

Построение модели прогнозирования продаж в единицах товара (на примере внедрения модели в сеть крупного спортивного ритейлера).

1. Business understanding (Бизнес постановка)

1.1. Определение бизнес-целей

Достижение такого высокого качества модели, позволяющего использовать прогноз для системы пополнения товара.

1.2. Оценка текущей ситуации

Определение временных возможностей экспертов из бизнеса для выдачи консультаций.

Определение наличия необходимого железа для проекта.

Оценка возможных рисков.

1.3. Определение целей аналитики

Определение используемых методов прогнозирования для определения baseline для проекта.

Выставление критериев качества модели, например, R2 общий и детальный для конкретного товара или группы товаров, MSE, RMSE, высокий прогноз в периоды крупный промокампаний.

1.4. Подготовка плана проекта

Определение сроков достижения результата.

2. Data Understanding (Анализ данных)

2.1. Сбор данных

Получение доступа к хранилищу данных, получение необходимых знаний о структуре данных, качестве, источниках, времени репликаций.

2.2. Описание данных

Описание всех данных в источниках, сбор статистики по данным.

2.3. Изучение данных

Выявление важностей переменных, гистограмм, попытка найти закономерности в данных.

Выявление особенностей данных, связанных с работой Компании (например, влияние крупных промоактивностей на всплески продаж).

2.4. Проверка качества данных

Проверка полноты данных после периодических репликаций.

3. Data Preparation (Подготовка данных)

3.1. Выборка данных

Выявление максимально значимых атрибутов.

Выявление корреляции с другими атрибутами.

3.2. Очистка данных

Избавление от выбросов, например, выбрасывание данных об оптовых непланируемых закупках юридических лиц с целью оставления информации только о физических клиентах.

3.3. Генерация данных

Создание новых описательных переменных вроде флагов промо активностей на определенную дату и скользящих средних по продажам товара за предшествующие периоды.

3.4. Форматирование данных

Приведение данных к необходимому для использования алгоритмов вид (например, шкалирование).

4. Modeling (Моделирование)

4.1. Выбор алгоритмов

Перебор основных алгоритмов, таких как случайный лес, градиентный бустинг, нейронные сети.

4.2. Подготовка плана тестирования

Обучение на исторических данных, тестирование на свежих данных.

4.3. Оценка качества моделей

Выбор модели с наилучшими показателями R^2 и RMSE.

5. Evaluation (Оценка решения)

5.1. Оценка результатов

Презентация результатов аналитикам из бизнеса для получение оценки результата.

5.2. Оценка процесса

Поиск неудачных мест, выявленных в ходе проекта, с целью недопущения подобного в будущем.

5.3. Определение следующих шагов

Принятие решения по внедрению модели или проведении последующих доработок.

Создание новых моделей для других сопутствующих задач, например, продаж товара в деньгах.

6. Deployment (Внедрение)

6.1. Внедрение

Настройка периодических репликаций данных, планового инкрементального дообучения модели и выгрузки результатов в приемлемый для заказчика вид.

Проведение периодического тюнинга модели.

6.2. Планирование мониторинга и поддержки

Создание системы контролинга за качеством работы модели и репликации данных.

6.3. Подготовка отчета и ревью проекта.