

Домашнее задание 2

В этом домашнем задании вам предстоит покрутить несколько алгоритмов машинного обучения, которые мы проходили на семинарах, попытаться применить техники улучшения качества моделей, а также сделать выводы о том, какая же модель лучше справляется с бизнес-задачей и почему она лучше других.

Описание кейса:

Сеть супермаркетов готовится к новогодней распродаже. Они запускают новый платный продукт — Gold Membership:

- даёт 20% скидку на все покупки в течение года;
- предложение доступно только для текущих клиентов;
- коммуникация через телефонные звонки / персонифицированные контакты.

Контакт-центр ограничен по ресурсам, а каждый звонок стоит денег (оклад операторов, время, инфраструктура).

Цель бизнеса:

Не «звонить всем подряд», а выбрать тех клиентов, которым с высокой вероятностью можно продать Gold Membership, чтобы:

- поднять выручку от программы;
- не сжечь бюджет на бесполезные звонки.

Ваша задача:

- 1) Сформулировать задачу и выбрать бизнес метрики, которые хочется улучшить с использованием модели. Выбрать ML-метрики, на которые будете опираться при моделировании и работе с результатами.
- 2) Построить минимум 2 модели машинного обучения: одна «простая» модель (линейная модель или дерево решений) и одна «сложная» модель (случайный лес или бустинг). Построение большого числа моделей дополнительных баллов не дает, однако может помочь в достижении лучшего результата и более подробного объяснения вашего решения.
- 3) Реализовать 6 подходов к улучшению качества каждой модели (каждый из них обсуждался на семинарах, всегда можно обратиться к материалам занятий). Насколько подходы действительно улучшают модель – вторично. Важно поэкспериментировать с их применением.
- 4) Подготовить файл с техническими расчетами, в котором будет описание каждого шага:
 - i. Выбор метрик (что хотим, какие подходят, какие не подходят, обоснование выбора итоговых)
 - ii. Описание разбиения данных на выборки

- iii. Краткое описание выбора моделей (одной простой, одной сложной: почему их, в чем их ожидаемые плюсы и минусы в этой задаче)
 - iv. Объяснение подходов к улучшению (что ожидаем перед применением, как проводим, какие результаты получаем)
 - v. Описание результатов работы моделей и объяснение, какую модель берем в работу
- 5) Подготовить презентацию с результатами вашей работы. Помимо результатов вашей работы,

!Важно! Насколько высокие метрики вы получите при построении модели будет минимально влиять на оценку (если крутые результаты – дадим дополнительные баллы; если метрики непривлекательные – не расстраивайтесь, объясните, почему делали именно так и почему могло не получиться). Основное, что от вас требуется – выстроить обоснованную логику действий в решении задачи и сделать выводы о том, какую модель, с какими улучшениями выбрать, основываясь и на результатах сравнительного анализа моделей, и на технических плюсах/минусах выбранных алгоритмов.

!!Важно!! ЧатГПТ при технической реализации – норм. При формулировании выводов – стрем.

Данные в приложенном .csv файле.

Описание данных:

Response (target) - 1, если клиент принял предложение в последней кампании, 0 в противном случае

ID - Уникальный идентификатор клиента

Year_Birth - Год рождения клиента (для определения возраста)

Complain - 1, если клиент жаловался за последние 2 года

Dt_Customer - Дата регистрации клиента в компании

Education - Уровень образования клиента

Marital - Семейное положение клиента

Kidhome - Количество маленьких детей в домохозяйстве клиента

Teenhome - Количество подростков в домохозяйстве клиента

Income - Годовой доход домохозяйства клиента

MntFishProducts - Сумма, потраченная на рыбные продукты за последние 2 года

MntMeatProducts - Сумма, потраченная на мясные продукты за последние 2 года

MntFruits - Сумма, потраченная на фрукты за последние 2 года

MntSweetProducts - Сумма, потраченная на сладости за последние 2 года

MntWines - Сумма, потраченная на вино за последние 2 года

MntGoldProds - Сумма, потраченная на золотые изделия (ювелирные украшения) за последние 2 года

NumDealsPurchases - Количество покупок, совершенных со скидкой (по акции)

NumCatalogPurchases - Количество покупок, совершенных через каталог (заказ товаров по почте)

NumStorePurchases - Количество покупок, совершенных непосредственно в магазинах

NumWebPurchases - Количество покупок, совершенных через веб-сайт компании

NumWebVisitsMonth - Количество посещений веб-сайта компании за последний месяц

Recency - Количество дней с момента последней покупки