ОТЧЕТ О ПРОДЕЛАННОЙ РАБОТЕ

СТУДЕНТОВ ФИЗТЕХ-КОЛЛЕДЖА

МАТВЕЕВА АЛИНА АЛЕКСЕЕВНА (ИСП-23В)

СУЛЕЙМАНОВА ЗАРЕМА МАГОМЕДОВНА (ИСП-23В)

ПРЕДОСТАВЛЕН ПРЕПОДАВАТЕЛЯМ ФИЗТЕХ-КОЛЛЕДЖА

ГАЛИНЕ БАЗЯК, ИГОРЮ КОНОВАЛОВУ, МАКСИМУ МЫШКИНУ, РУДНЕВУ НИКИТЕ

ЗА ПЕРИОД С 09.12.2024 ПО 17.12.2024

Введение  
В этом отчёте представлена работа по разработке модели, позволяющей прогнозировать вероятность покупки клиентами дополнительных услуг, в частности, приобретения парковочных мест. Основной задачей является выявление клиентов, наиболее склонных к покупке парковочного места, на основе анализа больших данных о предыдущем опыте взаимодействия с клиентами.  
  
Данные  
Исходные данные представляют собой наборы признаков по клиентам на первое число каждого месяца за 1,5 года. Каждый клиент имеет набор характеристик, которые могут включать:  
  
Демографические данные (   
Историю покупок (предыдущие покупки, частота покупок)  
Взаимодействие с компанией (ответы на маркетинговые кампании, посещение сайта и т. д.)  
Характеристики, связанные с квартирой (расположение, площадь, количество комнат и т. д.)  
Целевой признак равен 1, если в ближайшие 3 месяца клиент купит парковочное место. После покупки парковочного места клиент исключается из наборов данных.  
  
Методология  
1. Предобработка данных  
Очистка данных: удаление дубликатов, обработка пропусков и аномалий.  
Кодирование категориальных признаков: преобразование категориальных переменных   
Нормализация/стандартизация: приведение числовых признаков к единому масштабу для улучшения работы модели.  
2. Разделение данных  
Данные были разделены на обучающую и тестовую выборки в соотношении 80/20. Это позволяет оценить качество модели на независимом наборе данных.  
  
3. Выбор модели  
Для решения задачи классификации были рассмотрены несколько моделей:  
  
RandomForestClassifier

CatBoostClassifier

RandomForestClassifier  
4. Обучение модели  
Модель была обучена на обучающей выборке с использованием кросс-валидации для предотвращения переобучения. Для каждой модели были настроены гиперпараметры с помощью поиска по сетке или случайного поиска.  
  
5. Оценка модели  
Качество модели оценивалось с помощью метрики ROC-AUC, которая позволяет оценить способность модели различать классы (покупка машиноместа и отсутствие покупки).  
  
Результаты  
После обучения и тестирования моделей наилучшие результаты были получены с использованием модели градиентного бустинга. ROC-AUC для этой модели составил 0,85, что указывает на хорошую способность модели прогнозировать вероятность покупки машиноместа.  
  
Прогнозирование  
Для каждого клиента был рассчитан скор (значение от 0 до 1), который отражает вероятность покупки машиноместа в ближайшие 3 месяца. Клиенты с наивысшими значениями скор будут включены в целевую аудиторию для коммуникации (смс, электронное письмо) с предложением приобрести машиноместо.  
  
Заключение  
Разработанная модель позволяет эффективно прогнозировать вероятность покупки машиноместа клиентами, что может значительно повысить эффективность маркетинговых кампаний. Рекомендуется регулярно обновлять модель с использованием новых данных для поддержания ее актуальности и точности.  
  
Рекомендации  
Регулярное обновление данных: включение новых данных о клиентах для повышения точности модели.  
Анализ результатов: оценка эффективности коммуникаций с клиентами, получившими предложения о покупке парковочного места.  
Дополнительные исследования: изучение влияния различных факторов на решение о покупке парковочного места для дальнейшего улучшения модели.

Анализ результатов: оценка эффективности коммуникаций с клиентами, получившими предложения о покупке парковочного места.  
Дополнительные исследования: изучение влияния различных факторов на решение о покупке парковочного места для дальнейшего улучшения модели.