Порядковая логистическая регрессия используется для моделирования отношения между несколькими независимыми переменными и одной зависимой переменной, которая имеет три или более категорий. В работе зависимая переменная имеет четыре категории, поэтому для анализа используется именная порядковая логистическая регрессия.

Использование логистической регрессии требует предварительной обработки данных. Во-первых, она не может работать с пропусками в данных, поэтому пропущенные значения в объявлениях платформы «ЦИАН» были заполнены либо из текстовых описаний к объявлению, либо – если текст объявления не содержал нужной

информации – средним значением для непрерывных данных и модой для категориальных данных. Во-вторых, логистическая регрессия не может работать напрямую с категориальными данными, поэтому такие признаки как тип материала дома, состояние ремонта, район города были перекодированы в числовые. В-третьих, логистическая регрессия чувствительна к масштабу признаков, поэтому числовые признаки были стандартизированы.

Случайный лес (Random Forest) основан на методе ансамблевого обучения, который объединяет множество слабых моделей в одну сильную модель. Каждое дерево в случайном лесу строится на основе случайной выборки из обучающей выборки с заменой (bootstrap sample). При этом, на каждом шаге разбиения узла дерева используется только подмножество признаков, выбранное случайным образом из всего набора признаков. Это позволяет уменьшить корреляцию между деревьями и увеличить разнообразие моделей в ансамбле.

При классификации каждое дерево в лесу голосует за один из классов, и класс, получивший наибольшее количество голосов, считается окончательным решением.

Случайный лес имеет ряд преимуществ по сравнению с другими алгоритмами машинного обучения:

- обрабатывает как категориальные, так и числовые признаки;
- может обрабатывать пропуски в данных;
- устойчив к переобучению и шуму в данных;
- может автоматически определять взаимосвязи между признаками;
- может использоваться для оценки важности признаков.

Однако, случайный лес также имеет некоторые недостатки, среди которых можно выделить требовательность к ресурсам при работе с большими объемами данных, меньшую точность, чем у некоторых других алгоритмов, таких как, например, градиентный бустинг, также он может быть менее интерпретируемым, чем простые

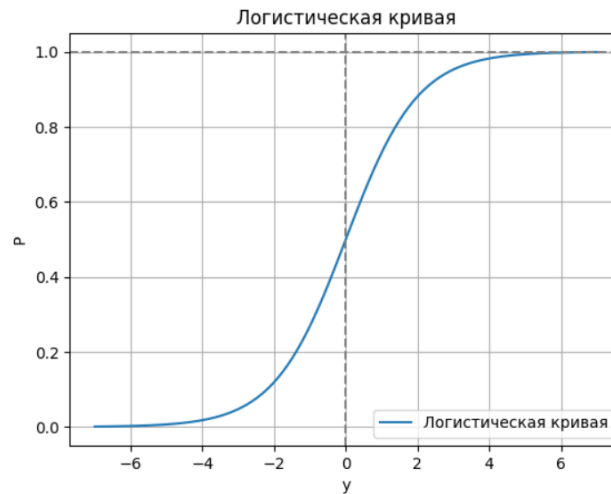


Рис. 1. Логистическая кривая.

модели, например, линейная регрессия.

Градиентный бустинг основан на методе ансамблевого обучения, который объединяет множество слабых моделей в одну сильную модель. Градиентный бустинг строит деревья последовательно, при этом каждое новое дерево строится таким образом, чтобы минимизировать ошибку предыдущего дерева. Алгоритм использует градиентный спуск для оптимизации функции потерь, которая измеряет разницу между предсказанными значениями и фактическими значениями целевой переменной.

На каждом шаге алгоритм вычисляет градиент функции потерь и строит новое дерево, которое наилучшим образом приближает градиент. Затем алгоритм обновляет значения целевой переменной, используя предсказания нового дерева, и повторяет процесс для следующего дерева. Этот процесс продолжается до тех пор, пока не будет достигнуто заданное число итераций или пока не будет достигнута заданная точность.

Стоит отметить преимущества градиентного бустинга для анализа данных. Он может работать как с числовыми, так и с категориальными признаками, присутствие пропусков в данных не является ограничением для его использования.

Для интеллектуального анализа данных особенно ценно то, что градиентный бустинг позволяет достигать очень высокой точности на многих задачах.

Однако градиентный бустинг также имеет некоторые недостатки, частично совпадающие с методом случайного леса. Он также бывает требовательным к ресурсам при работе с большими объемами данных и часто менее интерпретируемым, чем более простые модели. Помимо этого, он может быть чувствителен к переобучению, если не использовать регуляризацию или раннюю остановку.

Рынок жилой недвижимости в последние годы претерпел значительные изменения, адаптируясь к новым экономическим условиям, меняющимся потребительским предпочтениям и технологическим инновациям. Современный рынок характеризуется высоким уровнем конкуренции, разнообразием предложения и динамическими колебаниями цен.

В настоящее время исследователи все большее внимание уделяют разработке и развитию аналитических методов и методов оценки, основанных на искусственном интеллекте.

**Библиографический список**

1. Варфоломеева В. А. Актуальные проблемы и перспективы развития экономики: российский и зарубежный опыт. – 2021.
2. Васильева И. В., Алексеева Л. И., Соколов Е. А. Обзор рынка недвижимости в России и за рубежом после пандемии и мобилизации // Инновации и инвестиции. – 2023. – № 3. – С. 304–308.
3. Кропотухина Н. А., Хиревич С. А. Анализ факторов, влияющих на стоимость объектов жилой недвижимости // Инвестиции, градостроительство, недвижимость как драйверы социально-экономического развития территории и повышения качества жизни населения : Материалы XI Международной научно-практической конференции. – 2021. – С. 131–140.
4. Леонтьев Е. В., Майбуров И. А. Оценка влияния доступности общественного транспорта на стоимость городской жилой недвижимости // Journal of Applied Economic Research. – 2021. – Т. 20, № 1. – С. 62–83.
5. Малиновская А. В. Мировой опыт оценки недвижимости. Методология // Теория и практика эффективности государственного и муниципального управления : Сборник научных статей 2-й Всероссийской научно-практической конференции. – 2020. – С. 210–213.
6. Мирошниченко Н. В., Комиссарова И. В., Мирошниченко Д. А. Влияние экологических факторов на стоимость объектов недвижимости // Актуальные проблемы экологии и природопользования. – 2021. – С. 100–105.
7. Никитина Н. С. Прогнозирование индекса цен на недвижимость в России // Экономическое развитие России. – 2022. – № 29. – С. 23–28.
8. Николаева И. О., Ложкина Д. О., Домбровская О. В. Определение факторов, влияющих на стоимость квадратного метра жилой недвижимости г. Екатеринбурга // Агропромышленный комплекс: состояние, проблемы, перспективы. – 2022. – С. 759–761.
9. Оценка стоимости жилья в Российской Федерации. Методические рекомендации / Росстат. – 2019. – URL: <https://rosstat.gov.ru/folder/14458> (дата обр. 20.03.2024).
10. Проблемы моделирования оценки стоимости жилой недвижимости / Т. К. Богданова [и др.] // Бизнес-информатика. – 2020. – 14 (3). – С. 7–23.
11. Cupal M. The Comparative Approach Theory for Real Estate Valuation // Procedia – Social and Behavioral Sciences. – 2014. – No. 109. – P. 19–23.