**­Кирилл:**

**1 Слайд: (Титульный слайд)** Добрый день, уважаемые участники конференции! Меня зовут Рязанов Кирилл, мой коллега – Скуратов Александр мы студенты второго курса и тема нашего проекта: «Распознавание мошеннических операций в мобильных банковских системах с использованием алгоритмов машинного обучения, и информирование пользователей о рисках».

**Саня:**

**2 Слайд: (Цели, задачи работы)**

Цель нашего проекта описана на слайде. Там же приведены задачи, которые нам потребовалось решить, для достижения поставленной цели.

На одном из мастер-классов от банка ВТБ, участников опрашивали о качестве их мобильного приложения и платежной системы в общем. В процессе дискуссии обсуждалось тема возможных улучшений. Нас заинтересовал данный вопрос и мы провели своё собственное исследование.

**Кирилл:**

**3 Слайд: (График хищений и возврата за 3 года)**

Как показал опрос, проведенный ВТБ среди клиентов, свыше 80% россиян сталкивались с мошенниками, которые пытались украсть деньги с их банковского счёта. Статистика за последние 3 года представлена на слайде. Можно заметить закономерность: уровень возврата средств падает, на фоне роста объёма хищений.

**Саня:**

**4 Слайд: (Причины роста объёма хищений?)**

По данный ЦБ уровень социальной инженерии в контексте информационной безопасности в России продолжает расти. Это связанно с низкой киберграмотностью граждан. Именно низкая осведомлённость и неосторожность пользователей заняла первое место с показателем 29% из всех рисков, с которыми люди сталкиваются при использовании мобильных устройств для совершения транзакций.

**Кирилл:**

**5 Слайд: (Почему падает уровень возврата хищений?)**

Мошенники используют новые схемы для обмана пользователей банковских систем, которые в большинстве случаев даже не догадываются что их обманули. В следствии этого, люди обращаются в банк слишком поздно или не обращаются вовсе.

Так же имеет место сама форма мошенничества, при которой пользователь в большей степени причастен к краже его средств, что осложняет распознавание мошеннических транзакций.

На основе этого и выражается столь малый процент возвращенных средств пользователям банковских систем.

Как следствие уровень доверия населения к дистанционным платежным сервисам падает.

**Саня:**

**6 Слайд: (Постановка задачи)**

Машинное обучение - это область Искусственного интеллекта, стоящая на пересечении математической статистики и компьютерных наук. Оно позволяет машинам обучаться на основе данных, без явного программирования.

Постановка задачи машинного обучения представлена на слайде. В нашем случае задача сводится к бинарной классификации: нужно предсказать класс транзакции 1 или 0, в зависимости от того, мошенническая транзакция или нет.

**Кирилл:**

**7 Слайд: (Скриншот датасета)**

Для машинного обучения нужен набор данных или датасет. Датасет который используем мы есть в свободном доступе, он состоит из трёхсот тысяч транзакций с тридцатью параметрами в каждой. Для определения мошеннических транзакций ЦБ регламентирует такие параметры, как: частота и время переводов, сумма транзакции и количество адресатов. В нашем датасете большинство из этих параметров анонимизированы, что сделано для поддержания секретности данных пользователей.

Важной особенностью нашего набора данных, является дисбаланс классов. Из трёхсот тысяч транзакций, только 500 являются мошенническими. Далее будут представлены методы решения этой проблемы.

**Саня:**

**8 Слайд: (А теперь, анализ!)**

Для того чтобы научиться определять класс транзакции, нужно исследовать датасет, и только после этого выбирать алгоритм. Существует много методик для исследования данных, например, борьба с выбросами и аномалиями, кластерный и разведочный анализ, понижение размерности и т.д. Некоторые из них были использованы и в нашем случае, остальные представлены на слайде.

Для решения проблемы дисбаланса классов, самым эффективным оказался метод недостаточной выборки. Суть его в следующем: мы выбираем из нашего набора данных все мошеннические транзакции, и добавляем к ним такое же количество не мошеннических. Таким образом мы создаём новый датасет, в котором отношение объектов двух классов будет один к одному.

**Саня:**

**9 Слайд: (Алгоритмы)** Когда мы говорили про постановку задачи, было сказано, что мы решаем задачу бинарной классификации. Это правда, но не вся правда. За место того, чтоб предсказывать класс, 1 или 0, лучше сообщать вероятность транзакции принадлежать к классу мошеннических. Так пользователь будет проинформирован о рисках.

Исходя из этого, было решено выбирать вероятностный алгоритм. Мы, естественно, перебрали и кучу других алгоритмов, но самыми удачными оказались: Логистическая регрессия, наивный байесовский классификатор и Случайный лес. Информацию о них можно увидеть на слайде.

**10 Слайд: (Наш алгоритм)** Сравнив и перебрав параметры этих алгоритмов, лучшим оказался алгоритм под названием Случайный лес. Это, так называемый, ансамбль алгоритмов, в нашем случае, состоящий из ста решающих деревьев. Каждое дерево выдает свой прогноз для класса, а итоговым ответом модели будет отношение ответов одного класса, на всё количество ответов. Например, если 79 деревьев из ста, решили, что конкретная транзакция принадлежит к классу мошеннических, а оставшееся 21 дерево, отнесло эту же транзакцию к противоположному классу, то итоговая вероятность этой транзакции быть мошеннической, будет 79/100 или 79%.

Существует огромное количество различных метрик, для замера точности алгоритма. По одной из них, точность нашего алгоритма оказалась в районе 98%.

**Кирилл:**

**11 Слайд: (Описание структуры приложения)**

Для демонстрации определения мошеннических транзакций мы разработали приложение имитирующие мобильный банк. На слайдах описан и представлен его функционал.

**12 Слайд: (Описание работы приложения)**

Для реализации совершения транзакций мы случайным образом выбрали и сохранили 40 записей из нашего набора данных, исключив их из процесса обучения алгоритма.

При совершении транзакции одну из этих записей будем посылать алгоритму и принимать данные о вероятности принадлежности транзакции к мошенническим. Для пользователей эту информации будем отображать в истории и на графике.

Также, если риск мошеннической транзакции превысит 70% в приложении высветится уведомления, указывающее на необходимость обратиться в банк.

**Саня:**

**14 Слайд: (Итоги)** Подведём итог: нам удалось разработать систему распознавания мошеннических операций и уведомления о них пользователей, которая работает с достаточно высокой точностью. Данная система может помочь в решении проблем мошенничества и неблаговременного обращения его жертв в банк, для повышения процента возвращенных средств пользователям.

**15 Слайд: (Заключительный слайд)** Благодарим вас за внимание!