**Формирования предиктивной модели**

**покупок в интернет магазине на основе CRISP-DM**

1. **BUSINESS UNDERSTANDING**

* Обозначить метрику конверсии корзины в покупку. Определить типичные значения, проверить наличие сезонности
* Определить как заказчик видит использование полученной модели, сформулировать минимально необходимое качество
* Оценить ожидаемый эффект от такой модели и сравнить его с ожидаемыми трудозатратами  
  Для успешного выполнения данного этапа потребуется Аналитик и Владелец продукта

**2. DATA UNDERSTANDING**

* Загрузить необходимые данные из базы сеансов посещений с конверсией в корзину. Для идентификации сеансов формируется идентификатор на основе файлов cookie. Каждая сессия обозначается ориентиром, где ориентир может принимать n различных атрибутов с m различными значениями.
* Определить удельный вес влияния канала входа на решение о покупке и отказе от покупки. Выделить ориентиры, описывающие сеанс или взаимодействие с клиентом, как например переход на страницу и добавление чего-либо в корзину, чтобы уменьшить размер набора данных.
* На данном этапе потребуются ETL-специалисты, Web - аналитики, DS-специалисты

**3**. **DATA PREPARATION**

* Из данных необходимо удалить сеансы, вызванные не людьми, а ботами. Обнаружить и устранить трафика, генерируемый ботами. Данные должны быть преобразованы таким образом, чтобы каждое действие посетителя в течение сеанса записывалось в одну строку базы данных. Рассчитать ориентирующую метрику показывающую вероятность покупки каждый раз, когда в течении санса товар добавляется в корзину Определить профиль клиента и выбрать переменные: возраст, пол и последнее устройство входа.
* С помощью аналитической выборки из данных, сформировать набор функций и параметров для создания входных данных для алгоритмов машинного обучения.
* На данном этапе потребуются DS-специалисты, Data Engineers

**4. MODELING**

* Реализовать обработку массива данных используя определенные методы моделирования:Boosted Tree, Random Forest, Logistic Regression, Recurrent Neural Network. Построить оптимальную модель на результатах сравнительных оценок и тестов. Обучить модель.
* Получить прогнозные значения конверсии корзины в покупку.
* На данном этапе потребуются DS-специалисты, Data Engineers

**5. EVALUATION**

* Оценить точность предсказания модели на различных пороговых значениях. Сравнить результаты полученные на разных моделях по затратам, эффективности, релевантности в целевых группах покупателей. Понять насколько модель справилась с поставленной задачей, достаточна ли выборка и требуется ли откат на предыдущие стадии.
* Получить результат в форме матрицы ошибок и на основе показателей целевой метрики сделать вывод о надежности результатов. Ответить на вопрос какая модель машинного обучения лучше всего подходит для решения задачи прогнозирования конверсии пользовательской корзины в покупку.
* На данном этапе потребуются Data Engineers, а также присутствие Владельца продукта(Заказчика) для получение обратной связи по оценке эффективности полученных прогнозных значений и степени их применимости в бизнес-процессе

**6. DEPLOYMENT**

* Совместно с Владельцем продукта провести планирование мониторинга конверсии для формирование последующих бизнес-стратегий.
* Провести итоговый обзор проекта и реализовать в бизнес-продукт.
* Для успешного выполнения данного этапа потребуется DS, Web-разработчик и Владелец продукта.