

1. Содержание: Для расчета справедливой стоимости одного дериватива в среднем уходит 10 секунд на одном процессоре (моделируются стохастические случайные процессы, на каждой случайной траектории определяется стоимость дериватива, справедливая стоимость получается усреднением по всем траекториям). Справедливая стоимость зависит от текущих рыночных данных (что составляет примерно 50 параметров) и внутренних параметров каждого дериватива (тоже примерно 50 параметров). В портфеле банка находится 1000000 деривативов. Вычислительные мощности выделенные для расчета справедливой стоимости составляет 100 процессоров. Итого, для расчета справедливой стоимости всего портфеля потребуется больше 24 часов работы сервера. Для оперативного управления портфелем, из-за быстрого изменения рыночных данных, требуется ускорить процесс расчета до 30 минут..

Вопросы:

- Какое решение вы бы предложили (кроме тривиального - нарастить вычислительные мощности) ?
- Чем, как вы считаете, можно поступить для ускорения процесса? Какие выборы между показателями расчета (trade-off) возникают в данной задаче?
- Формализуйте задачу.
- Какими бы конкретными методами вы бы решали ее?

2. Содержание: Сейчас клиентский менеджер из общения с клиентом чувствует, нужно ли сделать скидку по ставке по кредиту, если клиент сомневается, брать ли у нас кредит, или наоборот, установить ставку побольше, так как клиент готов получить деньги на любых условиях.

Вопросы:

- Как сформулировать задачу в терминах машинного обучения для автоматического назначения цены?
- Какие данные нужно собрать?
- С какими проблемами при определении оптимальной ставки для клиента можно столкнуться?
- Предложите дизайн пилота для проверки результата модели.

3. Содержание: В крупной компании наблюдается высокий уровень текучести кадров. Руководитель отдела HR поручил Вам заняться этим вопросом. Руководство хочет снизить отток и:

- Определять склонного к увольнению сотруднику
- Оценивать результативность применения модели (её влияние на отток)
- Снизить отток

Вопросы:

- Какие данные будете собирать
- Будет ли выборка сбалансированной, какие у этого последствия.
- Предложите архитектуру модели
- Какие метрики качества следует использовать для оценки результатов предсказания
- Как построить процесс тестирования модели
- Как определить факторы, ведущие к увольнению, на каждом конкретном сотруднике
- Как понять, что полученное снижение оттока не случайно, и вызвано результатом применения модели.

4. Содержание: Банк заинтересован в том, чтобы клиенты не допускали "просрочку" по кредитам, т.е. совершали платежи по кредиту вовремя, согласно графику платежей. Для этого Банк рассылает клиентам напоминания о предстоящем платеже за несколько дней до даты планового платежа. Ваша задача - определить, каким клиентам нужно напоминать о предстоящем платеже и через какой канал - SMS, EMAIL, PUSH-уведомление, Звонок "Робота" (разные типы коммуникаций имеют разную стоимость). Цель - максимизировать долю клиентов, которые совершили платеж вовремя и ограничить затраты на рассылки.

Вопросы:

- Почему делать рассылку всем подряд может быть не выгодно ?
- Как можно применить модели машинного обучения для решения этой задачи ?
- Как можно сформулировать целевой признак (или целевые признаки) для этой модели (или моделей)?
- Какие метрики качества стоит использовать для оценки качества этих моделей?
- Какие данные могли бы быть полезны для обучения таких моделей?
- Как бы вы сформировали выборки train/valid/test для обучения таких моделей?
- Если предположить, что потери Банка от пропуска платежа зависят так же от величины кредита, по которому клиент допустил просрочку, то как эту информацию можно использовать при обучении моделей?

5. Содержание: У вас есть датасет клиентов, которым была одобрена заявка на кредит за 2020 год. Описание датасета: ключ (уникальный клиент), дата одобрения заявки на кредит, процентная ставка по кредиту предложенная клиенту (клиенту предлагалась случайная ставка), регион (Москва, Хабаровский край и т.п.), факт согласия взять кредит (бинарный признак: согласился взять кредит = 1 или нет = 0). Введем понятие чувствительности (эластичность) клиента к процентной ставке: изменение вероятности согласиться на предложенные условия по кредиту с ростом предложенной процентной ставки. Средняя чувствительность по всему потоку равна 5% (уменьшение вероятности взять кредит на 5 п.п с увеличением ставки на 1 п.п. - это процентный пункт). Задача: дать заключение о возможности выделить чувствительные и нечувствительные регионы, т.е. регионы, в которых клиенты имеют большую или меньшую среднюю чувствительность.

Вопросы:

1. Почему есть основания считать, что регионы могут быть разные по чувствительности? (логическое обоснование)
2. Предложить алгоритм определения чувствительных и нечувствительных регионов
- 2.1. Как для региона посчитать среднюю чувствительность?
3. Как использовать знание о чувствительных и нечувствительных регионах для увеличения прибыли банка
4. Какой показатель (помимо классических метрик, показывающих точность прогноза модели) отражает финансовый эффект (с обоснованием)
5. Предложите алгоритм для определения стабильности эффекта этой модели с учетом изменения ситуации на рынке
6. Какие дополнительные факторы (неописанные в датасете) еще могут влиять на эластичность (обоснование)

6. Содержание: В банке существует модель, которая определяет потенциал дохода, который клиент может принести банку по каждому из банковских продуктов.

Расчет по модели производится каждые полгода. Потенциал клиента влияет на то, будет ли у клиента персональный менеджер, и какой план этому менеджеру поставят.

Необходимо, не погружаясь в код модели, дать оценку ее работе с точки зрения бизнеса.

Вопросы:

- Как данный процесс может влиять на бизнес банка
- Как бы вы оценивали финансовый эффект от модели, какие данные для этого нужны
- На каких этапах процесса могут возникнуть проблемы для бизнеса
- Как эти проблемы можно выявить, какие данные потребуются