1. Uplift моделировани

Содержание:

Мы являемся крупным ритейлером. Мы регулярно проводим рекламные и скидочные кампании путём рассылки СМС, email, сообщений в мессенджерах и тд. Мы хотим оптимизировать рассылку уведомлений клиентам, увеличив их отклик и минимизировав затраты на проводимые кампании. Мы видим выход в разработке ML решения. У нас есть данные о прошлых кампаниях: различные характеристики клиента, флаг получал ли он рассылку, флаг купил ли клиент промотируемый товар.

Вопросы:

* Предложите архитектуру ML решения?
* Какую нам выбрать целевую?
* Какие переменные по клиенту потенциально могут быть полезными?
* Как тестировать получаемые модели?
* Какой метрикой оценивать качество получаемых моделей?
* Как вы видите развёртывание полученного решения, его режимы работы, потребляемые ресурсы?

1. Исправитель адресов

Содержание:

Наша компания использует продукт от вендора, главный функционал которого - нормализовывать адреса и раскладывать их на географические объекты. К сожалению, сейчас нормализатор от вендора успешно справляется только с 70% приходящих адресов. Мы хотим улучшить этот показатель, разработав решение, которое позволило бы предобрабатывать адреса, тем самым повышая результативность нормализатора. У нас есть датасет из адресов с указанием, справился ли нормализатор. Разрабатываемое решение должно быть способн вписать в наш технологический ландшафт и выдержать имеющуюся нагрузку. Нормализатор работает на кластере с данными в HDFS. Интенсивность обработки данных ~ 800000 адресов в день

Вопросы:

* Какие данные для построения решения вам потребуются?
* Какой вы видите архитектуру возможного решения?
* Какие методы обработки адресов вы видите возможным использовать для решения задачи?
* Какие технологии вы считаете подходящими для решения поставленной задачи?
* В случае использования ML, как вы видите процесс обучения моделей?

1. Сеть пунктов выдачи

Содержание:

Мы крупный ретейлер. В наших планах - создать сеть пунктов выдачи заказов, разместив их в наших магазинах: на кассах, в виде постаматов и тд. Мы провели пилот, разместив пункты выдачи в наших магазинах в Н-ской области. На основании полученных после пилота данных нам необходимо определить, в каких точках и в каких форматах размещать пункты выдачи в М-ской области. Предложите решение поставленной задачи. Данные о пилоте содержат информацию о количестве выданных за время пилота товаров в каждой точке, координаты точки, данные об типе точки. Данные о точках в М-ской области содержат полную информацию о точках: координаты, описание окрудения и тд.

Вопросы:

* Какие методы вы использовали бы?
* Какие дополнительные данные вы взяли бы?
* Какую метрику и функцию ошибки будете оптимизировать?
* Предложите способ визуализации полученных результатов

1. Модель прогнозирования внутреннего мошенничества

Содержание:

Перед Вами стоит задача выявления мошеннических действий сотрудниками Банка по доступным к отчетной дате данным. Необходимо предложить подход к решению данной бизнес-задачи.

Вопросы:

* Каким образом можно формализовать зачаду?
* Какие данные могут быть полезны для решения этой задачи?
* Как Вы будете собирать разметку для обучения модели?
* Как оценить финансовый результат использования данной модели?
* Какие проблемы могут возникнуть при моделировании?

1. Определение активности работы сотрудников и их эффективности на удаленной работе.

Содержание:

Имеются разного рода данные о сотруднике (орг. структура, опыт и т.д). Также имеются данные о его активности (время входа/выхода в банк, ВПН, почтовая переписка, активность по задачам (jira, bitbucket). Необходимо оценить, как удаленная работа повлияла на активность сотрудников, выделить "плохишей" и наиболее загруженных.

Вопросы:

* Как бы вы подошли к решению данной задачи?
* Какие новые признаки можно создать?
* Как классифицировать сотрудников на основе имеющихся данных?

1. Модель отклика на предложение кредита

Содержание:

Вы создаете модель оценки вероятности отклика на предложение банковского кредита. Клиентская база насчитывает 1 млн. клиентов, при этом клиентские менеджеры за 1 неделю могут довести до сведения предложения физически только для 10 000 клиентов. Пусть для всех клиентов одинаковая маржа от выдачи кредита (т.е. одинаковые сумма, срок, ставка, вероятность дефолта, фондирование) в размере 50 000 р. Издержки на взаимодействие менеджера с одним клиентом пусть так же одинаковы для всех менеджеров и клиентов - 1000 р.

Вопросы:

1. Сформулируйте задачу для моделирования. Как определите целевое событие?

2. Какие исторические данные вы бы использовали для включения в модель? Предположим, мы можем достать любые данные.

3. Какую метрику качества модели выберите?

4. Какое решение на основе прогноза модели вы сделаете, т.е. по каким именно образом определите список клиентов для предложения кредита?

5. Предположим, что до модели клиенты отбирались экспертно. Какие шаги можно проделать, чтобы оценить финансовый эффект от внедрения принятия решений на основе модели?

1. Автоматизация работы диспетчеров центра поддержки

Содержание:

В центр поддержки пользователи (внутренние сотрудники компании, либо клиенты) пишут сообщения с описанием проблем, возникших при работе с системами банка. Каждое сообщение анализируется диспетчером и направляется в профильное подразделение для решения проблемы. Предположим, что на классификацию сообщения в среднем диспетчер тратит 15 мин и число сообщений на 1 диспетчера составляет 160 в неделю. Всего сейчас работает 120 диспетчеров. Вам дали задачу - построить, провалидировать и внедрить модель, которая автоматизирует процесс маршрутизации сообщений и позволит держать меньшее число диспетчеров.

Вопросы:

1. Сформулируйте задачу для моделирования. Как определите целевое событие? Какой это типа задачи с точки зрения ML?

2. Какие исторические данные вы бы использовали для включения в модель? Предположим, мы можем достать любые данные.

3. Какую метрику качества модели выберите?

4. Какое решение на основе прогноза модели вы сделаете, т.е. по каким именно образом определите список клиентов для предложения кредита?

5. Как вы измените процесс после создания модели и как можно оценить финансовый эффект от внедрения принятия решений на основе модели? Предположим, что если сообщение пришло в профильное подразделение по ошибке, его возвращают на ручной разбор.