

Распознавание агентских объявлений об операциях с недвижимостью с использованием методов машинного обучения

И.С. Зеленский

Аннотация—В настоящей статье описано решение задачи валидации открытых данных о состоянии жилого фонда с целью дальнейшего использования проверенных данных в процессе мониторинга состояния городской среды. Задача валидации в контексте настоящего исследования состоит в автоматическом распознавании потенциально недостоверных объявлений, опубликованных риелторами и/или риелторскими агентствами (сокращенно «агентами») в сети Интернет. Автоматизация задачи распознавания осложнена отсутствием формальных критериев принадлежности объявлений агентам, зависимостью существующих методов от дополнительных источников информации и необходимостью привлечения дополнительных ресурсов. Рассмотрены существующие методы решения данной, а также подобных задач распознавания недостоверной информации на примере решения задачи выявления ложных новостей. Выделены две группы методов решения задачи: социально-ориентированные и на основе анализа содержимого. Первые методы слабо применимы к задаче распознавания агентских объявлений, вторые же позволяют делать вывод о принадлежности данных к множеству достоверных либо недостоверных на основе методов анализа естественного языка и решения задачи классификации, что является синонимом задачи валидации, поставленной в рамках настоящей работы. Результатом проведенного анализа является постановка задачи двоячной классификации объявлений на основе методов машинного обучения. Для обучения и тестирования моделей классификации использована заранее размеченный набор данных, состоящий из объявлений об операциях с квартирами г. Москва из открытых источников. Проведено сравнение шести моделей классификации по целевым метрикам: accuracy, precision, recall, F1 и ROC AUC метрикам. Выбрана лучшая по метрикам модель и проведена оптимизация её гиперпараметров.

Ключевые слова—обучение, искусственный интеллект, анализ данных, классификация, недвижимость, распознавание агентов, валидация данных.

Статья получена 14 января 2025.

Исследование выполнено за счет гранта Российского научного фонда и Администрации Волгоградской области № 22-11-20024, <https://rscf.ru/project/22-11-20024/>.

Зеленский Илья Сергеевич, аспирант кафедры «Цифровые технологии в урбанистике, архитектуре и строительстве» Волгоградского государственного технического университета; ORCID 0000-0001-6057-4549; Волгоград, Россия. (e-mail: timkaa525@yandex.ru).

I. ВВЕДЕНИЕ

Благосостояние человечества во многом определяется качеством жизни в городах. Структура современных городов характеризуется сложным пространственным размещением людей и видов деятельности, неравномерностью охвата инфраструктурой, что приводит к повышению сложности задач управления развитием городов. Для решения этих задач необходимо анализировать множество факторов, влияющих на качество жизни в городах, и совершенствовать способы получения информации о них.

Стремление к созданию качественной, гармонично развитой и устойчивой городской среды зафиксировано в законодательстве РФ [1], а также на мировом уровне 11-й позицией в списке целей устойчивого развития ООН [2]. Таким образом, возникает потребность в регулярном мониторинге состояния городской среды для своевременного получения полной информации о ней и принятия необходимых управленческих решений. Задача является актуальной для РФ и имеет ряд существующих решений [3, 4].

Под состоянием городской среды понимается, в том числе, состояние жилого фонда [5], наиболее актуальная информация о котором содержится в пользовательских объявлениях об операциях с недвижимостью, публикуемых в открытых источниках [6].

При работе с открытыми данными, источником которых является человек (в контексте настоящей работы – пользователи сети Интернет, публикующие объявления об операциях с недвижимостью), необходимо принимать во внимание вероятность публикации заведомо недостоверных сведений. Таким образом, возникает задача контроля достоверности или задачи валидации данных. В контексте объявлений об операциях с недвижимостью одним из возможных источников недостоверных данных могут служить объявления, опубликованные риелторами или риелторскими агентствами (далее – агентами). В случае если автором объявления является агент, возможно завышение характеристик относительно реальных либо описание объекта, не существующего в принципе. Целью создания подобных «виртуальных» объектов со стороны агента может быть, в частности, навязывание услуг и привлечение клиентов. Утверждать, что любые

объявления агентов несут недостоверную информацию, нельзя, однако однозначно подтвердить или опровергнуть недостоверность, основываясь только на открытых данных, также не представляется возможным. В контексте же мониторинга состояния городской среды важно опираться строго на реально существующие объекты с целью получить достоверную оценку реального положения вещей.

Таким образом, задача настоящей работы состоит в автоматическом разделении множества объявлений на два подмножества: объявления, опубликованные агентами, и объявления, опубликованные рядовыми пользователями. При этом объявления в части открытых источников, как правило, уже отмечены принадлежностью к одному из двух подмножеств, что сводит постановку задачи распознавания агентских объявлений к задаче классификации [7]. Целью настоящего исследования является получение модели, способной определить потенциально агентское объявление, принадлежность которого к определенному классу заранее не известна, отталкиваясь от данных, представленных в объявлении. Полученную модель можно будет использовать для автоматического определения и разметки потенциально недостоверных данных.

II. СУЩЕСТВУЮЩИЕ РЕШЕНИЯ

Существующие подходы к выявлению недостоверной информации в сети Интернет, противодействию мошенничеству в целом и применительно к сделкам с недвижимостью в частности часто основаны и возложении ответственности за достоверность публикуемой информации на администрацию площадки на которой публикуется информация (выделенных модераторов) и/или конечных пользователей, а в качестве методов валидации данных предлагается применять человеко-ориентированные подходы на основе косвенных критериев с привлечением дополнительных внешних данных и ресурсов [8].

Когда речь заходит о методах, слабо поддающихся формализации, основанных на неявных и/или нечетких зависимостях, предлагается применение методов искусственного интеллекта и машинного обучения [9]. Так, для перекрестной сверки различных открытых источников исследователи применяют методы парсинга веб-страниц [10], работы с естественными языками (методы text mining) [11], а для анализа изображений, прилагаемых к объявлениям – нейронные сети [10], а также хэш-алгоритмы [12] с целью выявить дубликаты изображений, приложенные к публикациям, описывающим разные объекты. Однако данные методы требуют наличия нескольких источников информации и вхождения в них объявлений с дублирующейся информацией (графической или текстовой), и не способны определить потенциально недостоверные данные в отдельно взятых источниках.

Целесообразно рассмотреть также опыт выявления недостоверных новостей (англ. fake news) в сети Интернет. В данном случае исследователи различают

методы анализа, определяющие достоверность информации на основе содержания новости (англ. content-based), и социально-ориентированные (англ. social-based) методы, в основе которых лежит анализ распространения новости, информации об авторе и т.д. [13]. Социально-ориентированные методы слабо применимы к задаче выявления недостоверных объявлений об операциях с недвижимостью, т.к. анализируемые данными методами критерии, например, дата регистрации автора на соответствующей интернет-площадке, применительно к объявлениям малозначимы. Пользователь может быть зарегистрирован на площадке лишь для того, чтобы разместить единственное объявление, наполненное достоверной информацией, провести операцию по нему, и более к площадке не возвращаться. Методы на основе содержания, в свою очередь, отталкиваются от текстов новостей, их стилистических особенностей и смысловой нагрузки, а суть данного подхода сводится к решению задач классификации различными методами машинного обучения и обработки естественных языков [14, 15]. Таким образом, задача настоящего исследования сводится к разработке метода анализа данных на основе содержания.

III. РЕШЕНИЕ ЗАДАЧИ КЛАССИФИКАЦИИ

A. Постановка задачи

Принцип определения принадлежности объявления конкретному классу заранее неизвестен, однако на тематических сайтах для публикации объявлений имеется значительный массив объявлений, уже отнесенных к одному из классов. Таким образом, для решения задачи целесообразно было прибегнуть к методам искусственного интеллекта и машинного обучения [16].

В рамках настоящего исследования была поставлена задача создания и обучения модели для двоичной классификации [7] объявлений об операциях с недвижимостью с целью разделить множество объявлений на два класса: объявления, опубликованные агентами, и объявления, опубликованные рядовыми пользователями.

Основой обучения модели стали данные о недвижимости, собранные из открытых источников [17], представленные в виде пользовательских объявлений об операциях с квартирами (сдача, продажа) города Москвы. Размер итогового набора данных составил 6457 записей. Описание структуры данных и возможных значений полей приведено в Таблице 1.

В составе набора данных следует выделить поле «Тип автора», отражающее принадлежность объявления физическим лицам («Арендодатель»/ «Частное лицо») либо агентам («Агентство»/«Застройщик»/«Риелтор»). Данное поле стало целевым параметром будущей модели.

Таблица 1 – Описание структуры набора данных

Название поля	Возможные значения	Описание
Тип автора	«Агентство», «Арендодатель», «Частное лицо», «Застройщик», «Риелтор»	Указывает, кем является автор объявления
Тип операции	«Сдам», «Продам»	Вид операции с квартирой
Заголовок	Текст на русском языке в свободном изложении	Содержит информацию о количестве комнат в квартире, общей площади и этаже расположения
Описание	Текст на русском языке в свободном изложении	Подробное описание квартиры из объявления, написанное автором объявления
Цена	Числа	Сумма сделки (в рублях)
Регион	Названия регионов РФ	Регион, в котором расположена квартира
Адрес	Текст на русском языке	Адрес квартиры с указанием города, страны, ближайшей станции метро, названия улицы и номера дома
Дата объявления	Дата в формате ДД.ММ.ГГГГ	Дата публикации объявления на сайте
Широта	Число с плавающей точкой	Географические оординаты квартиры
Долгота		
Количество комнат	Целое число либо «студия»	Число комнат в квартире
Общая площадь	Число с плавающей точкой	Общая площадь квартиры в м ²
Жилая площадь	Число с плавающей точкой	Жилая площадь квартиры в м ²
Этаж	Целое число	Номер этажа, на котором расположена квартира
Этажей	Целое число	Количество этажей в доме
Техника	Текст на русском языке	Перечисление имеющейся техники через запятую
Интернет и ТВ	Текст на русском языке	Перечисление имеющихся опций интернета и телевидения через запятую
Парковка	Текст на русском языке	Перечислены имеющиеся опции парковки машины
Балкон	«Балкон», «Лоджия»	Если балконов несколько – перечисляются через запятую
Санузел	«Совмещенный», «Раздельный»	Тип санузла в квартире
Тип дома	«Блочный», «Кирпичный», «Монолитный» и т.п.	Материал стен дома
Лифты	Целое число	Количество лифтов
Год	Год ГГГГ	Год постройки дома

В контексте исследования поля характеристик описываемого объекта не могли быть использованы в качестве признаков классификации, т.к. не несли информации, относящейся к автору. Так, в объявлениях собственника и агента может быть описан один и тот же объект, как следствие, значения характеристик будут идентичны.

Однако была выдвинута гипотеза, что автор может проявить себя в поле «Описание», где информация об описываемом объекте описана автором объявления в свободном формате на естественном языке. Предпосылкой данной гипотезы послужило общее структурное и стилистическое подобие описаний, публикуемых агентами.

Таким образом, целевым параметром модели было выбран поле «Тип автора», а функции входных параметров выполняло содержимое поля «Описание».

Решаемая задача классификации, как результат, была определена как принадлежащая к задачам классификации текстов [18]. Для построения модели классификации при этом тексты объявлений из поля «Описание» необходимо было преобразовать к виду вектора признаков любой природы с обязательным сохранением семантической и стилистической информации.

В. Подготовка данных для обучения модели

Для обучения модели необходимо было предварительно осуществить нормирование и преобразование данных.

Так, значения целевого параметра были нормированы до двоичного формата, где 0 – не-агент («Арендодатель»/«Частное лицо»), 1 – агент («Агентство»/«Застройщик»/«Риелтор»). Итого, в наборе данных было представлено 3940 записей объявлений от агентов и 2520 объявлений от обычных

пользователей.

Ключевым для дальнейшего построения модели обучения стал выбор способа преобразование текстов из поля «Описание», представленных на естественном (русском) языке, в векторы признаков с сохранением семантической и стилистической информации, в них содержащейся.

Предварительно тексты необходимо было нормализовать (привести к нормальной форме) и очистить от слов, знаков, символов, необходимых для нормального восприятия текста, но не имеющих собственной смысловой нагрузки и обычно игнорируемых при осуществлении ранжирования или индексации материалов. К таким стоп-словам были отнесены предлоги, суффиксы, причастия, междометия, цифры, частицы и т. п. Для решения задачи с помощью существующих морфологических пакетов была установлена часть речи каждого отдельного слова, удалены стоп-слова, а значимые слова были приведены к нормальной форме. В результате были получены цепочки только значимых с точки зрения смысловой нагрузки слов, представленных в нормальной форме.

Для решения задачи преобразования цепочек слов в векторы признаков с сохранением смысловой нагрузки был использован подход на основе векторного представления слов (англ. Word Embedding) [19], который ставит в соответствие каждому слову из словаря вектор чисел. Использованная для достижения целей исследования модель Word2Vec [20], обученная на корпусе нормализованных текстов из исследуемого набора объявлений, а также корпусе нормализованных текстов классической русской литературы, позволила за счет статистических свойств обучающих корпусов текстов сохранить семантические свойства слов при преобразовании их в векторы одинаковой размерности. В качестве итогового вектора признаков для обучения модели классификации было решено использовать средний вектор по всем значимым словам каждого объявления.

Структура итогового набора данных для обучения модели классификации представлена в Таблице 2.

В обучающую выборку попали 3142 объявления от агентов и 2023 объявления от обычных пользователей. В тестовую – 798 и 494 соответственно.

Таблица 2 – Описание структуры обучающих данных

Название поля	Возможные значения	Описание
Тип автора	0 – не-агент 1 – агент	Указывает на принадлежность объявления конкретному классу, целевой параметр
Описание	Вектор	Средний вектор значимых слов текстового описания

IV. РЕЗУЛЬТАТЫ

С помощью сформированной обучающей выборки были обучены и протестированы несколько алгоритмов ансамблевого машинного обучения с целью определить модель с наилучшими метриками качества классификации.

Перечень протестированных алгоритмов и результаты эксперимента представлены в Таблице 3.

Как видно из результатов эксперимента (рис. 1), лучшими показателями обладает модель Random Forest (accuracy = 95%). Модели GBC и KNN также показали

достойную точность (88%), однако в дальнейшем было решено сосредоточиться на модели Random Forest.

После оптимизации гиперпараметров модели Random Forest, в частности, числа деревьев решения, точность модели удалось поднять до 94%. Также важными метриками являются precision (показывает долю истинно-положительных выводов классификатора) и полнота recall (показывает, какую долю положительных меток классификатор распознал корректно). Для модели Random Forest после оптимизации гиперпараметров эти метрики равны 96 % и 0,95 соответственно.

Таблица 3 – Результаты эксперимента с моделями классификации

Алгоритм	Accuracy	Precision	Recall	F1-score	ROC AUC
Random Forest (RF)	0.91	0.92	0.92	0.92	0.91
Stochastic Gradient Descent (SGD)	0.73	0.70	0.97	0.82	0.66
Logistic Regression (LR)	0.82	0.85	0.87	0.86	0.81
Gaussian Naive Bayes (GNB)	0.70	0.73	0.80	0.77	0.66
Gradient Boosting Classifier (GBC)	0.88	0.9	0.91	0.91	0.87
K Nearest Neighbours (KNN)	0.88	0.89	0.91	0.90	0.87

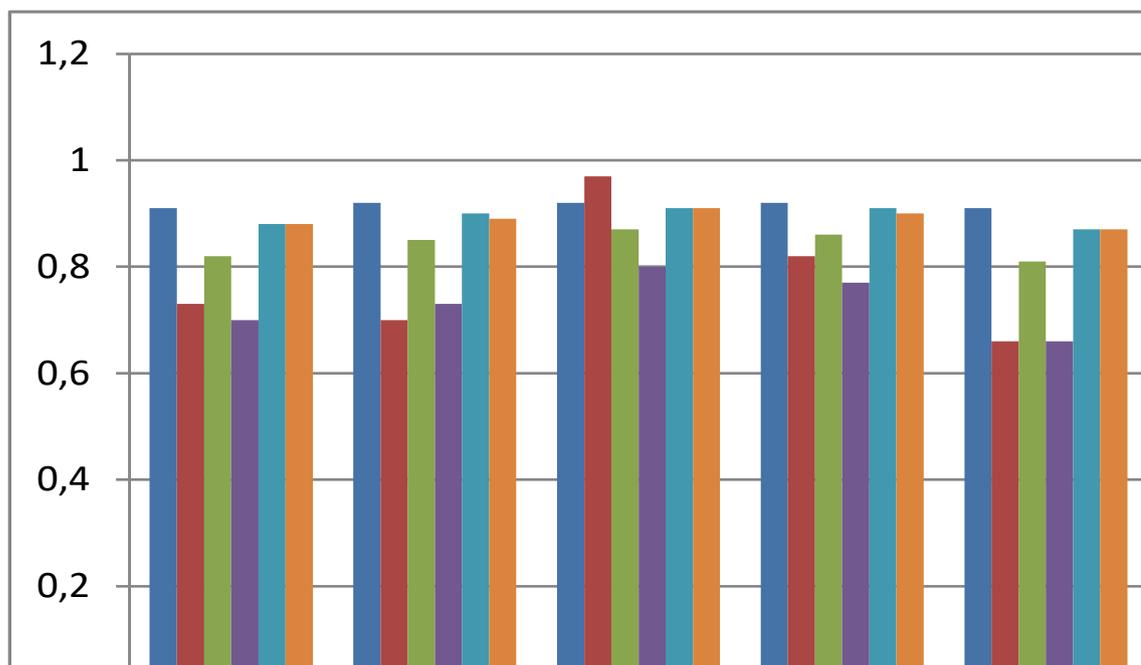


Рис. 1. Сравнение обученных моделей классификации объявлений по метрикам

Таким образом, метод ансамблевого машинного обучения Random forest превзошел другие алгоритмы в части решения задачи распознавания авторства объявлений об операциях с квартирами. Полученные после оптимизации гиперпараметров модели показатели метрик говорят о высоком качестве данных и их разметки в обучающем наборе данных.

Также полученные метрики говорят о том, что обученная и оптимизированная модель корректно распознает объявления, опубликованные агентами, в 96% случаев. Доля ложноположительных срабатываний классификатора составляет 4%, т.е. среди общего количества объявлений, отнесенных моделью к классу агентских, только 4% отнесены к данному классу ошибочно. Из всех агентских объявлений в наборе данных модель корректно распознает 95%. Вышеперечисленные показатели говорят о высокой точности распознавания, которую, тем не менее, целесообразно повысить в будущем.

В дальнейшем предложенный метод валидации данных о недвижимости из открытых источников будет использован для эффективной реализации комплексного развития территорий [21] и в качестве одного из методов обработки данных в интеллектуальной поддержке решений по использованию объектов недвижимости для управления урбанизированными территориями [22].

V. ЗАКЛЮЧЕНИЕ

В настоящем исследовании предложен новый метод определения объявлений об операциях с недвижимостью, опубликованных риелторами и/или риелторскими агентствами (агентами), на основе решения задачи двоичной классификации и семантического анализа текстов на естественном (русском) языке.

Задача была решена на примере заранее размеченных объявлений об операциях с квартирами г. Москва,

собранных из открытых источников. Модель классификации была подобрана с помощью различных алгоритмов классификации по метрикам accuracy, precision, recall, f1-score и ROC AUC. После оптимизации гиперпараметров моделей наилучший результат показала модель Random Forest.

Целесообразно попробовать повысить точность распознавания агентских объявлений с помощью применения дополнительных методов валидации данных, например, перекрестное сравнение данных из разных источников.

Полученную модель можно применять для автоматического прогнозирования принадлежности автора объявления к агентам (риелторам или застройщикам), чтобы затем учитывать возможную недостоверность информации при мониторинге городской среды, частью которой является описываемый в объявлении объект.

БЛАГОДАРНОСТИ

Исследование выполнено за счет гранта Российского научного фонда и Администрации Волгоградской области № 22-11-20024, <https://rscf.ru/project/22-11-20024/>. Автор выражает благодарность коллегам по кафедре "Цифровые технологии в урбанистике, архитектуре и строительстве" ИАиС ВолгГТУ, принимавшим участие в разработке проекта.

Обработка данных, обучение и тестирование моделей производились на "Вычислительном комплексе высокой производительности" внедренном на кафедре "Цифровые технологии в урбанистике, архитектуре и строительстве" ВолгГТУ в рамках реализации программы стратегического академического лидерства "Приоритет 2030".

БИБЛИОГРАФИЯ

- [1] Градостроительный кодекс Российской Федерации от 29.12.2004 N 190-ФЗ (ред. от 02.07.2021) (с изм. и доп., вступ. в силу с 01.10.2021).
- [2] Sustainable Development Goals <https://www.un.org/sustainabledevelopment/ru/sustainable-development-goals/> Retrieved: Dec, 2024
- [3] Рейтинг качества жизни https://asi.ru/government_officials/quality-of-life-ranking/ Retrieved: Dec, 2024
- [4] Индекс качества городской среды — инструмент для оценки качества материальной городской среды и условий её формирования <https://xn----dtbcccotsypabxk.xn--plai/#/> Retrieved: Dec, 2024
- [5] Методика формирования индекса качества городской среды <https://docs.cntd.ru/document/553937399?marker=6560IO/> Retrieved: Dec, 2024
- [6] Housing prices dataset <https://ru-brightdata.com/products/datasets/real-estate/housing-prices/> Retrieved: Dec, 2024
- [7] Журавлев Ю.И., Рязанов В.В., Сенько О.В. "Распознавание". Математические методы. Программная система. Практические применения. Москва: ФАЗИС, 2006 (М.: Типография "Наука" РАН). 176 с.
- [8] Горенбургов М.А., Гончаров В.В. Выработка правил и форм изложения бизнес-информации в сети интернет как средство противодействия мошенническим схемам // Региональная информатика (РИ-2020): XVII Санкт-Петербургская международная конференция. Материалы конференции. 2020. Т. 1. С. 296–297.
- [9] Епрынцева Н.А., Соколова А.В., Руднева А.А. Искусственный интеллект в сфере недвижимости // Информационные технологии в строительных, социальных и экономических системах. 2018. № 4(14). С. 47–50.
- [10] Ou Ts.Y., Lin G.Yu., Fu H.P. et al. An Intelligent Recommendation System for Real Estate Commodity. *Computer Systems Science and Engineering*, 2022, vol. 42, no. 3, pp. 881–897. doi: 10.32604/csse.2022.022637.
- [11] Sinyak N.G., Tajinder S., Madhu K.Ja., Kozlovskiy V.V. Predicting real estate market trends and value using pre-processing and sentiment text mining analysis. *Real Estate: Economics, Management*, 2021, no. 1, pp. 35–43.
- [12] Golubev A., Zelenskiy I., Parygin D. et al. Validation of real estate ads based on the identification of identical images. *Proceedings of the 2018 International Conference on System Modeling and Advancement in Research Trends, SMART 2018: 7*, 2018, pp. 274–279. doi: 10.1109/SYSMART.2018.8746926.
- [13] Vedova M.L.D., Tacchini E., Moret S. et al. Automatic Online Fake News Detection Combining Content and Social Signals. *Conference of Open Innovations Association, FRUCT*, 2018, no. 22, pp. 272–279.
- [14] Prachi N.N., Habibullah Md., Rafi Md.E.H. et al. Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms. *Journal of Advances in Information Technology*, 2022, vol. 13, no. 6. doi: 10.12720/jait.13.6.652-661.
- [15] Mohapatra A., Thota N., Prakasam P. Fake news detection and classification using hybrid BiLSTM and self-attention model. *Multimedia Tools and Applications*, 2022, vol. 81, no. 13, pp. 18503–18519. doi: 10.1007/s11042-022-12764-9.
- [16] Anderson J. R. *Machine Learning: An Artificial Intelligence Approach*. Elsevier Science, 1983. 572 p.
- [17] <https://ru-brightdata.com/products/datasets/real-estate/housing-prices>
- [18] Gasparetto A., Marcuzzo M., Zangari A., Albarelli A. A Survey on Text Classification Algorithms: From Text to Predictions. *Information (Switzerland)*, 2022, vol. 13, no. 2, p. 83. doi: 10.3390/info13020083.
- [19] Wadud Md.A.H., Mridha M.F., Rahman M.M. Word Embedding Methods for Word Representation in Deep Learning for Natural Language Processing. *Iraqi Journal of Science*, 2022, pp. 1349–1361. doi: 10.24996/ij.s.2022.63.3.37
- [20] Rakshit, P. A supervised deep learning-based sentiment analysis by the implementation of Word2Vec and GloVe Embedding techniques / P. Rakshit, A. Sarkar // *Multimedia Tools and Applications*. – 2024. – DOI 10.1007/s11042-024-19045-7. – EDN GVJSKV.
- [21] Zelenskiy, I., Parygin, D., Savina, O., Finogeev, A., Gurtyakov, A. (2022) 'Effective Implementation of Integrated Area Development Based on Consumer Attractiveness Assessment', *Sustainability*, 14(23), art. no. 16239. doi: 10.3390/su142316239. available at: <https://www.mdpi.com/2071-1050/14/23/16239/pdf> (accessed November 06, 2024).
- [22] Zelensky, I.S., Parygin, D.S., Savina, O.V., Finogeev, A.A., Shuklin, A.A., Antyufeev, A.Yu. (2020) 'Intelligent decision support on real estate objects use for urbanized territories management', *Proc. International Journal of Open Information Technologies*, 8(11), pp. 13–29.

Recognizing real estate agent ads using machine learning methods

Ilya Zelenskiy

Abstract— A solution of housing stock state open data validation problem for further using the verified data in the urban environment monitoring process is described. The validation problem in this study is the automatic recognition of potentially unreliable ads published by real estate agents. Automation is complicated by the lack of formal criteria for ads affiliation, the existing methods dependence on additional resources and information sources. Existing methods for solving this and similar problems of recognizing unreliable information are considered using fake news identifying as an example. Two groups of methods are distinguished: social-based and content-based. The first are poorly applicable to this work's task, while the latter marks data being reliable or unreliable based on natural language analysis and classification methods, which is synonymous with the problem posed in this work. The result is the problem formulation as the binary classification of ads based on machine learning methods task. A pre-labeled dataset of Moscow apartments ads from open sources was used to train and test the classification models. Six classification models were compared by target metrics: accuracy, precision, recall, F1 and ROC-AUC metrics. The best model by metrics was selected and its hyperparameters were optimized.

Keywords—learning, artificial intelligence, data analysis, classification, real estate, agents recognition, data validation.

REFERENCES

- [1] Gradostroitel'ny`j kodeks Rossijskoj Federacii ot 29.12.2004 N 190-FZ (red. ot 02.07.2021) (s izm. i dop., vstup. v silu s 01.10.2021).
- [2] Sustainable Development Goals <https://www.un.org/sustainabledevelopment/ru/sustainable-development-goals/> Retrieved: Dec, 2024
- [3] Rejting kachestva zhizni https://asi.ru/government_officials/quality-of-life-ranking/ Retrieved: Dec, 2024
- [4] Indeks kachestva gorodskoj sredy` — instrument dlya ocenki kachestva material'noj gorodskoj sredy` i uslovij eyo formirovaniya <https://xn---dtbccdtsypabxk.xn--p1ai/#/> Retrieved: Dec, 2024
- [5] Metodika formirovaniya indeksa kachestva gorodskoj sredy <https://docs.cntd.ru/document/553937399?marker=6560IO/> Retrieved: Dec, 2024
- [6] Housing prices dataset <https://ru-brightdata.com/products/datasets/real-estate/housing-prices/> Retrieved: Dec, 2024
- [7] Zhuravlev Yu.I., Ryazanov V.V., Sen'ko O.V. "Raspoznavanie". Matematicheskie metody`. Programmnyaya sistema. Prakticheskie primeneniya. Moskva: FAZIS, 2006 (M.: Tipografiya "Nauka" RAN). 176 s.
- [8] Gorenburgov M.A., Goncharov V.V. Vy`rabortka pravil i form izlozheniya biznes-informacii v seti internet kak sredstvo protivodejstviya moshennicheskim sxemam // Regional'naya informatika (RI-2020): XVII Sankt-Peterburgskaya mezhdunarodnaya konferenciya. Materialy` konferencii. 2020. T. 1. S. 296–297.
- [9] Epry`nceva N.A., Sokolova A.V., Rudneva A.A. Iskusstvenny`j intellekt v sfere nedvizhimosti // Informacionny`e texnologii v stroitel'ny`x, social'ny`x i e`konomicheskix sistemax. 2018. № 4(14). S. 47–50.
- [10] Ou Ts.Y., Lin G.Yu., Fu H.P. et al. An Intelligent Recommendation System for Real Estate Commodity. Computer Systems Science and Engineering, 2022, vol. 42, no. 3, pp. 881–897. doi: 10.32604/csse.2022.022637.
- [11] Sinyak N.G., Tajinder S., Madhu K.Ja., Kozlovskiy V.V. Predicting real estate market trends and value using pre-processing and sentiment text mining analysis. Real Estate: Economics, Management, 2021, no. 1, pp. 35–43.
- [12] Golubev A., Zelenskiy I., Parygin D. et al. Validation of real estate ads based on the identification of identical images. Proceedings of the 2018 International Conference on System Modeling and Advancement in Research Trends, SMART 2018: 7, 2018, pp. 274–279. doi: 10.1109/SYSMART.2018.8746926.
- [13] Vedova M.L.D., Tacchini E., Moret S. et al. Automatic Online Fake News Detection Combining Content and Social Signals. Conference of Open Innovations Association, FRUCT, 2018, no. 22, pp. 272–279.
- [14] Prachi N.N., Habibullah Md., Rafi Md.E.H. et al. Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms Journal of Advances in Information Technology, 2022, vol. 13, no. 6. doi: 10.12720/jait.13.6.652-661.
- [15] Mohapatra A., Thota N., Prakasam P. Fake news detection and classification using hybrid BiLSTM and self-attention model. Multimedia Tools and Applications, 2022, vol. 81, no. 13, pp. 18503–18519. doi: 10.1007/s11042-022-12764-9.
- [16] Anderson J. R. Machine Learning: An Artificial Intelligence Approach. Elsevier Science, 1983. 572 p.
- [17] <https://ru-brightdata.com/products/datasets/real-estate/housing-prices>
- [18] Gasparetto A., Marcuzzo M., Zangari A., Albarelli A. A Survey on Text Classification Algorithms: From Text to Predictions. Information (Switzerland), 2022, vol. 13, no. 2, p. 83. doi: 10.3390/info13020083.
- [19] Wadud Md.A.H., Mridha M.F., Rahman M.M. Word Embedding Methods for Word Representation in Deep Learning for Natural Language Processing. Iraqi Journal of Science, 2022, pp. 1349–1361. doi: 10.24996/ij.s.2022.63.3.37
- [20] Rakshit, P. A supervised deep learning-based sentiment analysis by the implementation of Word2Vec and GloVe Embedding techniques / P. Rakshit, A. Sarkar // Multimedia Tools and Applications. – 2024. – DOI 10.1007/s11042-024-19045-7. – EDN GVJSKV.
- [21] Zelenskiy, I., Parygin, D., Savina, O., Finogeev, A., Gurtyakov, A. (2022) 'Effective Implementation of Integrated Area Development Based on Consumer Attractiveness Assessment', Sustainability, 14(23), art. no. 16239. doi: 10.3390/su142316239. available at: <https://www.mdpi.com/2071-1050/14/23/16239/pdf> (accessed November 06, 2024).
- [22] Zelenskiy, I.S., Parygin, D.S., Savina, O.V., Finogeev, A.A., Shuklin, A.A., Antyufeev, A.Yu. (2020) 'Intelligent decision support on real estate objects use for urbanized territories management', Proc. International Journal of Open Information Technologies, 8(11), pp. 13–29.