**Домашнее задание №9**

Вам нужно построить модель, предсказывающую, оформит ли пользователь заказ, сформированный в корзине интернет-магазина.

В качестве источников данных доступно несколько систем, включая:

* информацию о пользователях и истории их товаров,
* информацию о товарах,
* информацию о корзине.

Разложите задачу по CRISP-DM. Для каждого шага сформулируйте:

* что планируете на нём делать,
* каких результатов от него ожидаете,
* какие роли специалистов вам нужны на этом шаге.

**План проекта по Crisp-DM**

|  |  |  |
| --- | --- | --- |
| **Этап** | **Задачи** | **Роли** |
| Business Understanding | * 1. Вывести метрику конверсии корзины в покупку. Определить типичные значения, проверить наличие сезонности.   2. Определить, как заказчик видит использование полученной модели, сформулировать минимально необходимое качество.   3. Оценить ожидаемый эффект от такой модели и сравнить его с ожидаемыми трудозатратами. | Владелец продукта, аналитик |
| Data Understanding | 1. Собрать все имеющиеся данные воедино (информация о пользователях и история их товаров, информация о товарах, информация о корзине) 2. Описать собранные данные (количество, типы, значения, связи). 3. Провести оценочное исследование данных. 4. Оценить качество данных на предмет наличия ошибок, отсутствующих значений, полноты. | Аналитик |
| Data Preparation | 1. Очистить данные от пропусков, повторяющихся значений и ошибок. 2. Обогатить данные для увеличения количества признаков, на основании которых будет предсказываться результат. 3. Разделить данные на выборки test и train 4. Сохранить выборки в отдельные датафоеймы. | DS, DE |
| Modeling | 1. Выбор модели – поскольку предсказываем целевую переменную (оформит клиент или не оформит заказ) – это будет модель линейной регрессии. 2. Построить модель линейной регрессии. 3. Произвести оценку качества модели. 4. Описать результат моделирования. | DS |
| Evaluation | 1. Оценить результат. Для оценки используем сравнение конверсии корзины в покупку, полученную при моделировании со значениями аналогичной конверсии, полученными на этапе Business Understanding. 2. Если результат не соответствует ожиданиям – анализ процесса моделирования, поиск альтернативных решений (например, до обогащение данных за счет дополнительных источников) 3. Определить дальнейшие шаги (либо переход на этап Business/Data Understanding, подготовка в внедрению). | Аналитик |
| Deployment | 1. Создание плана внедрения для каждой модели (их будет несколько – с учетом товарных групп, сезонности, времени суток и политико-экономической ситуации в стране) 2. Определение возможных проблем при внедрении. Описать пути их решения. 3. Составить план мониторинга и технического обслуживания (при необходимости) 4. Провести итоговый обзор проекта. | Владелец продукта, Аналитик, DS, Разработчик |