

## ИННОВАЦИОННЫЕ ТЕХНОЛОГИИ МАССОВОЙ ОЦЕНКИ ЖИЛОЙ НЕДВИЖИМОСТИ

*А.Л. Острикова<sup>1</sup>, В.В. Селютин<sup>2</sup>*

<sup>1</sup> Южный федеральный университет, г. Ростов-на-Дону

<sup>2</sup> Федеральный исследовательский центр Южный научный центр РАН, г. Ростов-на-Дону  
annaostrikova00@yandex.ru, vvs1812@gmail.com

**Аннотация.** Массовая оценка является необходимым инструментом определения кадастровой стоимости и формирования налоговой базы регионов. В то же время действующие методы государственной кадастровой оценки опираются на традиционные статистические модели и не учитывают возможности, предоставляемые современными информационными технологиями.

В работе дан обзор современных подходов к массовой оценке урбанизированных территорий с использованием ГИС-технологий и искусственного интеллекта. Основная часть статьи посвящена демонстрации возможности улучшения качества массовой оценки путем использования инновационной технологии машинного обучения на примере жилой недвижимости Ростова-на-Дону. Для этого используются такие программные средства, как LGBost (Light Gradient Boosted Machine), XGBost (eXtreme Gradient Boosting) и CatBoost. В качестве зависимой переменной использовалась цена одного квадратного метра общей площади, в качестве независимых применялись такие параметры, как площадь кухни, общая площадь, тип этажа (первый, последний, остальные), количество этажей в доме, год постройки здания, тип объекта (новостройка, вторичный рынок), материал фасада дома, район города и микрорайон. Точность регрессионной модели, характеризуемая коэффициентом детерминации, составляет 0,68, в то время как модели машинного обучения дают точность 0,85.

Таким образом, модели машинного обучения можно рекомендовать в качестве практического инструментария для повышения качества оценки кадастровой стоимости жилой недвижимости.

**Ключевые слова:** массовая оценка, машинное обучение, жилая недвижимость.

Для эффективного управления земельными ресурсами и планирования регионального бюджета необходима кадастровая оценка стоимости недвижимости и земель населенных пунктов, поскольку именно она является базой для налогообложения. Налог на недвижимость является одним из важнейших источников бюджетных поступлений, позволяет регулировать рынок недвижимости и способствует перераспределению доходов, чтобы смягчить их поляризацию.

Создание системы налогообложения недвижимости является сложным процессом, включающим в себя не только соответствующую политику и законы, но и механизм и методы оценки. Большое количество оценок налоговой базы недвижимости необходимо проводить за относительно короткий промежуток времени. В то же время эта оценка должна соответствовать законодательству. На практике важно внедрить такую модель массовой оценки, которая соответствует структуре рынка недвижимости конкретной страны или ре-

гиона, а также является адаптивной к её изменениям с течением времени.

Массовая оценка характеризуется использованием методов, в которых осуществляется построение единых моделей определения кадастровой стоимости для групп объектов недвижимости, имеющих схожие характеристики. Главное отличие от индивидуальной оценки заключается в работе не с отдельно взятыми объектами, а набором таких объектов, что обуславливает применение стандартных унифицированных методик. Индивидуальная оценка оперирует более конкретными параметрами, которые являются характерными для локального рынка недвижимости, в то время как массовая оценка более широко использует методы статистического анализа для определения цен на основе расчетов по множеству локальных рынков недвижимости.

Массовая оценка недвижимого имущества изначально подразумевает выявление параметров группы объектов на определенную дату с исполь-

зованием общих данных, стандартизированных методов и статистического тестирования, которые в большей степени характеризуют стоимость недвижимого имущества, а также очевидных и скрытых зависимостей, объединяющих данные параметры [1].

Классические методы экономической и математической статистики подходят для моделирования реальной ситуации на рынке жилого недвижимого имущества, когда есть возможность проводить анализ на основе качественных исходных данных, при помощи которых несложно получить статистические допущения о характере распределения. Однако практика показывает, что далеко не всегда можно получить хорошую выборку исходной информации, имеющую полные, непротиворечивые данные. Эта сложность обуславливает неробастность модели – неустойчивость к выбросам и помехам, которые нередко можно встретить на реальном рынке недвижимости. Помимо этого, использование модели, построенной на маленьком наборе статистических данных, заведомо опасно, поскольку невозможно на их основе с помощью законов распределения определить истинные закономерности характеристик. Сравнения продаж, доходов и затрат, которые являются классическими подходами к оценке, при массовой оценке недвижимого имущества также не могут являться достаточными, поскольку на региональной основе требуется организация большого объема пространственных данных. С этой задачей могут справляться современные информационные технологии, основанные на машинном обучении с использованием нейронных сетей и географических информационных систем.

**1. Краткий обзор современных компьютерных технологий массовой оценки.** В последние годы в связи с появлением новых технологических возможностей внимание к совершенствованию теории и практики массовой оценки возросло. Это относится как к формированию критериев, так и методов оценки. При этом также необходимо учитывать коллизии, нередко возникающие в связи с недовольством собственников недвижимости результатами оценки.

В работе [2] об искусственных нейросетях в массовой оценке недвижимости предлагается сосредоточить внимание на теоретическом развитии искусственных нейронных сетей и массовой оценке недвижимости, эволюции моделей искус-

ственных нейронных сетей и совершенствовании существующих алгоритмов, а также практике и применении искусственных нейронных сетей. Исследование приводит ряд предложений массовой оценки недвижимого имущества в Китае.

В работе [3] авторы проводят сравнительный анализ нейросетевых моделей для оценки рыночной стоимости. Результаты оценок, полученные на основе моделей с помощью геоинформационных систем, проанализированы методом пространственного анализа. Сформулирован вывод о влиянии ценообразующих характеристик.

В статье [4] предлагается к разработке множество моделей автоматической оценки, использующих различные статистические и математические модели, такие как регрессия, нейронные сети и нечеткая логика. Анализ, проведенный на основе данных румынского города Орадя, позволяет сделать выводы о применимости геоинформационной системы (ГИС) при оценке недвижимости.

В исследовании [5] ученых Стамбульского технического университета представлен гибридный подход, основанный на интеграции ГИС и машинного обучения для автоматизированной массовой оценки жилой недвижимости в Турции и Соединенном Королевстве Великобритании. Оценки пространственных критериев, полученные в результате ГИС-анализа, включаются в данные для прогнозирования цен с целью дополнения характеристик. Результаты демонстрируют значительное увеличение точности прогнозирования при добавлении факторов местоположения к данным о ценах на недвижимость. Также авторы установили, что «случайный лес» является наиболее успешной регрессионной моделью по сравнению с другими ансамблевыми методами, использованными в работе.

В работе [6] была поставлена цель на основе данных района Пендик в Стамбуле определить потенциальные факторы, влияющие на цены жилья, путем применения разнообразных регрессионных моделей, включая обычные модели наименьших квадратов (OLS), модель пространственного запаздывания (SLM) и модель пространственной ошибки (SEM), а также изучить их географические вариации с помощью локальных регрессионных подходов, таких как географически взвешенная регрессия (GWR) и многомасштабная географически распределенная регрессия. Для этих целей была создана база геоданных с 28 атрибутами:

структурными, географическими переменными и переменными соседства. Первоначально 9 значимых переменных, используемых в качестве входных данных в других моделях, были выбраны с использованием поэтапного подхода OLS.

Результаты моделирования показали, что локальные модели достигают более высокой производительности по сравнению с глобальными моделями. Кроме того, были сопоставлены коэффициенты локального модельного фактора и локальные значения коэффициента детерминации  $R^2$  для 9 переменных. Географические вариации этих переменных в отношении цен на жилье были проанализированы для моделей GWR и MGWR. Данное исследование предоставило всеобъемлющую методологию для рынка недвижимости и городских властей, позволяющую понять географические различия факторов, влияющих на рынок недвижимости в городской географии.

Исследование автоматизации процесса оценки земли [7], утверждает, что консолидация земель является эффективной процедурой управления земельными ресурсами, направленная на устойчивое развитие. Важнейшим вопросом в схемах консолидации земель является оценка земель, которая представляет собой тип массовой оценки, обычно проводимой с использованием эмпирического процесса. В результате этот процесс отнимает много времени, является дорогостоящим, а результаты не являются достаточно точными, надежными и последовательными. Решением этих проблем является использование автоматизированных моделей оценки (AVM).

В данной статье представлена разработка, внедрение и оценка двух гедонических ценовых моделей, основанных на линейной и нелинейной функции в сочетании с географической информационной системой, примененных к тематическому исследованию на Кипре. Модели протестированы с точки зрения обеспечения качества в соответствии с международными стандартами. Оценка, основанная на международных стандартах, показала, что обе модели дали очень хорошие результаты, а небольшая выборка стоимости земли, предоставленная Комитетом по оценке земель, достаточна для автоматизации процесса оценки земли, что приводит к достаточной точности, надежности и согласованности. Следовательно, AVM является высокоэффективным по сравне-

нию с традиционными методами оценки земли, поскольку он может значительно сократить используемое время и ресурсы и обеспечить прозрачность, поскольку процесс был преобразован из эмпирического в систематический, аналитический и стандартизированный.

**2. Построение и сравнительный анализ моделей.** Для анализа были использованы бесплатные источники, данные собраны и систематизированы с помощью метода синтаксического анализа – парсинга веб-ресурса. Одними из наиболее популярных сайтов для продажи, покупки и аренды недвижимости являются «Циан» (rostov.cian.ru) и «Рестэйт» (restate.ru), сведения из баз данных которых и легли в основу исследования. В результате удаления объектов с незаполненными признаками и тримминга итоговая выборка для построения модели массовой оценки жилой недвижимости Ростова-на-Дону составила 6847 объектов.

### 2.1. Регрессионная модель

В качестве первичного набора факторов, определенных экспертным путем и представленных на указанных сайтах, использовались следующие характеристики:

- стоимость квартиры, тыс. руб. –  $Y_1$ ;
- цена за  $1 \text{ м}^2$ , тыс. руб. –  $Y_2$ ;
- общая площадь, кв. м –  $X_1$ ;
- площадь кухни, кв. м –  $X_2$ ;
- жилая площадь, кв. м –  $X_3$ ;
- количество комнат –  $X_4$ ;
- тип этажа (первый, последний, остальные) –  $X_5$ ;
- количество этажей в доме –  $X_6$ ;
- год постройки здания –  $X_7$ ;
- тип объекта (новостройка, вторичный рынок) –  $X_8$ ;
- материал фасада дома –  $X_9$ ;
- район города –  $X_{10}$ ;
- микрорайон –  $X_{11}$ .

Анализ корреляционной матрицы позволил исключить два фактора –  $X_3$  (жилая площадь) и  $X_4$  (количество комнат), значимо коррелированных с переменной  $X_1$  (общая площадь).

С помощью функции “LinearRegression” библиотеки Python “sklearn” из модуля “linear\_model” была построена модель линейной регрессии по оставшимся 9 факторам, где в качестве зависимой переменной была выбрана цена за один квадратный метр квартиры ( $Y_2$ ). В качестве тестовой выборки использовалось 80 % от общего объема

датафрейма. Результаты расчета представлены на диаграмме рассеяния (рис. 1).

## 2.2. LGBBoost

На следующем этапе была построена модель методом машинного обучения Boosting с помощью библиотеки LGBBoost. Light Gradient Boosted Machine – это библиотека Python с открытым исходным кодом, предоставляющая эффективную реализацию алгоритма градиентного бустинга. Градиентный бустинг является продвинутым методом машинного обучения для решения задач классификации и регрессии. Он предполагает построение предсказания как ансамбль слабых моделей прогнозирования, которыми обычно являются деревья решений. Из нескольких таких моделей в итоге получается единая, но уже эффективная модель.

LGBBoost является расширенным алгоритмом градиентного бустинга, который добавляет тип автоматического выбора объектов и фокусируется на примерах бустинга с большими градиентами. Это может приводить к резкому ускорению обучения, а также заметному улучшению показателей прогнозирования.

Градиентный бустинг относится к классу ансамблевых алгоритмов машинного обучения. Ансамбли строятся на основе моделей дерева решений. Деревья по одному добавляются в ансамбль и обучаются с целью исправления ошибок прогнозирования, которые были сделаны предыдущими моделями.

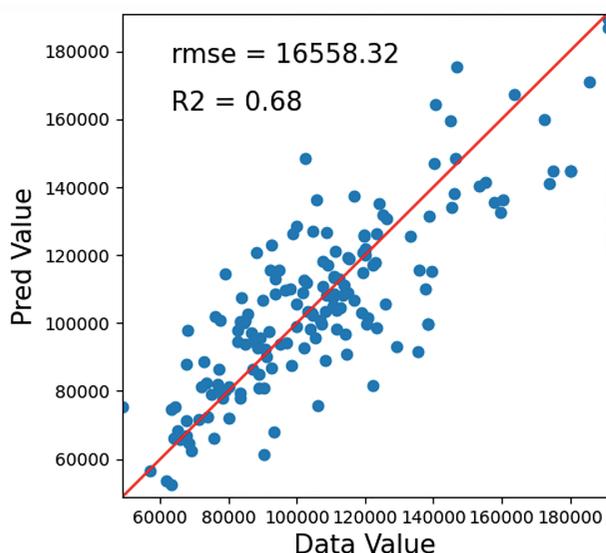


Рис. 1. Диаграмма рассеяния для модели множественной регрессии

Модели могут обучаться с использованием любой произвольной дифференцируемой функции потерь, а также алгоритма оптимизации градиентного спуска. Всё это дает методу название градиентного бустинга, поскольку градиент потерь по мере обучения модели минимизируется подобно нейросети.

В качестве зависимой переменной принималась стоимость квартиры за квадратный метр «price\_per\_m2», а на основе анализа статистической модели к независимым переменным отнесены следующие:

- площадь кухни, кв. м –  $X_1$ ;
- общая площадь, кв. м –  $X_2$ ;
- тип этажа (первый, последний, остальные) –  $X_3$ ;
- количество этажей в доме –  $X_4$ ;
- год постройки здания –  $X_5$ ;
- тип объекта (новостройка, вторичный рынок) –  $X_6$ ;
- материал фасада дома –  $X_7$ ;
- район –  $X_8$ ;
- микрорайон –  $X_9$ .

Данные разбиваются на обучающую и тестовую выборки с помощью кросс-валидации  $k$ -fold: данные разбиваются на  $k$  частей, на  $k - 1$  частях производится обучение модели, а оставшаяся часть используется для тестирования. Полагаем  $k$  равным 5.

Методом “LGBMRegressor” библиотеки “lightgbm” формируется модель, для которой задаются параметры “objective = ‘regression’” и “metrics = ‘mape’”. Первый означает, что метод используется для регрессионной задачи. Второй задает метрику ошибки, которой оценивается модель, – MAPE. Далее происходит обучение, после чего формируется оценка полученной модели и строится график с признаками в порядке убывания значимости. Среднее значение MAPE = 0,1349, откуда получаем точность модели, определяемую коэффициентом детерминации  $R^2 = 86,5$  %. Это значительно более высокое качество аппроксимации, чем у регрессионной модели. Наиболее важным признаком является переменная  $X_2$  – общая площадь. Наименее значимым –  $X_6$  тип объекта.

## 2.3. XGBoost

XGBoost (eXtreme Gradient Boosting) является разновидностью Boosting. Это пример ансамблевого метаалгоритма машинного обучения, который может применяться с целью снижения смещения и дисперсии в обучении с учителем, а

также семействе алгоритмов машинного обучения, создающие из слабых моделей более сильные. Данный метод отвечает на вопрос о возможности «слабых» обучающихся алгоритмов с результатами ненамного лучше случайного гадания в PAC модели (вероятно, приблизительно правильной) усовершенствоваться до «сильного» обучающегося алгоритма произвольной точности.

Таким образом, основополагающий принцип метода бустинга заключается в последовательном применении слабых алгоритмов обучения с каждым последующим уменьшением смещения всей модели. Такой алгоритм позволяет преобразовать слабые модели в мощную ансамблевую. Существует много различных вариаций метода бустинга, такие как, например, LPBoost, градиентный бустинг, а также AdaBoost – адаптивный бустинг, подстраивающийся под слабые алгоритмы обучения.

Схема градиентного бустинга реализована на языке Python в библиотеке с открытым исходным кодом “XGBoost”. В ней модели строятся поэтапно, аналогично тому, как это происходит при использовании других методов бустинга. Обобщает слабые обучающиеся алгоритмы, допуская оптимизацию произвольной дифференцируемой функции потерь (функции потерь с вычислимым градиентом).

XGBoost включает оригинальный алгоритм машинного обучения, основанный на решающих деревьях и пригодный для работы с разреженными данными – теми, которые в основном содержат неиспользуемые элементы, не несущие никакой информации, например, набор данных, в котором большинство значений элементов равны нулю. Теоретически обоснованная процедура позволяет работать с весами различных элементов в обучении деревьев. Алгоритм обладает следующими преимуществами:

1. Поддерживает реализацию параллельных и распределенных вычислений. Главное отличие от многих других алгоритмов бустинга заключается в скорости работы обучения – здесь оно может проводиться параллельно, значительно сокращая время;
2. XGBoost предоставляет надежные, готовые к использованию средства регуляризации наряду с набором параметров по настройке этого процесса. Среди параметров представлены вес для L1-регуляризации, вес для L2-регуляризации,

минимальное уменьшение функции потерь для дальнейшего деления дерева, максимальная глубина дерева, минимальная сумма весов всех наблюдений для дочернего объекта.

В XGBoost суммирование ответов по всем деревьям ансамбля описывается следующей формулой:

$$F(x) = \sum_{k=1}^K f_k(x), \quad (1)$$

где  $\{x, y\}_{i=1}^N$  – обучающая выборка,  $K$  – количество деревьев в ансамбле,

$f_k: x \rightarrow y$  – функция  $k$ -го дерева ансамбля,  $F: x \rightarrow y$  – функция всего ансамбля.

Функция потерь выглядит следующим образом:

$$total_{loss} = \sum_{i=1}^N loss(y_i, F(x_i)) + \gamma \sum_{k=1}^K T_k + \frac{1}{2} \lambda \sum_{k=1}^K \|w_k\|^2, \quad (2)$$

где  $\gamma, \lambda$  – гиперпараметры.

Первое слагаемое задает основную функцию потерь. Второе – штрафы деревьев за слишком большое количество листьев. Третье обеспечивает минимальный вклад в результат каждым деревом, что не является типичным в машинном обучении. Результат моделирования в виде диаграммы рассеяния представлен на рисунке 2.

Среднее значение MAPE = 0,1346, откуда получаем точность модели 86,5 %, как и в модели LGBBoost. Наиболее важным признаком является переменная  $X_8$  – административный район. Наименее значимым –  $X_3$  тип этажа.

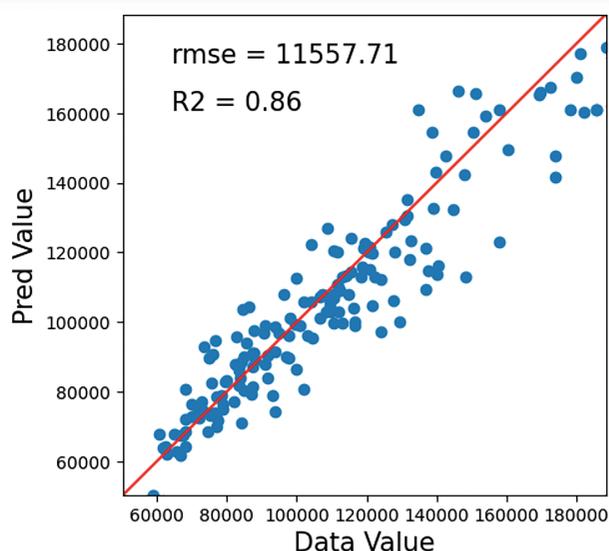


Рис. 2. Диаграмма рассеяния для модели машинного обучения XGBoost

## 2.4. CatBoost

CatBoost – открытая программная библиотека градиентного бустинга на Python, созданная разработчиками Яндекса. В ней используются «забывчивые» (oblivious) деревья решений, при помощи которых дерево получается сбалансированным.

Главными преимуществами метода являются улучшенная скорость вычислений и поддержка категориальных входных переменных. Второе достоинство обосновывает название CatBoost, от «Category Gradient Boosting» (Категориальный градиентный бустинг).

Среди наиболее распространенных способов обработки категориальных данных в машинном обучении являются «one-hot» кодирование и кодирование лейблов. CatBoost позволяет использовать категориальные признаки без необходимости предварительной обработки. Здесь не требуется пользоваться «one-hot» кодированием или любым другим, поскольку это влияет на скорость обучения и на качество прогнозов. Вместо этого необходимо передать категориальные признаки в параметр «cat\_features».

Данный метод имеет ряд недостатков: чувствительность алгоритма к выбросам, переобучение при слишком большом количестве деревьев (данная проблема существует в любом алгоритме с деревьями, решается правильной настройкой параметра «n\_estimators»), а также большое количество времени, затрачиваемого на обучение модели. Поэтому если входной набор данных является большим, то требуется корректировать параметр «min\_samples\_leaf».

Данные разбиваются на обучающую и тестовую выборки с помощью кросс-валидации  $k$ -fold на 5 частей.

Методом «CatBoostRegressor» из библиотеки «catboost» формируется модель с параметром «eval\_metric='MAPE'»; задающим в качестве метрики для оценивания ошибки «MAPE». Далее происходит обучение и формируется оценка полученной модели с построением графика признаков в порядке убывания их значимости. Как и в двух рассмотренных выше моделях, точность аппроксимации данных составила 86,5 %. Наиболее важным признаком здесь оказалась переменная  $X_4$  – количество этажей. Наименее значимым –  $X_6$  тип объекта.

**Выводы.** Проведенное исследование моделей массовой оценки жилой недвижимости на приме-

ре Ростова-на-Дону позволяет сделать следующие выводы.

Использование моделей машинного обучения существенно повышает качество массовой оценки по сравнению с традиционными статистическими моделями на основе линейной регрессии. При этом все три использованные модели дают одинаковую точность аппроксимации исходных данных.

Однако скорость работы библиотек существенно различается. Если на данном объеме выборки модели для первых двух методов обучались за считанные секунды, то метод «CatBoost» обучал модель пару минут. Это может быть существенным недостатком при значительном увеличении количества данных.

Следует также отметить, что значимость ценообразующих факторов в использованных моделях различна (табл. 1).

Статистическая модель и LGBost определили общую площадь  $X_2$  как признак, наиболее влияющий на формирование стоимости квадратного метра площади квартиры. XGBoost поставил в приоритет переменную  $X_8$  – административный район, CatBoost – количество этажей в здании  $X_4$ . Порядок наименее значимых признаков у моделей отличается, однако топ-3 наименее значимых факторов образован одинаковыми переменными: тип этажа  $X_3$ , тип объекта  $X_6$ , материал фасада дома  $X_7$ . Статистическая модель линейной регрессии по ранжированию факторов очень близка к LGBost, единственное отличие заключается в порядке определения значимости переменных «площадь кухни»  $X_1$  и «количество этажей в доме»  $X_4$ .

**Таблица 1. Ценообразующие факторы моделей**

Признаки в порядке убывания значимости	Статистическая модель	LGBost	XGBoost	CatBoost
1	$X_2$	$X_2$	$X_8$	$X_4$
2	$X_5$	$X_5$	$X_4$	$X_8$
3	$X_4$	$X_1$	$X_5$	$X_2$
4	$X_1$	$X_4$	$X_2$	$X_5$
5	$X_8$	$X_8$	$X_9$	$X_9$
6	$X_9$	$X_9$	$X_1$	$X_1$
7	$X_7$	$X_7$	$X_7$	$X_7$
8	$X_3$	$X_3$	$X_6$	$X_3$
9	$X_6$	$X_6$	$X_3$	$X_6$

Таким образом, модели машинного обучения можно рекомендовать в качестве инновационного инструментария для повышения качества оценки кадастровой стоимости жилой недвижимости.

*Публикация подготовлена в рамках ГЗ ЮНЦ РАН, № ГР проекта 122020100349-6.*

#### Список литературы

1. Wang D., Li V.J. Mass Appraisal Models of Real Estate in the 21<sup>st</sup> Century: A Systematic Literature Review // Sustainability. 2019. Vol. 11. No. 24. P. 1–14. DOI: 10.3390/su11247006
2. Zhou G., Ji Y., Chen X., Zhang F. Artificial Neural Networks and the Mass Appraisal of Real Estate // International Journal of Online Engineering. 2018. Vol. 14. No. 3. P. 180–187. DOI: 10.3991/ijoe.v14i03.8420
3. Yalpir S., Durduran S.S., Unel F.B., Yolcu M. Creating a valuation map in GIS through artificial neural network methodology: A case study // Acta Montanistica Slovaca. 2014. Vol. 19. No. 2. P. 79–89.
4. Droj G., Droj L. GIS based automated valuation models – a genuine solution for real estate valuation in Romania // “1 Decembrie 1918” University of Alba Iulia, RevCAD 20/2016. P. 45–52.
5. Mete M.O., Yomralioglu T. Mass Valuation of Real Estate Using GIS-based Nominal Valuation and Machine Learning Methods // European Real Estate Society, ERES 2022\_177. P. 1–7.
6. Sisman S., Aydinoglu A.C. A modelling approach with geographically weighted regression methods for determining geographic variation and influencing factors in housing price: A case in Istanbul // Land Use Policy. 2022. Vol. 119. 106183. DOI: 10.1016/j.landusepol.2022.106183
7. Demetriou D. Automating the land valuation process carried out in land consolidation schemes // Land Use Policy. 2018. Vol. 75. P. 21–32. DOI: 10.1016/j.landusepol.2018.02.049

## INNOVATIVE TECHNOLOGIES FOR MASS VALUATION OF RESIDENTIAL REAL ESTATE

A.L. Ostrikova<sup>1</sup>, V.V. Selyutin<sup>2</sup>

<sup>1</sup> Southern Federal University, Rostov-on-Don,

<sup>2</sup> Federal Research Centre the Southern Scientific Centre of the Russian Academy of Sciences, Rostov-on-Don,  
annaostrikova00@yandex.ru, vvs1812@gmail.com

**Abstract.** Mass valuation is a necessary tool for determining the cadastral value and forming the tax base of the regions. At the same time, the current methods of state cadastral valuation are based on traditional statistical models and don't take into account the opportunities provided by modern information technologies. The first part of this study gives a brief overview of modern approaches to mass valuation of urbanised territories using GIS-technologies and artificial intelligence. The main part of the work is devoted to demonstrating the possibility of improving the quality of mass assessment by using innovative machine learning technology on the example of residential real estate in Rostov-on-Don. For this purpose such software tools as LGBost (Light Gradient Boosted Machine), XGBoost (eXtreme Gradient Boosting) and CatBoost are used. The dependent variable was the price of one square metre of total floor area, independent variables were kitchen area, total floor area, type of floor (first, last, remaining), number of floors in the house, year of construction of the building, type of object (new construction, secondary market), material of the house facade, city district and microdistrict. The accuracy of the regression model, characterised by the coefficient of determination, is 0.68, while machine learning models give an accuracy of 0.85. Thus, machine learning models can be recommended as a practical toolkit to improve the quality of cadastral value assessment of residential real estate.

**Keywords:** mass valuation, machine learning, real estate.

### References

1. Wang D., Li V.J. Mass Appraisal Models of Real Estate in the 21<sup>st</sup> Century: A Systematic Literature Review. *Sustainability*. 2019. Vol. 11. No. 24. P. 1–14. URL: <https://doi.org/10.3390/su11247006>
2. Zhou G., Ji Y., Chen X., Zhang F. Artificial Neural Networks and the Mass Appraisal of Real Estate. *International Journal of Online Engineering*. 2018. Vol. 14. No. 3. P. 180–187. URL: <https://doi.org/10.3991/ijoe.v14i03.8420>
3. Yalpir S., Durduran S.S., Unel F.B., Yolcu M. Creating a valuation map in GIS through artificial neural network methodology: A case study. *Acta Montanistica Slovaca*. 2014. Vol. 19. No. 2. P. 79–89.
4. Droj G., Droj L. GIS based automated valuation models – a genuine solution for real estate valuation in Romania. “1 Decembrie 1918” University of Alba Iulia, RevCAD 20/2016. P. 45–52.
5. Mete M.O., Yomralioglu T. Mass Valuation of Real Estate Using GIS-based Nominal Valuation and Machine Learning Methods. *European Real Estate Society, ERES 2022\_177*. P. 1–7.
6. Sisman S., Aydinoglu A.C. A modelling approach with geographically weighted regression methods for determining geographic variation and influencing factors in housing price: A case in Istanbul. *Land Use Policy*. 2022. Vol. 119. P. 106183. URL: <https://doi.org/10.1016/j.landusepol.2022.106183>
7. Demetriou D. Automating the land valuation process carried out in land consolidation schemes. *Land Use Policy*. 2018. Vol. 75. P. 21–32. URL: <https://doi.org/10.1016/j.landusepol.2018.02.049>