

# Сервис для оценки стоимости недвижимости

Проект в рамках всероссийской научно-технологической программы по решению проектных задач в области искусственного интеллекта и смежных дисциплин «Сириус.ИИ»

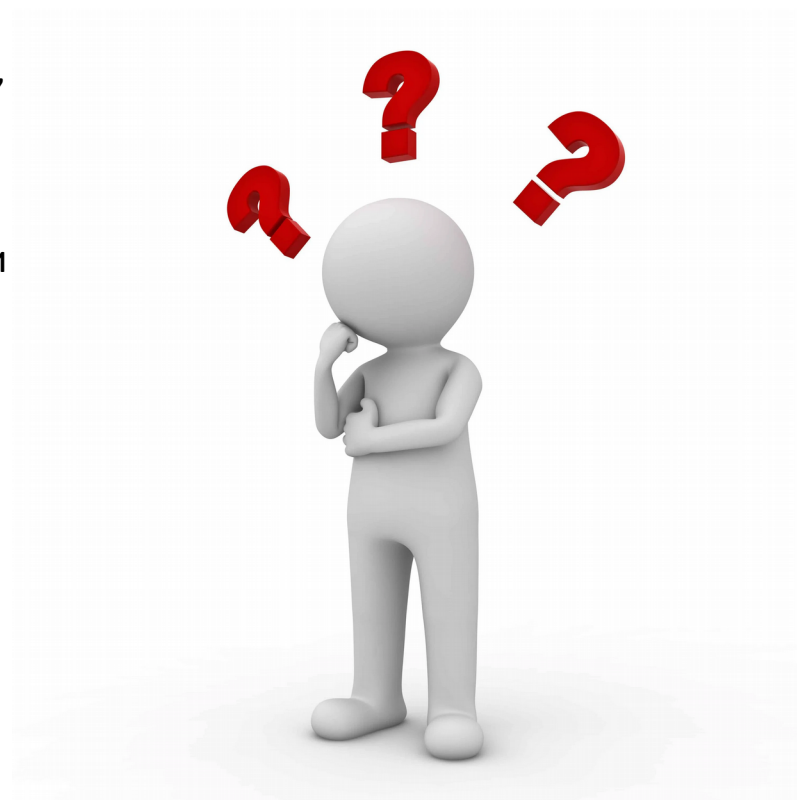
# Состав команды

- Капитан: Хорошилова Анастасия Евгеньевна(10 класс, МБОУ «Гимназия 42», город Барнаул, Алтайский край)
- Участники: Речкина Виктория Сергеевна(10 класс, КГБОУ Алтайский краевой педагогический лицей-интернат, город Барнаул, Алтайский край); Гостеев Прохор Игоревич(10 класс, КГБОУ Алтайский краевой педагогический лицей-интернат, город Барнаул, Алтайский край); Ефимов Игорь Андреевич(10 класс, КГБОУ Алтайский краевой педагогический лицей-интернат, город Барнаул, Алтайский край)

# Формулировка проблемы

Существует огромное количество проблем, связанных со стоимостью недвижимости: сложности с получением ипотеки, мошенники обманывают покупателей квартир, продавец не может самостоятельно оценить стоимость квартиры в связи с изменениями на рынке.

Но все эти проблемы решает онлайн сервис оценки стоимости недвижимости, ведь он доступен каждому, является бесплатным и с высоким качеством анализирует цены на квартиры, с помощью машинного обучения



# Кому будет полезен сервис для оценки стоимости недвижимости?

Продавцам(Вы сравните предложение с похожими в одной ценовой категории и том же доме, поймёте, какая стоимость ускорит сделку и повысит на неё спрос или же можете быть спокойны, что не продешевили)

Покупателям(Рыночная цена станет аргументом для торга и поможет не переплатить. Выбор увеличится: вы увидите похожие предложения по той же цене)

Риелторам(Поможет быстро собрать информацию о предложениях, похожих на квартиру клиента, и договориться о стоимости продажи)

Банкам(Важно понимать, на какую собственность он выдает деньги и насколько быстро ее можно продать, если заемщик перестанет погашать задолженность. Если заемщик не в состоянии выплачивать взятую ипотеку, кредитор продает жилье по рыночной или ликвидационной (бывает реже) цене, зафиксированной в документе.)

# Другие случаи, когда приложение будет полезно

- 1 - получение кредита под залог недвижимости;
- 2 - обмен квартирами;
- 3 - получение жилья в наследство;
- 4 - страхование и приватизация объекта;
- 5 - процедура ликвидации, реорганизации или банкротства предприятия;
- 6 - обращение в суд по имущественным спорам;
- 7 - оценка ущерба при пожаре и стихийных бедствиях;
- 8 - претензии со стороны налоговой службы;
- 9 - вклад недвижимости в уставный капитал в качестве доли одного из партнеров;
- 10 - привлечение в бизнес инвестиций.



# Цель проекта

- Целью нашей работы станет создание приложения, способного предсказывать стоимость недвижимости, основываясь на её особенностях и находить варианты недвижимости в данном ценовом диапазоне.
- Оригинальность идеи: Программа будет оценивать стоимость объектов в реальном времени, учитывая экономическую ситуацию. Также цену недвижимости можно будет посмотреть сразу в нескольких валютах. В случае, если квартира не подошла пользователю, он может посмотреть похожие варианты.

# Анализ области. Компании, занимающиеся оценкой стоимости недвижимости

Компания	Плюсы	Минусы
1) Дом.рф — финансовый институт развития в жилищной сфере России. Компания развивает жилищное строительство, разрабатывает субсидированные ипотечные программы, выплачивает региональные субсидии.	<p>Оценка стоимости недвижимости - это одна из полностью независимых, предоставляемых услуг, напрямую связанная с их основной деятельностью.</p> <p>Сервис работает для очень большой территории (вся Россия и даже за её пределами)</p>	<p>Не предлагает оценку стоимости для конкретной квартиры с учётом её особенностей (например, качественного ремонта)</p> <p>Совсем не предлагает конкретных вариантов для совершения сделки</p>

# Анализ области. Компании, занимающиеся оценкой стоимости недвижимости

Компания	Плюсы	Минусы
<p>2) Циан</p> <p>Основная функция Циан - предложение посетителям самую качественную базу объявлений. Циан помечает объявления, цены в которых попадают в привлекательный диапазон оценки объекта, специальным значком «Хорошая цена». Поэтому у Циан оценка стоимости недвижимости практически основная задача, но всё же второстепенная.</p>	<p>Сервис оценки квартиры «ЦИАН» учитывает не только расположение и метраж жилья, но и тип и год постройки дома, этажность, высоту потолков в квартире, вид ремонта, наличие балкона и даже куда выходят окна.</p>	<p>Узнавать серию дома, год его постройки и другие детали нужно на отдельных сервисах, что иожет занять время.</p>



# Анализ области. Компании, занимающиеся оценкой стоимости недвижимости

Компания	Плюсы	Минусы
3) Яндекс Недвижимость	<p>На сайте можно посмотреть, какие объявления в том же доме публиковали за последние несколько лет</p> <p>Есть подборка текущих объявлений о продаже похожих квартир в том же районе</p>	<p>«Яндекс Недвижимость» не покажет примерную стоимость квартиры в моменте</p> <p>Конкретной оценки, хорошая цена на объект или плохая нет.</p>

# Анализ области. Компании, занимающиеся оценкой стоимости недвижимости

Компания	Плюсы	Минусы
<p>4) Авито</p> <p>Авито - сервис для продажи различных объектов. Недвижимость не является основной концепцией компании, а следовательно и оценка стоимости тоже.</p>	<p>Позволяет оценить стоимость недвижимости за счёт сотен объявлений, когда-либо опубликованных в сервисе</p> <p>При введении параметров, пользователю предоставляют варианты недвижимости, соответствующие указанным критериям, а также цены, ниже и выше которых объект считается "недооценён" и "переоценён" соответственно.</p>	<p>«Авито» учитывает не так много данных о квартире</p> <p>Из-за маленького количества введённых критериев на объект, недвижимость имеет большой разброс характеристик.</p>

# Анализ области

- При анализе области оценки стоимости недвижимости наша команда определила подход к решению поставленной задачи. Созданная нами программа предоставит пользователям как возможности, которые есть на других платформах и сервисах, так и новые, которые ранее не были представлены.
-

# Описание предлагаемого решения. В каком виде сервис будет доступен пользователю.

Сервис представляет собой мобильное приложение на Android. На первом экране будет две кнопки, дающие возможность запустить программу в разных режимах: анализ стоимости жилья на основе его характеристик, анализ возможных особенностей квартиры на основе допустимого бюджета.

При переходе в **первый режим** открывается новое окно, где пользователю предлагается указать особенности квартиры. После чего он нажимает на кнопку «Проанализировать стоимость покупки» или «Проанализировать стоимость аренды» и на экран выводится приблизительная цена.

При переходе во **второй режим** пользователь вводит в текстовые поля цену, которую он готов потратить на недвижимость и населённый пункт, где она будет находиться, а также выбирает одну из кнопок «Проанализировать стоимость покупки» или «Проанализировать стоимость аренды». Открывается новое окно, где предложены различные варианты квартир.

В обоих режимах будет кнопка «Показать похожие варианты», где пользователь сможет посмотреть другую недвижимость с похожей ценой или характеристиками.

# Критерии, которые необходимо ввести в первом режиме

Особенности квартиры:

частный дом/квартира  
площадь(в квадратных метрах),  
количество комнат,  
этаж,  
наличие/отсутствие балкона,  
раздельный или совмещенный санузел,  
сделан ли в квартире ремонт,  
вторичное это жилье или новое.

Параметры дома:

тип(панельный, кирпичный и тд),  
год постройки,  
количество этажей  
количество лифтов,  
удаленность от центра(в км или минутах езды)  
наличие парковки у дома



# Характеристики, которые будут указаны для недвижимости во втором режиме

площадь(в квадратных метрах)

количество комнат

наличие ремонта

удаленность от центра

новое/вторичное жилье



# Отображение стоимости в разных валютах

- Также в приложении можно будет выбрать валюту, в которой указана стоимость жилья(рубли, доллары, евро), что очень удобно, особенно для инвесторов или людей, хранящих капитал в различной валюте.



# Высокоуровневый план решения задачи

Основные шаги при создании сервиса:

## 1) Подготовка данных

Предполагается собрать данные о квартирах из открытых источников с достаточно большим количеством характеристик (цена, географические данные, площадь, количество комнат, этаж, данные о ремонте, тип комнат, тип санузла, год постройки и т.д.) для использования в модели линейной регрессии.

## 2) Обучение модели

Использование Линейной Регрессии с помощью библиотеки `scikit-learn` в нашем случае имеет свои преимущества, которые делают ее предпочтительным выбором.

## 3) Создание приложения

Приложение на Android с удобным и простым интерфейсом — идеальный вариант взаимодействия пользователя и программы. Реализация планируется с помощью языка программирования Kotlin и сервиса Android studio.

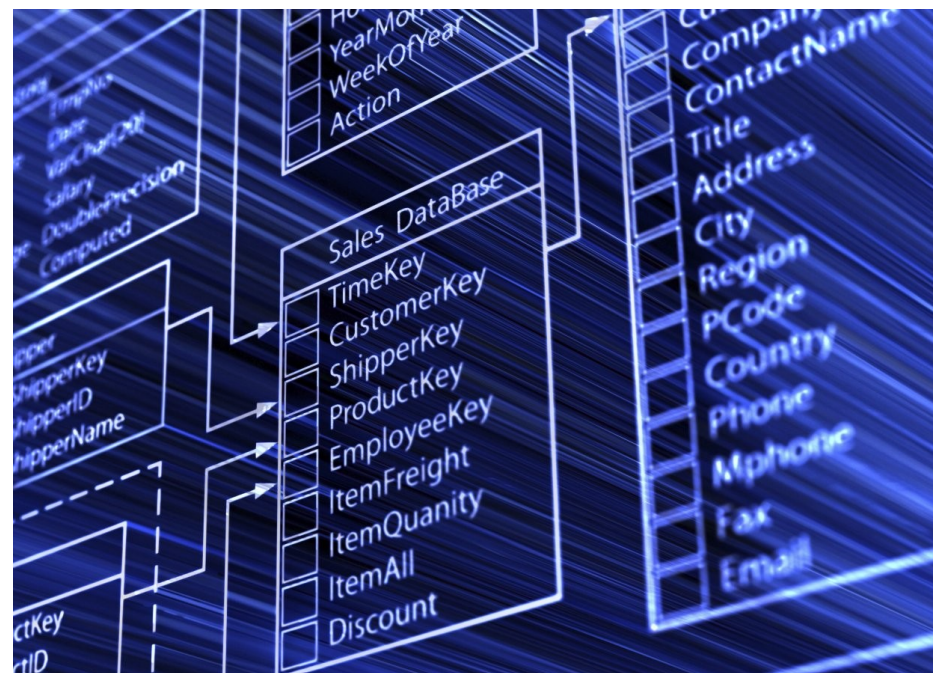


# Сбор данных о недвижимости

Предполагается собрать данные о квартирах из открытых источников (сервис авито) с достаточно большим количеством характеристик (цена, географические данные, площадь, количество комнат, этаж, данные о ремонте, тип комнат, тип санузла, год постройки и т.д.).

Данный этап включает в себя:

- 1) Парсинг данных о квартирах с использованием библиотеки BeautifulSoup4 и Windscribe VPN (VPN приходится использовать из-за ограниченного количества запросов к сервису авито).
- 2) Представление данных о квартирах в виде таблицы pandas.
- 3) Предварительная обработка данных: заполнение пустых значений, удаление ошибочных данных, перевод категориальных признаков в числовые (One-hot encoding), возможно какие-то преобразования числовых данных.
- 4) Разбиение выборки на обучающую и тестовую.



# Обучение модели

Использование Линейной Регрессии с помощью библиотеки scikit-learn в нашем случае имеет свои преимущества, которые делают ее предпочтительным выбором. аргументами в пользу использования Линейной Регрессии с scikit-learn стали:

## **Простота реализации и понимания:**

Линейная Регрессия - один из самых простых и понятных алгоритмов машинного обучения, легко интерпретируется.

Библиотека scikit-learn предоставляет простой и интуитивно понятный интерфейс для работы с моделями Линейной Регрессии, что делает процесс разработки и экспериментов более простым и удобным.

## **Быстрая разработка и обучение модели:**

Реализация и обучение модели Линейной Регрессии с помощью scikit-learn требует минимального количества кода и времени.

## **Хорошая производительность на простых данных:**

Линейная Регрессия хорошо подходит для решения задачи прогнозирования цен на недвижимость, так как данные не имеют сложной структуры и зависимостей.

## **Меньшая сложность модели:**

По сравнению с более сложными моделями, такими как нейронные сети, модель Линейной Регрессии имеет меньше параметров, что может снизить риск переобучения и упростить ее интерпретацию.

# Обучение модели. Линейная регрессия

В качестве основных показателей при оценке точности модели будем рассматривать среднюю квадратичную ошибку ( $MSE^1$ ) и коэффициент детерминации ( $r\text{-squared}^2$ ).

При сравнении моделей нам будет предпочтительней та, у которой меньше грубых ошибок. Насколько часто модель совершает грубые ошибки хорошо отражает MSE, так как ошибка прогноза возводится в квадрат. MSE не ограничен сверху, поэтому бывает тяжело понять насколько хорошо его значение удовлетворяет нашим требованиям, поэтому чаще будем рассматривать коэффициент детерминации.

Коэффициент детерминации показывает, какая доля дисперсии целевой переменной объяснена моделью, по сути он является нормированной средней квадратичной ошибкой. Он принимает значения от 0 до 1. Чем ближе данное значение к 1, тем лучше модель объясняет изменчивость целевой переменной. Таким образом, коэффициент детерминации легко интерпретировать и использовать для сравнения различных моделей.

$$^1 MSE = \frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2$$

$$^2 R^2 = 1 - \frac{\sum_{i=1}^n (f(x_i) - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

# Обучение модели. Обработка данных на примере от заказчика.

- Первым шагом стало соединение train.csv и macro.csv в один dataframe по столбцу timestamp.
- Далее были удалены столбцы, у которых больше половины значений отсутствует.
- Если в столбце мало пропущенных значений, то просто удаляются строки с отсутствующими данными.
- Переведены все значения к типу float64 там где это возможно.
- Удалены столбцы timestamp и sub\_area, так как их нельзя представить в численных значениях, а в таблице уже есть данные о состоянии экономики по дате и характеристика о районах.
- Все оставшиеся отсутствующие данные заменены медианным значением соответствующего столбца.
- Таким образом данные будут обработаны наиболее точно
- После чего все данные проходят через линейную регрессию и случайный лес с помощью библиотеки sklearn.

# Подбор признаков

Большое количество признаков сильно усложняет модель, уменьшает скорость обучения, повышает риск переобучения, ухудшает интерпретируемость. Поэтому важно правильно подобрать набор признаков, чтобы улучшить качество модели.

При отборе признаков стоит учитывать:

Корреляцию: признаки с высокой корреляцией с целевой переменной будут более предпочтительны.

Мультиколлинеарность: признаки не должны сильно коррелировать между собой, так как это может привести к снижению точности оценки отдельных коэффициентов. В качестве численного показателя мультиколлинеарности используется VIF.

Значимость признаков: признаки, которые оказывают малое влияние на целевую переменную менее предпочтительны.

При добавлении или удалении признаков нужно сравнить показатели точности с предыдущими, чтобы убедиться, что выбранные признаки действительно улучшают качество модели.

# Создание приложения

- Приложение будет написано на языке Kotlin для операционной системы Android. В программе для разработки ПО «Android studio» можно размещать виджеты на экране вручную, после чего задавать им определенные свойства, что очень удобно в данном случае. В основу приложения лягут такие простые элементы как текстовая метка, кнопка и поле для ввода текста.

# Технические компетенции

- У участников команды есть все необходимые знания и навыки для создания данного приложения. Основные технологии, которые будут использованы при его создании: библиотека pandas в языке программирования Python3.0 для сбора данных, библиотека sklearn в языке программирования Python3.0 для использования машинного обучения при анализе данных, языки программирования Python и Kotlin, программа «Android studio» для создания мобильного приложения.

## Имеющийся результат

- Наша команда успела реализовать значительную часть продукта
- К нынешнему моменту готова система сбора данных о недвижимости с сервиса «Авито» с помощью библиотеки `pandas`
- Также нами был реализован анализ данных о недвижимости с использованием машинного обучения (технологий регрессии и случайного леса). Эта часть работы является основной, так как именно она включает в себя новизну и актуальности нашего приложения.
- В процессе разработки находится мобильное приложение, которое представит результаты оценки стоимости недвижимости в удобном для пользователя виде



# Чем приложение отличается от уже существующих?

- Сервис позволяет анализировать как стоимость аренды, так и стоимость продажи недвижимости
- Можно указать большое количество параметров, которые не учитываются в других сервисах. Следовательно, оценка стоимости будет более точной
- Сразу можно посмотреть похожие варианты и выбрать наиболее выгодный
- Цена указана сразу в нескольких валютах
- Есть как функция оценки цены недвижимости по её характеристикам, так и функция подбора недвижимости, основываясь на цене

Спасибо за внимание!