

цифровой
прорыв 

сезон: III

КЕЙС

Компания «Самолет»



Модель склонности
клиента к приобретению
машиноместа



Министерство
экономического развития
Российской Федерации



Кейсодержатель

Компания «Самолет»

01 Сфера деятельности

Компания «Самолет» реализует деятельность в сфере девелопмента и PropTech

02 Краткое описание кейса

Создание модели для поиска клиентов, наиболее склонных к приобретению машиноместа



Сайт организации

<https://samolet.ru/>
<https://career.samolet.ru/>

Постановка задачи

На основе больших данных о предыдущем опыте взаимодействия с клиентами разработать модель, позволяющую прогнозировать вероятность покупки клиентами дополнительных услуг, в частности, приобретения машиномест в паркинге.

Среди клиентов компании — владельцев квартир необходимо выделить покупателей, наиболее склонных к покупке машиноместа. С такими клиентами будет проводиться коммуникация (смс, эл. письмо) с предложением приобрести машиноместо.

Входные данные представляют собой наборы признаков по клиентам на первое число каждого месяца за 1,5 года (sample данных). Целевой признак равен 1, если в следующие 3 месяца клиент купит машиноместо. После покупки машиноместа клиент исключается из наборов данных.

Результатом предсказания модели должен стать скор (значение от 0 до 1) по каждому клиенту — вероятность, что клиент купит машиноместо в следующие 3 месяца. Оцениваться результат будет метрикой ROC-AUC.

Проблематика

Ключевую роль в продажах играет эффективная целевая рассылка. Рассылки позволяют оперативно информировать клиентов об актуальных предложениях и сервисах компании.

Однако каждая рассылка сопряжена с издержками:

- финансовые затраты - на подготовку и доставку сообщений (SMS, email и т.д.);
- временные затраты маркетологов и продавцов;
- риск раздражения получателей частыми сообщениями.

Это может негативно сказаться на лояльности клиентов в долгосрочной перспективе.

Поэтому компания хочет максимизировать отдачу от рассылок, ориентируясь только на наиболее заинтересованных и потенциально ориентированных на покупку клиентов.

Проблематика

Модель должна решать следующую проблематику:

- выделить только целевую аудиторию для каждой рассылки;
- избежать частых сообщений для неактивных клиентов;
- увеличить эффективность продаж благодаря точному целеполаганию.

Это позволит компании снизить затраты и улучшить лояльность клиентов за счет более персонифицированного подхода.

Решение

Ожидаемый результат: разработанная модель машинного или глубокого обучения, позволяющая на основе имеющихся данных о клиенте рассчитывать вероятность его склонности к конкретной услуге — покупке машиноместа.

В качестве подтверждения полученного результата ожидается csv-файл формата `submission_file.csv`. В нем необходимо обновить только колонку `score`.

Подготовленная презентация разработанной модели, содержащая основные аспекты:

- подход к решению бизнес-задачи;
- подробно изложенный процесс моделирования;
- полученные результаты;
- раскрытые инсайды, если есть.

Стек технологий, обязательных к использованию

Необходимые данные, дополнения, пояснения, уточнения

01

Jupyter notebook, Python, scikit-learn, pandas

02

Необходимо иметь в виду, что мы имеем дело с панельными данными - по каждому клиенту в датасете могут быть представлены несколько срезов за разные даты.

И в обучающем, и в тестовом датасете представлены не полные данные, а только подвыборки из ген совокупности.

При моделировании рекомендуется использовать кросс-валидацию.



Необходимые данные, дополнения, пояснения, уточнения

02

Будет плюсом, если презентация будет отражать следующую информацию:

- Какой предварительный анализ данных проводился? Есть ли какие-то интересные выводы?*
- Какие подходы к отбору признаков пробовали? Какой в итоге выбрали и почему?*
- Добавляли ли новые фичи?*
- Какую валидационную схему использовали?*
- Как подбирали гиперпараметры? Какие параметры сильно влияли на метрики?*
- Какие алгоритмы машинного обучения пробовали и какой выбрали для итоговой модели?*
- Работает ли модель стабильно? Наблюдалось ли переобучение? Как с ним боролись?*
- С какими сложностями столкнулись?*



Оценка

→ Для оценки решений применяется метод экспертных оценок и автоматизированные средства оценивания.

→ Жюри состоит из отраслевых экспертов и/или представителей кейсодержателя.

→ На основании описанных ниже характеристик, жюри выставляет оценки 0-3 балла.

→ Итоговая оценка определяется как сумма баллов всех экспертов: технического, отраслевого и/или представителя кейсодержателя, как значение, выданное автоматизированными средствами оценивания, либо как итоговый балл жюри, умноженный на оценку автоматизированной системы.

Технический эксперт оценивает решение по следующим критериям:

01

Запускаемость кода

02

Обоснованность выбранного метода (описание подходов к решению, их обоснование и релевантность задаче)

03

Точность работы алгоритма (возможность оценить формальной метрикой с обоснованием выбора)

04

Адаптивность/
Масштабируемость

05

Отсутствие в решении импортного ПО и библиотек, кроме свободно распространяемого с обоснованием выбора

06

Наличие интеграционных интерфейсов, в первую очередь интерфейсов загрузки данных

Автоматизированные средства оценивания точности работы предложенных участниками алгоритмов (решений) выставляют оценку в диапазоне 0-1, где 1 равно 100% точности работы решения.

Итоговая оценка определяется как итоговый балл жюри, умноженный на оценку автоматизированной системы.

цифровой
прорыв



сезон: ИИ



Министерство
экономического развития
Российской Федерации

