РАСПОЗНАВАНИЕ МОШЕННИЧЕСКИХ ТРАНЗАКЦИЙ В МОБИЛЬНЫХ БАНКОВСКИХ СИСТЕМАХ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ И ИНФОРМИРОВАНИЕ ПОЛЬЗОВАТЕЛЕЙ О РИСКАХ

А.Ю. Скуратов, К.В. Рязанов

Машинное обучение – область искусственного интеллекта, стоящая на пересечении математической статистики и компьютерных наук. На сегодняшний день, машинное обучение получило огромную популярность, за счёт недавних прорывов в области глубокого обучения [1]. Основное преимущество машинного обучения в том, что для него достаточно большого количества данных и правильно подобранного алгоритма, чтобы научиться предсказывать номинальную или континуальную величину. В машинном обучении не нужно в ручную программировать тысячи строк кода, чтобы отыскать существующую зависимость желаемого результата от входных данных. Существует десятки если не сотни алгоритмов, способных отыскивать зависимости в данных. Конечно то, что машинное обучение требует только данных, и не требует жёсткого задания логики программы, может быть как плюсом, так и минусом. Иногда бывает чрезвычайно трудно добыть данные, но часто это очень хорошо окупается. В настоящей работе предлагается использовать машинное обучение, для определения мошеннических транзакций с целью скорейшего информирования пользователей о возможных рисках. Используя набор данных из открытых источников, получилось достичь точности распознавания мошеннических транзакций в 93%.

*Ключевые слова: машинное обучение, мошеннические операции, мобильная разработка.*

**Введение**

На данный момент, машинное обучение используется банками, как средство определения подозрительных действий в личном кабинете пользователя. Но как можно заметить, в мобильных банковских приложения не реализована система распознавания и уведомления пользователей о мошеннических транзакциях. Именно этим и было принято заняться в настоящей работе. Было разработано мобильное приложение, использующие алгоритм машинного обучения, для визуализации того, как распознавание и информирование о мошеннических транзакциях могут выглядеть в реальном банковском приложении.

Актуальность данной темы не вызывает вопросов, достаточно лишь взглянуть на статистику. 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 рисунке 1. Можно заметить закономерность: уровень возврата средств падает, на фоне роста объёма хищений.

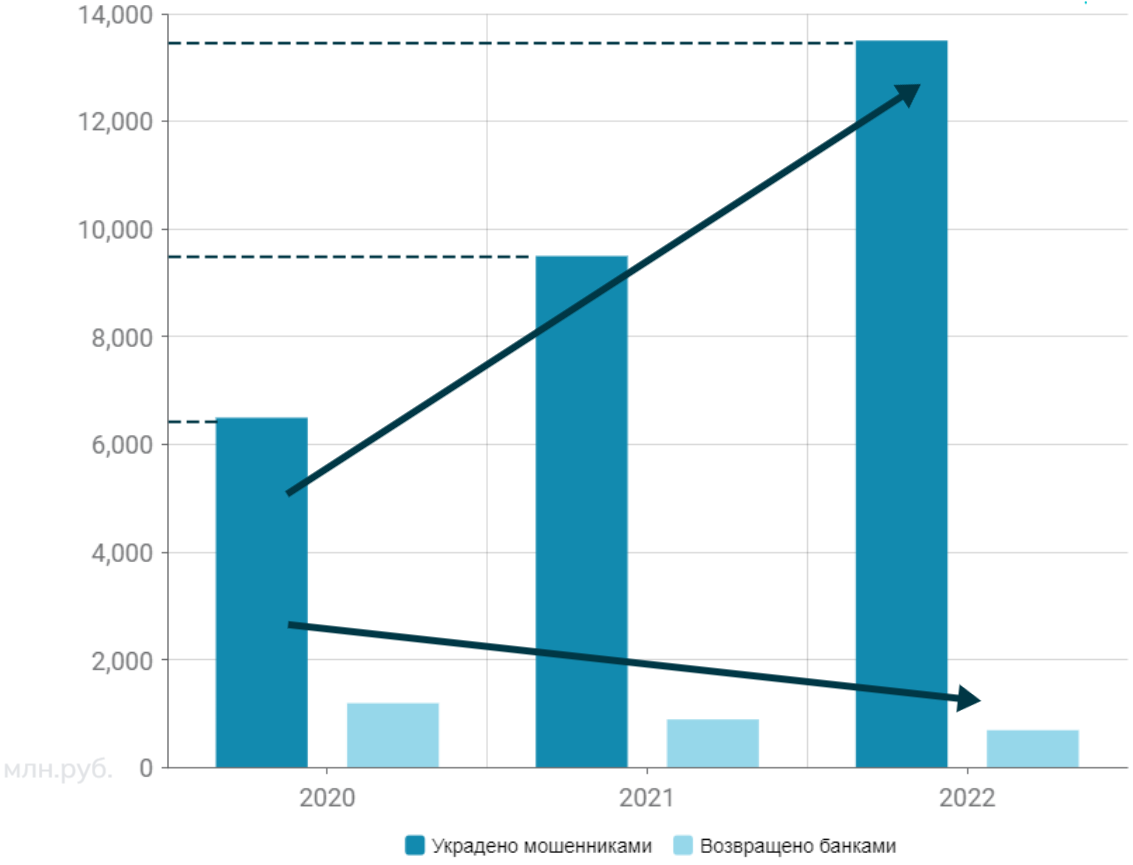


Рисунок 1 – Статистика хищений и возвратов за последние 3 года в млн. руб.

Возможные постановки задачи машинного обучения:

Дано:

Найти:

В нашем случае задача сводится к бинарной классификации: нужно предсказать класс транзакции 1 или 0, в зависимости от того, мошенническая транзакция или нет. Но используя вероятностный алгоритм, можно из бинарной классификации, получить предсказание вероятности класса мошеннических транзакций.

**Анализ данных**

Для того, чтобы можно было заниматься машинным обучением, нужно собрать данные. В свободном доступе был найден набор данных (датасет), который содержит в себе 300000 транзакций, 500 из которых – мошеннические, остальные нет. В глаза сразу бросается недостаток этого датасета – дисбаланс классов.

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 в настоящей работе.

Для решения проблемы дисбаланса классов, самым эффективным оказался метод под выборки (under-sampling) [2]. 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. Визуализация этого метода представлена на рисунке 2.

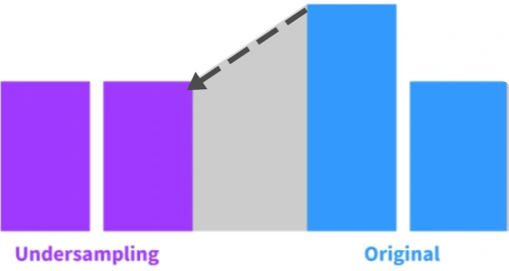


Рисунок 2 – визуализация метода under-sampling

**Выбор алгоритма машинного обучения**

Исходя из задачи предсказания вероятности класса мошеннической транзакции, было решено выбирать вероятностный алгоритм. В настоящей работе было решено использовать следующие алгоритмы: Логистическая регрессия, наивный байесовский классификатор, Случайный лес [3]. Основание этих алгоритмов описано на рисунке 3.

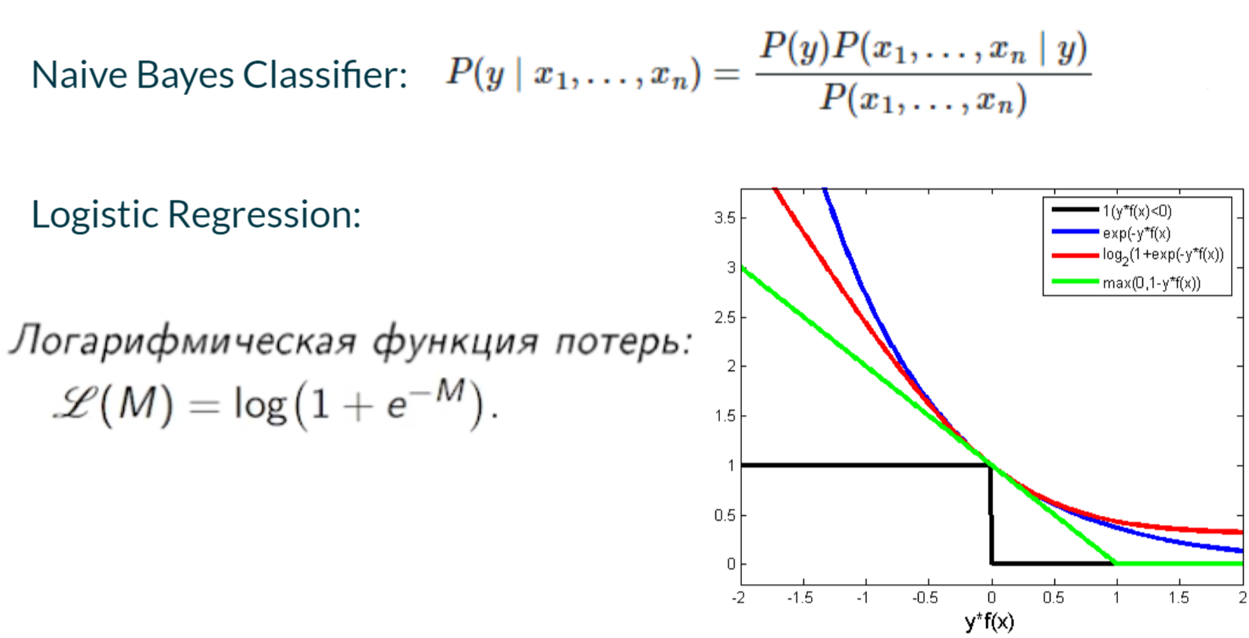


Рисунок 3 – математическое основание алгоритмов

Произведя перебор параметров 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, предсказали, что конкретная транзакция принадлежит к классу мошеннических, а оставшееся 21 дерево, отнесло эту же транзакцию к противоположному классу, то итоговая вероятность того, что эта транзакция мошенническая 79%.

Существует огромное количество различных метрик, для замера точности классификации, например: accuracy, precision, recall, ROC AUC, и др [4]. Используя метрику recall, точность выбранного алгоритма оказалась в районе 93%.

**Мобильное приложение**

Для демонстрации определения мошеннических транзакций было разработано приложение имитирующие мобильный банк. Для реализации совершения транзакций случайным образом было выбрано и сохранено 40 записей из набора данных, они же были исключены их из процесса обучения алгоритма. При совершении транзакции одна из этих записей посылается алгоритму хранящемуся на сервере, а после приложение принимает от алгоритма вероятность мошеннической транзакции. Для пользователей эта информация будет отображаться в истории и на графике. Также, если риск мошеннической транзакции превысит 70% в приложении высветится уведомления, указывающее на необходимость обратиться в банк.

**Анализ результатов и основные выводы**

Подведём итог: в настоящей работе 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. Итоговая точность системы составляет 93%.

**Список источников**

1. *Вейдман. С.* Глубокое обучение: легкая разработка проектов на Python. — СПб.: Питер, 2021. — 272 с.
2. *Шай Шалев-Шварц, Шай Бен-Давид.* Идеи машинного обучения: от теории к алгоритмам / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2019. – 436 с.
3. *Рашка С., Мирджалили В.* Python и машинное обучение: машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow 2, 3-е изд.
4. *Ясен С.* Машинное обучение для алгоритмической торговли на финансовых рынках. Практикум: Пер. с англ. — СПб.: БХВ-Петербург, 2020. — 560 с.

Скуратов Александр Юрьевич

студент СГУПС (630049, Новосибирск, ул. Дуси Ковальчук, 191), тел. (902) 997-22-58, e-mail: skuratov.1411@gmail.com.

Рязанов Кирилл Вадимович

студент СГУПС (630049, Новосибирск, ул. Дуси Ковальчук, 191), тел. (913) 899-66-40, e-mail: skuratov.1411@gmail.com.

**Recognition of Fraud Transactions in Mobile Banking Systems Using Machine Learning Algorithms And Informing Users About Risks**

1. **Skuratov, K. Ryazanov.**

Machine learning is an area of artificial intelligence at the intersection of mathematical statistics and computer science. Today, machine learning has gained immense popularity due to recent breakthroughs in the field of deep learning. The main advantage of machine learning is that it needs a large amount of data and a well-chosen algorithm to learn how to predict a nominal or continual value. In machine learning, you do not need to manually program thousands of lines of code to find the existing dependence of the desired result on the input data. There are dozens if not hundreds of algorithms that can find dependencies in data. Of course, the fact that machine learning requires only data, and does not require a rigid specification of program logic, can be both a plus and a minus. Sometimes it can be extremely difficult to get data, but often it pays off very well. In this paper, it is proposed to use machine learning to identify fraudulent transactions in order to inform users about possible risks as soon as possible. Using a set of data from open sources, it was possible to achieve an accuracy of recognizing fraudulent transactions of 93%.

*Keywords*: machine learning, fraudulent operations, mobile development.