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

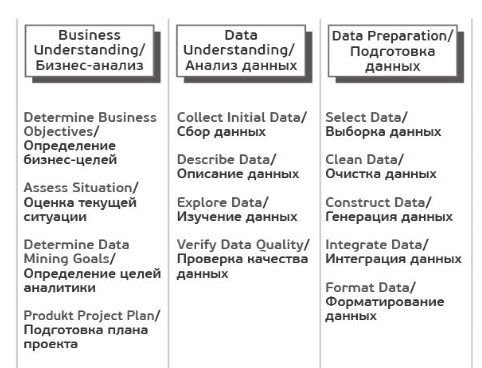
В качестве источников данных доступно несколько систем, включая:

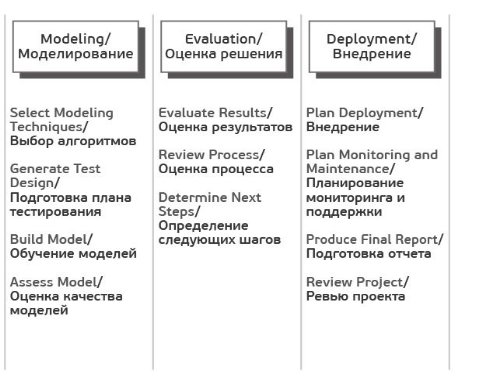
* информацию о пользователях и истории их товаров,
* информацию о товарах,
* информацию о корзине.

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.  
  Пример — business understanding.  
  На этом шаге планируется выявить основную боль бизнеса.

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.





**CRISP-DM**

|  |  |  |  |
| --- | --- | --- | --- |
| **Этап** | **Запланированные действия** | **Результаты** | **Роли привлекаемых на этап специалистов** |
| **Business Understanding ( бизнес анализ)** | 1.Вывести метрику конверсии корзины в покупку. Определить типичные значения, проверить наличие сезонности. | 1. Определены критически важные параметры необходимые для построения модели.  2. Показаны сезонные колебание в спросе на определённые группы товаров.  3. Определены типичные показатели конверсии | **BI- аналитик** |
| 2. Определить, как заказчик видит использование полученной модели, сформулировать минимально необходимое качество. | 1. Определены пределы применимости результатов модели для различных групп товаров. | **BI - аналитик Владелец продукта** |
| 3. Оценить ожидаемый эффект от такой модели и сравнить его с ожидаемыми трудозатратами. | 1. Получен планируемые экономический эффект. Показано, что эффект от внедрения модели за счет оптимизации связанных предложений на 15% процентов может увеличить показатель конверсии , что при планируемых затратах на разработку поможет увеличить общую прибыль на 10%. | **Экономист, Владелец продукта** |
| **Data Understanding (Анализ данных)** | 1. Сбор имеющиеся источников данных:  БД информации о пользователях и истории покупок товаров  БД товаров  Информацию о набранных пользователям корзинах | 1. Определены источники и способы доступа к данным.  2. Получены необходимые права доступа  3. Протестированы возможности выгрузки, выполнения запросов. При необходимости добвалены новые представления  4. Дана оценка производительности операций с хранилищами (базами) данных. | **Аналитик данных, ETL-специалист, Data Engineer, или Специалист по БД** |
| 2. Получить подробное описание данных: полей, типов данных, связей, количества записей. | 2. Получено подробное описание данных: полей, типов данных, связей, количества записей. | **ETL-специалист, Data Engineer, или Специалист по БД** |
| 3. Определить общие статистические показатели данных. Выполнить разведочный анализ данных | 3. Определены общие статистические показатели данных и выполнен разведочный анализ данных | **Аналитик данных** |
| 4. Изучить качество имеющихся данных: пропуски, выбросы, ошибки заполнения. | 4. Получены характеристики качества данных | **Аналитик данных** |
| **Data Preparation (Подготовка данных)** | 1. Провести очистку данных. Исключить некорректные данные | 1. Исключены ошибочно введённые, данные.  2. Устранено дублирование данных. | **Аналитик данных** |
| 2. Выбрать алгоритм заполнения пропусков и заполнить пропуски. | Из данных устранены пропуски | **Аналитик данных и(или) DS - специалист** |
| 3.Решить проблему выбросов. | Из данных устранены выбросы | **Аналитик данных и (или) DS - специалист** |
| 4. Провести обогащение данных. | В общий датафрейм сведены данные о существенных признаках приобретаемых товаров, информация о пользователях и корзинах | **Аналитик данных** |
| 5. Получить тестовый датафрейм для обучения модели. | Получен качественный датафрейм для обучения модели | **DS - специалист и (или) Аналитик данных** |
| 6. Выполнить разделение данных на обучающую и тестовую выборки. | Получены обучающая и тестовая выборки | **DS - специалист** |
| **Modeling (Моделирование)** | 1.Выбрать подходящие для решения задачи методы моделирования | Отобраны модели позволяющие решить поставленную задачу | **DS - специалист** |
| 2. Построить модели | Построен набор моделей | **DS - специалист** |
| 3. Получить результаты | Выполнено моделирование на тестовом датафрейме | **DS - специалист** |
| 4. Выбрать оптимальную модель по метрикам качества | Проведена оценка качества моделирования. Выбрана модель показавшая лучшие прогнозные свойства | **DS - специалист** |
| 5. Описать результаты | Результаты представлены в оптимальной и понятной для ЛПР форме в виде дашборда, динамического отчёта с необходимыми визуализациями | **DS – специалист и BI- аналитик** |
| 6. Оценка полученных результатов и их соответствия поставленным задачам. | Проведён анализ результатов и дана оценка полученным инстайтам | **BI- аналитик и Владелец продукта** |
| **Evaluation (Оценка решения)** | 2. Наметить круг новых вопросов и подзадач появившихся после оценки модели | Определён круг новых вопросов и подзадач появившихся после оценки модели | **BI- аналитик и Владелец продукта** |
| 1. Провести настройку продукта на площадке владельца продукта | Рабочая модель функционирует на реальных данных владельца продукта | **DS – специалист, Разработчик, Data Engineer** |
| **Deployment (Внедрение)** | 2. Предложить интеграцию полученного решения в ИС владельца продукта | Получены и отработаны спецификации для интеграции модели в ИС владельца продукта | **DS – специалист, Разработчик, Data Engineer** |
| 3. Разработать необходимый API для интеграции | Разработан и задокументирован API для интерграции модели в ИС владельца продукта | **DS – специалист, Разработчик** |
| 4. Предложить план актуализации модели при изменение показателей эффективности | Разработаны метрики для индикации актуальности модели, разработана и утверждена дорожная карта поддержки решения | **Аналитик данных**  **DS – специалист,**  **Data Engineer, Владелец продукта** |