

## **ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРЕДСКАЗАНИЯ ЦЕН НА НЕДВИЖИМОСТЬ В МОСКВЕ И МОСКОВСКОЙ ОБЛАСТИ**

*Каинов Егор Андреевич, студент 4 курса института Экономики и управления АПК, ФГБОУ ВО РГАУ–МСХА имени К.А. Тимирязева  
E-mail: egor.kainov@bk.ru*

*Научные руководители – Худякова Елена Викторовна, д.э.н., профессор, профессор кафедры прикладной информатики, ФГБОУ ВО РГАУ – МСХА имени К.А. Тимирязева, evhudyakova@rgau-msha.ru.*

*Лосев Алексей Николаевич, старший преподаватель кафедры прикладной информатики, ФГБОУ ВО РГАУ – МСХА имени К.А. Тимирязева, losev@rgau-msha.ru.*

**Аннотация:** В данной научной работе представлен проект, посвященный предсказанию цен на квартиры с использованием методов машинного обучения. Целью проекта было разработать модель, способную предсказывать стоимость жилья на основе различных характеристик квартир. Для достижения этой цели были использованы открытые данные о рынке недвижимости и широкий спектр алгоритмов машинного обучения, а также реализация телеграмм бота для упрощения взаимодействия пользователя с моделью.

**Ключевые слова:** машинное обучение, модели, предсказание, регрессия, анализ данных

Исследование по предсказанию цен на квартиры в Москве и Московской области с использованием методов машинного обучения представляет собой актуальную задачу по следующим нескольким причинам:

1. Рынок недвижимости в Москве и Московской области представляет собой сложную систему с большим объемом данных и множеством влияющих на цену факторов. Обработка всех этих данных вручную является трудоемкой и неэффективной по времени задачей.

2. Применение методов машинного обучения позволяет эффективно анализировать большие объемы данных и выявлять сложные закономерности и зависимости между различными объясняющими признаками и ценой, которая является целевой переменной

3. Данный проект имеет практическую значимость для конечного пользователя, поскольку предоставляет возможность быстро и удобно оценивать стоимость жилья по заданным признакам (характеристикам квартиры). Так же это может быть особенно важно для агентств недвижимости и потенциальных покупателей.

Для реализации проекта были взяты данные из открытых источников, которые предоставляют информацию о рынке недвижимости в Москве и Московской области. Эти данные включают в себя различные характеристики квартир и цен на них в определенных районах.

Набор данных включает в себя следующие признаки:

1. Цена на квартиру (руб.);
2. Тип квартиры (вторичное жилье или новостройка);
3. Станция ближайшего метро;
4. Минут до станции метро;
5. Регион (Москва или Московская область);
6. Адрес квартиры;
7. Количество комнат;
8. Площадь квартиры;
9. Площадь кухни;
10. Этаж, на котором расположена квартира;
11. Количество этажей в доме;
12. Ремонт (Косметический, евроремонт, без ремонта или дизайнерский);

Всего в фрагменте используется 12 столбцов и 22 676 объявлений о продаже квартир.

После получения данных из открытых источников, была проведена работы по обработке и подготовке набора данных для последующего анализа. В этот этап входили такие шаги, как очистка от ошибочных данных, структурирование данных, для удобства работы и анализа, преобразование на соответствующие типы данных и заполнение пропусков.

После обработки был проведен анализ данных [5] и выявлены следующие взаимосвязи из данных:

1. Данные предоставлены за ноябрь и декабрь 2023 года;
2. Двухкомнатные квартиры на рынке недвижимости преобладают;
3. Большинство квартир расположены в 13 минутах ходьбы от метро;
4. Квартир на нижних этажах больше (в основном квартиры расположены со 2 по 10 этаж);
5. Средняя цена на квартиры с косметическим ремонтом меньше остальных категорий ремонта;
6. Самая дорогая квартира стоит 2 455 020 000 рублей;
7. Квартиры, которые расположены от 5 минут до метро до 30 минут стоят дешевле чем те, которые расположены в 40-60 минутах;

Далее были проведены эксперименты с применением различных моделей машинного обучения для построения модели, которая предсказывает цены на квартиры в Москве и Московской области. В ходе экспериментов было реализовано более 10 различных подходов к моделированию алгоритмов.

Среди примененных моделей в экспериментах были – линейная регрессия, случайный лес и CatBoost. Каждая модель была настроена, а также было оценено качество прогнозов с использованием таких метрик, как средняя квадратическая ошибка, средняя абсолютная ошибка и среднее абсолютное процентное отклонение.

В ходе сравнения различных алгоритмов – лучшее качество показала модель от Яндекса – CatBoost – градиентный бустинг на решающих деревьях.

Данная модель показала следующие метрики:

1. RMSE (корень из средней квадратической ошибки) = 4 193 612
2. MAE (средняя абсолютная ошибка) = 1 975 044
3. MAPE (средняя процентная абсолютная ошибка) = 0.23

То есть модель в среднем ошибается на 0.23 процента, что является хорошим результатом работы модели

В результате исследовательской работы по предсказанию цен на квартиры в Москве и Московской области с применением моделей машинного обучения был разработан и реализован телеграмм-бот.

Цель создания данного бота – обеспечение удобного и эффективного взаимодействия с моделью, предоставляющей информацию о ценах на недвижимость.

Функционал телеграмм бота включает возможность прогноза цена на квартиру в зависимости от ее характеристик. Пользователи могут вводить данные о квартире непосредственно в чат с ботом, после чего модель проводит прогноз цены на основе имеющихся данных и выдает результат. Так же есть функция получения аналитики набора данных – для просмотра и изучения взаимосвязей и признаков, которые влияют на цену.

Фрагмент работы телеграмм бота с моделью машинного обучения представлен на рисунке 1.



Рисунок 1 – Средняя цена квартир в зависимости от ремонта

### Библиографический список

1. pandas - PythonDataAnalysisLibrary [Электронныйресурс]. – 2024. – <https://pandas.pydata.org/docs/>
2. seaborn: statistical data visualization [Электронныйресурс]. – 2024 – <https://seaborn.pydata.org/tutorial.html>
3. Telegram Bot API [Электронный ресурс]. – 2024 – <https://core.telegram.org/bots/api>
4. Cat Boost - open-source gradient boosting library [Электронныйресурс]. – 2024 – <https://catboost.ai/en/docs/>
5. Курс социально-экономической статистики : Учебник для вузов / В. Л. Соколин, М. Р. Ефимова, А. Л. Кевеш [и др.]. – Москва : Финстатинформ, 2002. – 976 с. EDN TDCKD