



Прогнозирование вероятности покупки машиномест

Данная презентация посвящена проекту по разработке модели машинного обучения для прогнозирования вероятности покупки машиномест клиентами. Цель проекта – повысить эффективность маркетинга и увеличить доход компании.



by Alexey Kostk

Цель проекта:

**Создать модель, которая сможет предсказывать
вероятность
того, что клиент купит машиноместо.**

Это поможет компании:

**Эффективнее направлять рекламу: сосредоточиться на клиентах,
которые с большей вероятностью совершат покупку.**

**Экономить ресурсы: не тратить деньги на незаинтересованных
клиентов.**

Увеличивать прибыль: повышать процент успешных продаж.

Предварительный анализ данных

Изучение данных

Данные содержат числовые и категориальные признаки, а также пропуски. Целевая переменная (target) сильно дисбалансирована: только 3% клиентов совершили покупку. Выявлены дополнительные полезные признаки, такие как дата обращения.

Обработка данных

Числовые признаки заполнили медианными значениями. Категориальные признаки заполнили наиболее часто встречающимися значениями. Применили One-Hot Encoding для категориальных признаков.

Создание новых признаков

1

1. Выделение месяца и квартала

Выделили месяц и квартал из даты обращения.

2

2. Агрегирование средних значений

Агрегировали средние значения числовых данных (`mean_value`).

3

3. Балансировка классов

Использовали SMOTE для создания синтетических данных и уравнивания классов.



Моделирование

Логистическая регрессия

Базовая интерпретируемая модель.

Градиентный бустинг

Учитывает сложные зависимости в данных.

Случайный лес

Ансамблевый метод, устойчивый к шуму.

Результаты моделирования

0.9307

Случайный лес

Лучшая модель.

0.8892

Градиентный бустинг

0.6794

Логистическая регрессия





Визуализация и интерпретация



Важность признаков

Наиболее значимые факторы:
Количество посещений сайта.
Признаки региона клиента.
Признаки активности по времени
(месяц, квартал).



ROC-кривые

Сравнили модели по ROC-кривым,
что наглядно демонстрирует
превосходство случайного леса.



Предсказания для тестового набора

Выполнили прогнозирование для тестового набора данных. Сохранили результаты в CSV-файл: client_id: идентификатор клиента. score: вероятность покупки.

Рекомендации

Использовать модель для приоритизации клиентов в маркетинговых кампаниях. Регулярно обновлять данные для переобучения модели. Внедрить дополнительные признаки для повышения точности (например, поведенческие данные клиентов).



ВОПРОСЫ:

возможные вопросы и ответы на них

?

?

?

*СПАСИБО ЗА
ВНИМАНИЕ*