**Доклад к ВКР: Веб-приложение для прогнозирования стоимости легкового автомобиля**

Слайд 1. Титульный

Здравствуйте, уважаемые члены комиссии. Меня зовут Дмитрий Колбанов, я обучающийся группы 6401‑020302 Института информатики и кибернетики Самарского университета. Представляю вашему вниманию выпускную квалификационную работу на тему «Веб‑приложение прогнозирования стоимости легкового автомобиля». Руководитель – доцент Гордеева О.А.

**Слайд 2. Цель и задачи**

Цель работы - разработать веб-приложение для прогнозирования стоимости легкового автомобиля с использованием нескольких алгоритмов машинного обучения и провести сравнительный анализ их эффективности.

Задачи:

* Провести анализ предметной области;
* Сделать обзор систем-аналогов в области прогнозирования стоимости автомобилей;
* Найти данные для обучения моделей машинного обучения;
* Провести сравнительный анализ примененных алгоритмов прогнозирования стоимости легкового автомобиля;
* Разработать и реализовать информационное и программное обеспечение;
* Провести тестирование и отладку разработанного веб-приложения.

Слайд 3. Актуальность

 Рынок подержанных автомобилей в России устойчиво растёт: вторичный рынок сейчас более чем в два раза превышает первичный по числу сделок (свыше 3,5 млн против 1,6 млн автомобилей в 2025 г.).

 Цены на новые автомобили растут из‑за политической обстановки, проблем с поставками комплектующих и высокой (оценочно +10 % за год), что повышает интерес пользователей к «б/у» сегменту и делает инструмент оценки цены крайне востребованным.

 Для частных лиц (продавцов и покупателей), дилеров и страховых компаний важно иметь объективную рыночную оценку автомобиля, чтобы минимизировать риски переплаты или недополучения прибыли.

 Всё это создаёт предпосылки для интеграции моделей машинного обучения в онлайн‑сервисы оценки авто.

Слайд 4. Предметная область

Прогнозирование стоимости легкового автомобиля – задача определения рыночной цены транспортного средства на основе его характеристик, т.е. задача регрессии: модель должна предсказывать непрерывную величину на основе входных признаков: марки и модели, год выпуска, пробега и др.

Могут быть различные сценарии использования:

* Частный продавец оценивает стоимость своего авто перед публикацией объявления.
* Компания‑дилер формирует цену при закупке машин.
* Страховые компании автоматически рассчитывают страховую стоимость.

Слайды 5-8. Анализ аналогов

Avito Авто – Сервис для размещения объявлений о продаже автомобилей, включающий функцию автоматической оценки стоимости на основе характеристик транспортного средства. Т.е. как такого инструмента прогнозирования нет, лишь подсказка при создании объявления о продаже или поиске авто. У Авито очень много объявлений по всей стране. Это предоставляет широкую базу для анализа с учетом региональных цен.

Auto.ru – популярный автомобильный портал, предоставляющий инструмент для прогнозирования стоимости автомобиля на основе его параметров и рыночных данных. Здесь уже конкретный инструмент прогнозирования, который, однако, предоставляет платный доступ к расширенному функционалу. Также у Auto.ru меньше объявлений, соответственно, менее объемная база данных.

Слайд 9-10. Источник данных

Для обучения моделей использованы данные, собранные с помощью веб-скрапинга с сайта Avito по Самарской области — почти 25 000 объявлений. Извлечены все основные технические и эксплуатационные характеристики, а также таргет – цена.

Слайды 11-12. Подготовка данных

Следующим важным этапом была подготовка данных. Она включала очистку данных от дубликатов и выбросов (количество данных после всех этапов – на слайде), кодирование категориальных признаков, анализ значимости и отбор признаков, а также разбиение на обучающую и тестовую выборки.

Слайды 13-14. Алгоритмы прогнозирования

Были использованы следующие алгоритмы:   
• Линейная регрессия — простая и интерпретируемая модель, которая пытается уловить линейную зависимость между признаками и целевой переменной, слабая при сложных зависимостях.   
• Метод k-ближайших соседей — очень простой алгоритм, который вычисляет расстояние до всех объектов из обучающей выборки в признаковом пространстве, и берет k ближайших.   
• Случайный лес — является ансамблевым методом, т.е. композицией алгоритмов, а именно бэггингом над решающими деревьями.   
• Градиентный бустинг — также ансамблевый метод, где происходит последовательное обучение моделей, и каждая следующая компенсирует ошибки предыдущих. В реализации CatBoost – неглубокие деревья.

Слайды 15-16. Метрики оценки

Оценка качества проводилась с помощью MAE, R² и MAPE. 

**MAE**– средняя абсолютная ошибка в рублях.

**R² (коэффициент детерминации)** – доля дисперсии цен, объясняемая моделью (1 – идеал, 0 – бесполезная модель).

**MAPE**– средняя абсолютная ошибка в процентах.

Слайды 17. Итоговая модель

По результатам оценки на тестовой выборке наилучший результат по всем метрикам показал градиентный бустинг от CatBoost. Соответственно, при реализации веб-приложения использовался именно он. Характеристики модели представлены на слайде.

Слайды 18. Ошибки системы

Модель была протестирована на различных ценовых сегментах, и выяснилось, что наиболее значительные ошибки системы встречаются у эксклюзивных и премиальных автомобилей, т.к. их было немного в обучающей выборке.

Слайды 19-21. Проект системы

Бэкенд приложения реализован на Flask, фронтенд на Vue. Схемы

Слайды 22-23. Экранные формы

Пользователь вводит параметры автомобиля, и приложение возвращает прогноз цены. Реализованы валидация, обработка ошибок.

Слайд 24. Выводы

В результате работы было разработано веб-приложение для прогнозирования стоимости легкового автомобиля с использованием нескольких алгоритмов машинного обучения. Выполнены следующие задачи:

* Проведен анализ предметной области;
* Сделан обзор систем-аналогов в области прогнозирования стоимости автомобилей;
* Собраны данные для обучения моделей машинного обучения;
* Проведен сравнительный анализ примененных алгоритмов прогнозирования стоимости легкового автомобиля, в ходе которого было выяснено, что оптимальным алгоритмом решения задачи является градиентный бустинг, показывающий высокую точность при прогнозировании;
* Разработано и реализовано информационное и программное обеспечение;
* Проведено тестирование и отладка разработанного веб-приложения.

Спасибо за внимание! Готов ответить на ваши вопросы.