



# Проект по курсу “ML System Design”

## “Оценивание стоимости жилой недвижимости”

Выполнили студенты М8О-406Б-21: Деревянко  
Екатерина, Волошинская Евгения, Леленков  
Никита



# Актуальность

В последние годы рынок недвижимости становится всё более динамичным и технологически насыщенным. Одной из ключевых проблем, с которой сталкиваются как покупатели, так и продавцы недвижимости, является правильная и объективная оценка стоимости объектов. Традиционные методы оценки, такие как анализ рыночных предложений, консультации с экспертами или использование оценочных компаний, часто оказываются медленными, дорогими и не всегда дают точные результаты.

Технологии искусственного интеллекта позволяют значительно ускорить процесс получения точной информации о стоимости объекта и делает его доступным для широкого круга пользователей, включая частных лиц, агентства недвижимости, а также финансовые и инвестиционные компании.

Система, основанная на нейросети, может учитывать множество факторов, влияющих на цену недвижимости, таких как её расположение, площадь, инфраструктура, состояние, история продаж в регионе и многие другие переменные. Внедрение подобных технологий позволяет повысить прозрачность рынка, снизить риски для участников сделок и улучшить качество обслуживания клиентов.

Таким образом, проект является актуальным и востребованным в современных условиях, поскольку он предлагает инновационное решение для точной и быстрой оценки недвижимости, что способствует улучшению пользовательского опыта и оптимизации сделок на рынке недвижимости.



# Бизнес цель

Создать веб-сайт, который использует нейросеть для предоставления мгновенных и точных оценок стоимости жилой недвижимости. Это позволит пользователям быстро получать оценки без необходимости обращаться к профессиональным оценщикам, что сэкономит время и деньги, а также ускорит сделку. Сайт будет полезен как для частных лиц, желающих продать или купить недвижимость, так и для финансовых учреждений, нуждающихся в надежных данных для кредитования и страхования.



# ML цель

Основной целью системы машинного обучения данного проекта является прогнозирование стоимости недвижимости на основе множества факторов, включая расположение, площадь, состояние, возраст здания, инфраструктуру и другие переменные. Для достижения этой цели необходимо решить несколько ключевых задач:

1. Сбор и обработка данных
2. Обучение модели
3. Обработка новых данных и адаптация модели
4. Составление прогнозов

# Исходные и выходные данные

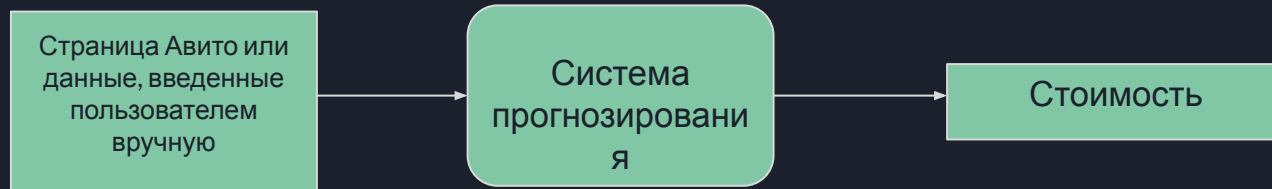
Входными данными являются объявления Авито, из которых парсер выделяет такие данные как фотографии квартиры, площадь, количество комнат, адрес, наличие ремонта и т.п. Можно выделить следующие важные признаки:

Квартира: число комнат, этаж, балкон, ремонт, площадь — общая, жилая, кухни

Расположение: удалённость от центра, расстояние до метро

Дом: год постройки, число этажей, материал, перекрытия, лифты, мусоропровод, газ

На выходе модель выдает одно число - спрогнозированная стоимость.





# Предобработка данных

1. Удаление пропущенных значений
2. Обработка текстов: Токенизация адресов
3. Нормализация числовых данных
4. Обработка мультимедиа: Извлечение признаков из изображений или видео (например, через модели компьютерного зрения).



# Выбор модели

Так как проект подразумевает обработку большого числа объявлений, а также возможную оценку по изображениям, в качестве категории стоит выбрать глубокое обучение (Deep Learning).

В качестве модели мы выбрали нейронную сеть, потому что нейронные сети отлично справляются с текстом и изображениями; категориальные данные (например, адрес или материал дома) можно преобразовать в эмбединги; такие модели, как BERT или ResNet, можно использовать для извлечения признаков из текста и изображений.

Также мы выбрали одну DNN, так как предсказываем только одно число.



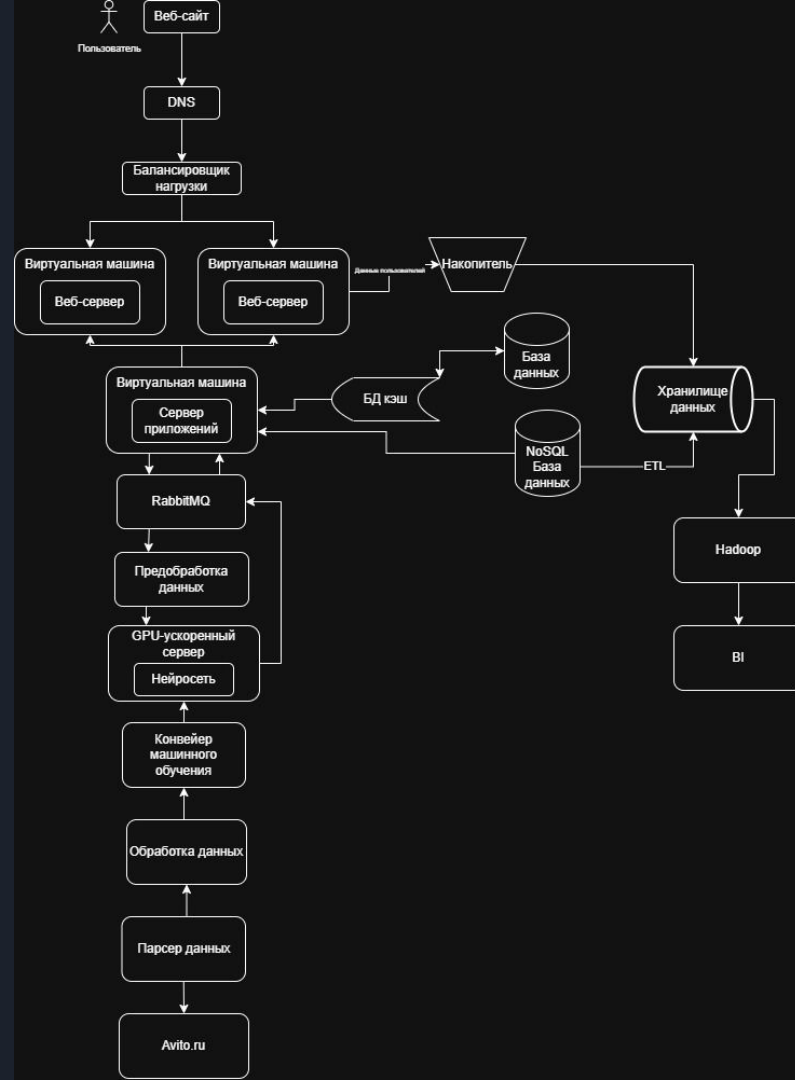
# Эксплуатация. Внедрение

1. Мониторинг производительности:
  - Необходимо отслеживать метрики точности предсказаний (для нашей задачи регрессии - MSE, MAE).
  - Анализируем поведение модели на новых данных.
2. Регулярное обновление модели:
  - Т.к. рынок динамичен, то необходимо периодически дообучать модель на новых данных.
3. Оптимизация вычислений:
  - Используем распределенное обучение на кластере GPU.



# Общий дизайн системы

Подробное описание схемы можно найти в Readme на Github (URL: <https://github.com/TooManyFandoms/ML-System-Design/blob/main/README.md>)





# Список литературы

Сюй А. System Design. Подготовка к сложному интервью. - СПб.: Питер, 2022. - 304 с.



Спасибо за внимание!