Design Doc Competition 2023

ODS

Team

Sergey Artiukhin

TG : @unipplyo
Content SIMULATOR ML

Dmitriy Naumenko

TG : @naumenko_ds
Simulator ML Student

Eduard Poliakov

TG : @EdwPol
Junior DS,
Souther Federal University,
Student of SimulatorML

Sabrina Sadiekh

TG : @sabrina_sadiekh
https://t.me/jdata_blog
PetrSU Student

Artem Savelyev

TG : @w3ban
Simulator ML Student

Цели и предпосылки

- **Бизнес-Цель:** увеличение эффективности системы антифрода для снижения финансовых и репутационных рисков связанных с мошенническими действиями.
- **Почему ML:** сравнение с правилами
- Что будем считать **успехом** с точки зрения бизнеса

Стоимость

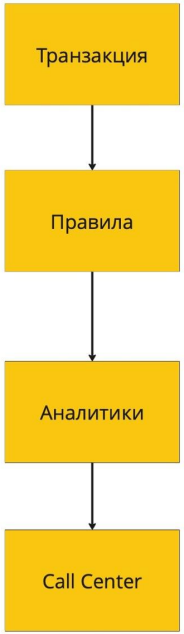
- **Затраты на работу датасаентистов:** затраты на разработку и внедрение
- **Затраты на инфраструктуру:** развертывание и поддержание
- **Финансовые издержки:** связанные с применением (штрафы, потеря лояльности клиентов) и поддержкой антифрод системы

Бизнес требования

- Повышение эффективности обнаружения и предотвращения фродовых операций путем внедрения ML-решения
- Время обработки не более 1 сек
- Интерпретируемость моделей
- Ограниченное количество аналитиков для проверки транзакций и тайминг
- Минимизация ложные положительные срабатывания

Стадии

- Оценка
- Пилот
- Продуктив



Ключевые показатели:

1. **Размер штрафа** в конце месяца: суммарный размер штрафов, за каждую пропущенную фрод транзакцию. Величина штрафа равна сумме транзакции.
2. **Профит от транзакций**: процент вознаграждения за успешно проведенные обычные транзакции
3. **Количество жалоб**: позволяет измерить, насколько успешно система антифрод предотвращает мошенническую активность и показывает удовлетворенность пользователей как проксиметрика потери клиентов.
4. **Затраты на систему:** работа аналитиков, является второй целью после качества

Бизнес метрика

$$profit = \lambda * \sum_{t}^{TN} p_t - (\sum_{t}^{FN} p_t + p_c * \beta * m + \mu * n)$$

Профит от транзакций
lambda - процент, который банк получает за успешную транзакцию
TN (True Negative) - к-во транзакций, помеченное моделью как не фрод, при этом в действительности фродом не являющеяся
p_t - сумма транзакции

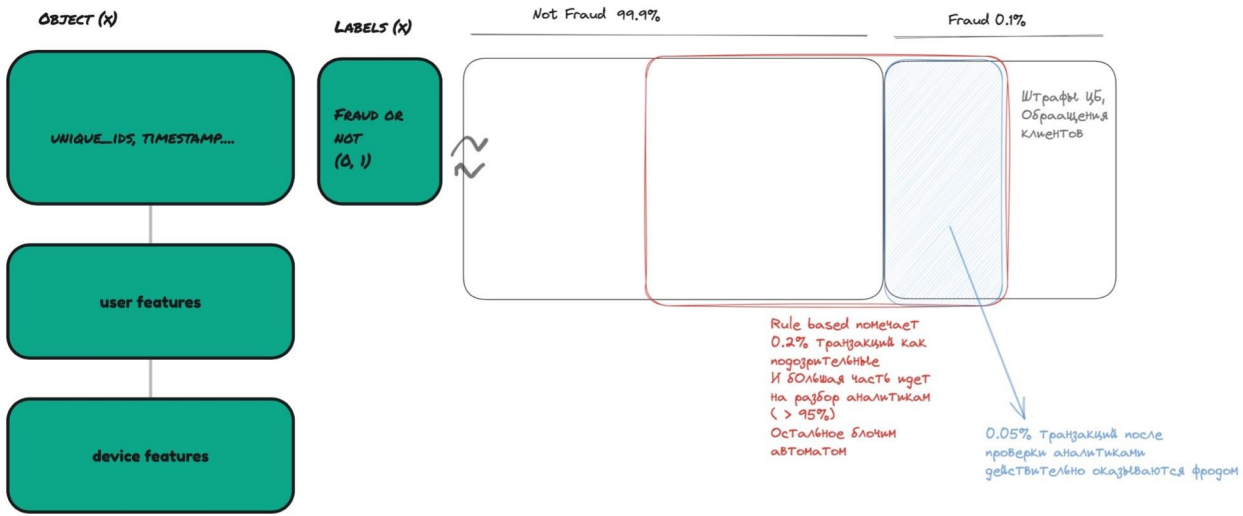
Сумма штрафов
FN (False Negative) - к-во транзакций, помеченное моделью как не фрод, при этом в действительности являющиеяся фродом

Оценка убытка от ушедших клиентов
beta - *средняя вероятность ухода клиента после жалобы*
p_c - средний убыток от потерянного клиента
m - количество жалоб на ложную блокировку транзакции

Затраты на аналитиков
n - *количество подозрительных транзакций, обрабатываемые аналитиками*
ти - средняя стоимость обработки одной подозрительной транзакции

$$nProfit = \frac{profit}{max\ profit} \qquad max\ profit = \lambda * \sum_{t}^{N} p_t$$

DATASET

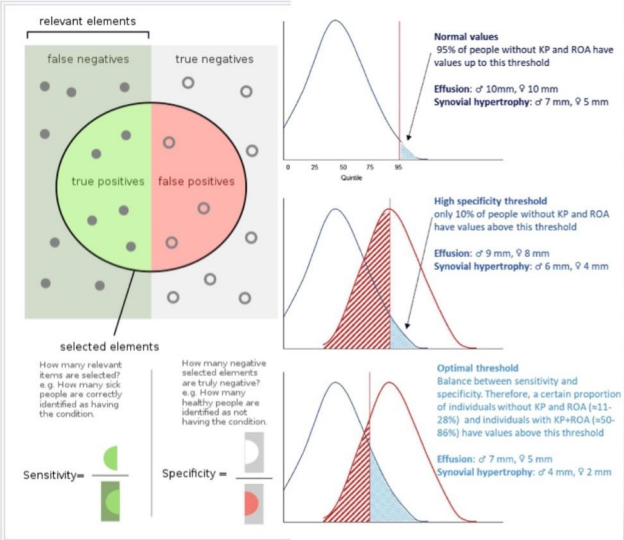


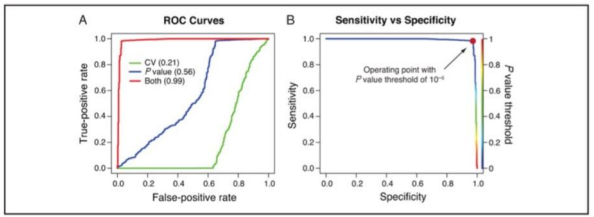
- Проверка качества временного ряда в рамках выбранного отрезка времени
- Оценка распределений целевого признака для типа перевода Card2Card/Account2Account. Проверка гипотез о специфичностях категорий (разные по сумме типичные транзакции, разная частота транзакций, разное соотношение мошеннических транзакций);
- Проверка гипотез о сезонности. Уточнение задачи: детекция аномалий или детекция аномалий с учетом временного ряда. Определение отрезка времени, необходимого для пересчета модели;
- Оценка корректности разметки данных. Проверка распределений по каждой переменной для каждого из имеющихся признаков;
- Оценка важности признаков для baseline и MVP, формулирование гипотез для новых признаков с целью улучшения качества системы (возможные варианты признаков – время между транзакциями одного пользователя, обыкновенная частота транзакций для времени/дня недели, неизвестный (новый) девайс пользователя);
- Признак времени с последний транзакции, количества транзакций за последние N минут
- Признак нового устройства, локации

METRICS

OFFLINE

RECALL@SPECIFICITY > 99.5





BUSINESS

NORMALIZED PROFIT

Определим profit в деньгах как основную бизнес метрику на основе этих показателей

$$profit = \lambda * \sum_{i}^{TN} p_i - (\sum_{i}^{FN} p_i + p_c * \beta * m + \mu * n)$$

Где

профит от транзакций (первое слагаемое):

lambda - процент, который банк получает за успешную транзакцию

TN(True Negative) - к-во транзакций, помеченное моделью как не фрод, при этом в действительности фродом не являющееся

p_i - сумма транзакции

сумма штрафов (первое слагаемое со знаком минус)

FN(False Negative) - к-во транзакций, помеченное моделью как не фрод, при этом в действительности являющиеся фродом

оценка убытка от ушедших клиентов (второе слагаемое со знаком минус):

beta - *средняя вероятность ухода клиента после жалобы*

p_c - средний убыток от потерянного клиента

m - количество жалоб на ложную блокировку транзакции

затраты на аналитиков (третье слагаемое со знаком минус)

n - *количество подозрительных транзакций, обрабатываемые аналитиками*

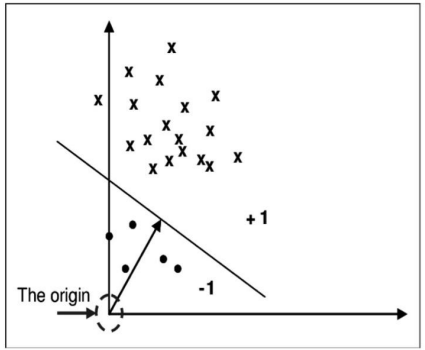
mu - средняя стоимость обработки одной подозрительной транзакции

Определяем ключевые показатели, которые далее агрегируем в единую формулу в деньгах, понятную бизнесу. Показателями являются:

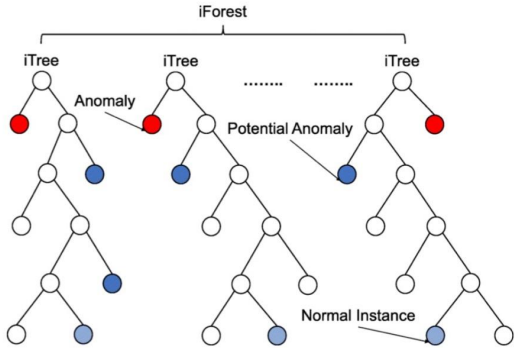
- Размер штрафа в конце месяца: эта метрика отражает суммарный размер штрафов, наложенных на клиентов за мошенническую активность. Успешная работа системы антифрод должна уменьшать этот размер, указывая на эффективность ее действий.
- Профит от транзакций: данная метрика позволяет оценить, насколько успешно система антифрод предотвращает мошеннические транзакции и снижает финансовые потери компании.
- Количество обращений в поддержку: это позволяет измерить, насколько успешно система антифрод предотвращает мошенническую активность и снижает необходимость клиентов обращаться в поддержку в связи с мошенническими операциями.

MODEL

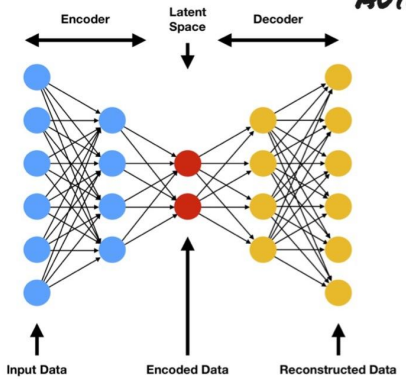
One-Class SVM (Support Vector Machine)



Isolation Forest



Auto-Encoder

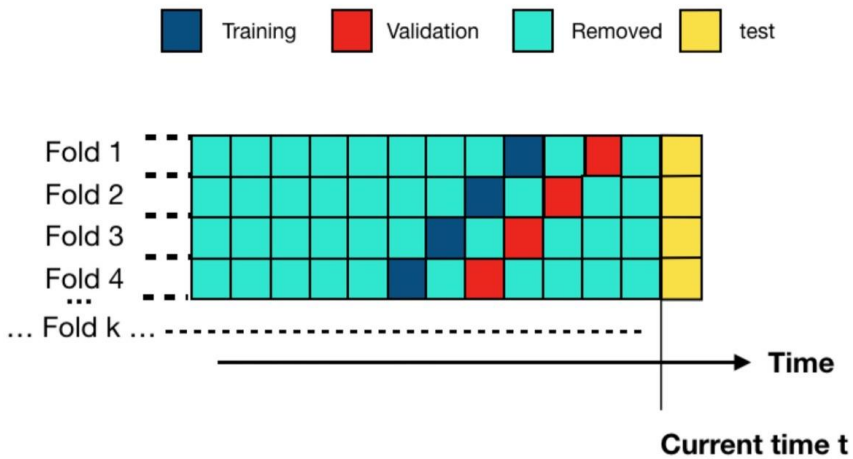


Представьте машину, которая может сжимать данные в маленькую коробку, а затем расширять их обратно в их исходный размер. Автоэнкодер учится делать именно это с транзакциями. Он стремится точно скопировать каждую транзакцию, сначала сжимая ее в более маленькую, а потом расширяя обратно. Если при восстановлении возникают ошибки, это может указывать на то, что транзакция отличается от типичной и, возможно, является подозрительной. Мы можем определить, какие аспекты транзакции вносят наибольший вклад в возникновение ошибок при восстановлении.

Auto-Encoder + GAN

VALIDATION

В качестве валидационной стратегии будет использоваться Time-Series Validation, которая учитывает временные аспекты данных и обеспечивает более реалистичную оценку ее производительности .



При использовании Time-Series Validation датасет делится на K фолдов, где каждый фолд представляет собой свой временной промежуток данных. Каждый фолд содержит информацию о транзакциях, произошедших в определенный период времени. Количество фолдов определяется длительностью временного промежутка, за который берется набор данных. Например, если мы имеем данные о транзакциях за 12 месяцев, то можно разделить их на 12 фолдов, по одному на каждый месяц.

100 000 тран/сут -> 1.5 тран/сек в среднем
Исходя из этого даже тяжелая нейросетевая модель успеет
На случай пиковых нагрузок возможно применение Kafka, Kubernetes