Применение алгоритмов машинного обучения к задаче выявления мошенничества при использовании пластиковых карт

Т.А. Осипова, К.С. Зайцев, В.О. Биферт

***Аннотация*. Сегодня наблюдается значительный рост числа инцидентов при использовании пластиковых карт, а также разнообразие мошеннических методов, применяемых злоумышленниками. Представленная статья посвящена применению методов машинного обучения для противодействия мошенническим операциям с использованием пластиковых карт. Целью статьи является исследование эффективности различных моделей машинного обучения при анализе транзакций с пластиковыми картами для выявления различных типов мошенничества. В статье последовательно анализируются такие методы машинного обучения, как RandomForest, CatBoost, LogisticRegression, 2-х слойный обыкновенный перцептрон и многослойные перцептроны Румельхарта L-BFGS и SGD. Значительное внимание уделено процессу подготовки данных для участия в моделировании, который является итерационным и включает операции отбора таблиц, атрибутов, записей, конвертацию, очистку данных, фильтрацию и их объединение в нужном формате. В качестве исходного взят набор данных, созданный компанией Worldline и группой машинного обучения ULB по интеллектуальному анализу больших данных и обнаружению мошенничества. Проблема с несбалансированностью классов решалась при помощи ресэмплинга. При использовании метрик «время реакции» и «точность» лучшие результаты в экспериментах показали перцептроны Румельхарта L-BFGS и SGD.**

***Ключевые слова* – мошеннические транзакции, пластиковые карты, машинное обучение, нейронные сети, подготовка данных.**

I. Введение

Правонарушения в сфере банковской деятельности представляют собой серьёзную угрозу мировым экономическим интересам. С каждым годом темп убытков от таких правонарушений увеличивается, что может быть связано как с увеличением числа держателей пластиковых карт (в среднем у каждого пользователя имеется 2,3 карты), так и с увеличением числа интернет-магазинов, оплатой по безналичному расчёту, начислением заработной платы на карты. Благодаря интенсивному развитию ИТ появились новые возможности решать эту задачу.

Наблюдается ежегодный прирост доли безналичных операций в расходных операциях по картам и доли безналичного торгового оборота в общих расходах граждан. В 2019 г. доля безналичных операций в России составила 69,4 % в то время как 10 лет назад аналогичный показатель составлял 10,9 %.

Каждый пятый факт кражи в России связан с хищением денежных средств с банковского счета - прежде всего, звонки от фальшивой службы безопасности [1]. По итогам января – августа 2020 года выявлено 107200 хищений. Это вдвое больше прошлогодних показателей. Раскрывается лишь 8% таких преступлений.

Общая доля мошенничества с использованием социальной инженерии в 2019 году составила около 70% от всех случаев фрода с реальными потерями клиентов [1]. При этом средняя сумма для всех попыток мошенничества — около 14 000 руб., за год эта цифра сократилась на 13%. Из неумолимой статистики видно, что задача борьбы с мошенничеством такого типа является крайне актуальной.

В настоящей статье описывается разработанное базовое решение задачи выявления случаев мошенничества в банковской сфере с использованием алгоритмов машинного обучения.

II. Анализ технологий ИИ

Разграничение искусственного интеллекта (ИИ) и глубокого обучения неоднозначное [4]. Педро Домингос, профессор Вашингтонского университета, соглашается с мнением, что глубокое обучение выступает гипонимом по отношению к термину «машинное обучение», которое соответственно является гипонимом по отношению к ИИ [4]. П.Домингос утверждает: на практике области их применения пересекаются достаточно редко. Согласно альтернативному мнению - Хуго Ларочелле, профессора Шербрукского университета, данные концепты мало связаны между собой. Ларочелле замечает, что ИИ фокусируется на цели, а глубокое обучение — на необходимой для ML технологии [4].

Глубокое обучение можно рассмотреть как подкласс машинного обучения, его более сложную и функциональную разновидность. В свою очередь машинное обучение может быть определено как подполе ИИ, связанное с разработкой алгоритмов, которые могут помочь сделать предсказания. Соотношение ИИ, машинного и глубокого обучения представлено на рис. 1.

Алгоритмы глубокого обучения основаны на искусственных нейронных сетях со скрытыми слоями. Глубокие нейронные сети могут применяться в моделировании нелинейных

отношений как в контролируемых (использование данные с целью прогнозирования дальнейшего результата), так и в неконтролируемых (поиск новых закономерностей, аномалий в данных, кластеризация).

Рисунок 1: Соотношение ИИ, машинного и глубокого обучения [4].

III. ПОСТАНОВКА ЗАДАЧИ И ТРЕБОВАНИЯ К ПРОГРАММНОМУ РЕШЕНИЮ

Задачи, решаемые в статье условно можно разделить на две части:
- подготовка данных для анализа (1);
- обнаружение мошенничества при использовании пластиковых карт (2).

Первая задача решается путем преобразования исходных данных с помощью фильтров и ресэмплинга (resampling) данных. Вторая - применением алгоритмов машинного, в т.ч. глубокого обучения.

Решение разрабатывалось с учетом требований:
- предсказательной силы и устойчивости решения на заданном временном интервале;
- возможности реализации программного решения на C++ для обеспечения более высокой скорости работы на больших объемах данных;
- количество строк используемых данных должно быть статистически достоверным;
- процент некорректных данных в массиве должен быть минимальным;
- использования метрик точности (accuracy) и времени исполнения алгоритмов;
- кроссплатформенность, обеспечиваемая языком реализации Python.

Созданное программное решение по обнаружению мошеннических транзакций, включает в себя такие алгоритмы машинного обучения, как RandomForest, CatBoost, LogisticRegression, нейронные сети, и представляет схему выявления фрода с этапами:
- подготовка исходных данных;
- использование технологии ресэмплинга;
- применение метрик машинного обучения;
- выявление важных признаков (feature importance);
- расчёт потенциального экономического эффекта.

IV. ВЫБОР ИНСТРУМЕНТОВ

В работе использовалось несколько алгоритмов машинного обучения.

Random Forest – как это хорошо зарекомендовавшее себя семейство сложных базовых классификаторов с высоким быстродействием и Logistic Regression - как быстро обучаемая и подходящая для бинарной классификации [2]. CatBoost – как сильная модель, способная к обобщению на разнородных данных [17].

Среди методов глубокого обучения в работе использовались полноценный перцептрон (Keras RELU) в связи с удобством применения [3], многослойный перцептрон Румельхарта L-BFGS, как часто применяемый способ найти минимум целевой функции с использованием её значения и многослойный перцептрон Румельхарта SGD в связи с его эффективностью при отсутствии обучающих данных.

V. ПОДГОТОВКА ДАННЫХ ДЛЯ УЧАСТИЯ В МОДЕЛИРОВАНИИ

Исходный набор исследуемых данных [7] был создан и проанализирован в ходе исследовательского сотрудничества компании Worldline и группы машинного обучения ULB по интеллектуальному анализу больших данных и обнаружению мошенничества. Набор включает в себя транзакции, совершенные европейскими держателями кредитных карт в сентябре 2013 года [7]. Представлены транзакции, произошедшие за 2 дня. Выявлено 492 мошенничества из 284 807 транзакций. Набор данных включает числовые входные переменные, являющиеся результатом преобразования метода главных компонент (МГК). Исходные данные не предоставлены, так как данные обезличены на стороне платежной системы. Существует множество клиентских данных, которые в соответствии с требованиями банков и законами о персональных данных вообще нельзя использовать вне организации, собравшей их. Целью обезличивания является подготовка данных, максимально похожих на реальные, хранящиеся в продуктивных базах.

Входной набор данных включает в себя 28 признаков. Характеристики V1, V2,… V28 являются основными компонентами, полученными с помощью метода главных компонент (principal component analysis, PCA). Этот метод является одним из основных способов снижения размерности данных с потерей наименьшего количества информации [7]. Единственными не преобразованными с помощью него функциями являются параметры «Time» и «Cost». «Time» содержит время в секундах, прошедшие между текущей и первой транзакциями в наборе данных. Переменная «Cost» используется для обучения с учетом реальных сумм и для расчета экономического эффекта. Переменная «Class»

является целевой и указывает, была ли транзакция мошеннической (значение 1) и 0 в противном случае. Таким образом, данные для использования алгоритмов машинного обучения являются размеченными, но требующими определенной подготовки.

Подготовка данных является значимым, итеративным, трудозатратным процессом, занимающим до 80% времени решения задачи и содержащая этапы генерации признаков; выборки, очистки, интеграции и форматирования данных.

Целью этого процесса является получение набора признаков, который далее будет применяться для моделирования в процессе решения задачи. Исходные данные обычно разнородны с разными форматами. Поэтому процесс подготовки данных является итерационным и включает операции отбора таблиц, атрибутов, записей, конвертацию, очистку данных, фильтрацию и их объединение в нужном формате.

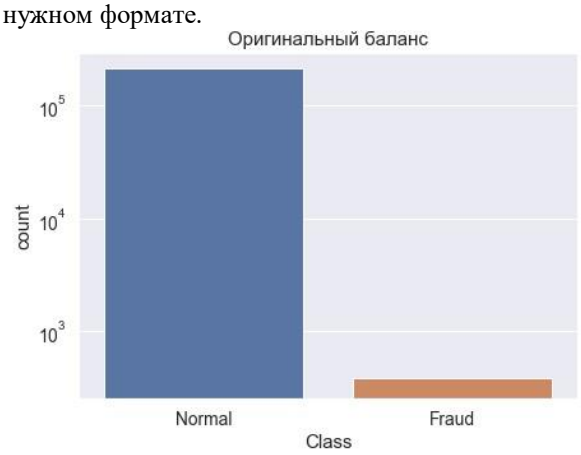

Рисунок 2: Баланс мошеннических и нормальных транзакций до ресэмплинга.

Рисунок 3: Баланс мошеннических и нормальных транзакций после ресэмплинга алгоритмом SMOTE.

С целью обеспечения баланса между обнаружением мошеннических и «нормальными» транзакций в исследовании были использованы алгоритмы ресэмплинга данных, т.е. генерации нового датасета - выборки из оригинального. Это делается по причине того, что оригинальная модель обладает явным дисбалансом в сторону «не мошенничество» и лишь малая часть содержит выявленные мошеннические транзакции, поэтому её использование может привести к переобучению.

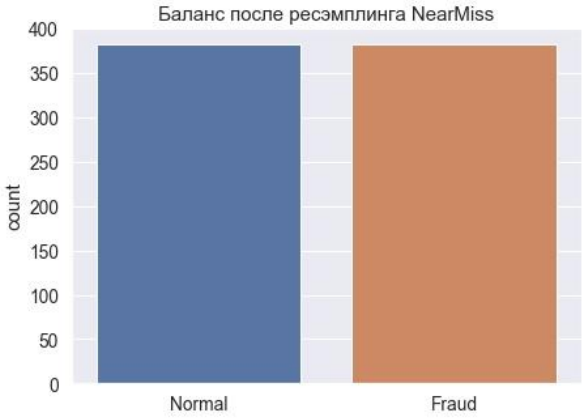

Рисунок 4: Баланс мошеннических и нормальных транзакций после ресэмплинга алгоритмом NearMiss.

Это показано при обработке методом LogisticRegression. Далее данные были преобразованы с помощью ресэмплинга. Для этого было использовано добавление (метод SMOTE) и изъятие (метод NearMiss) из выборки случайных семплов. Такие действия при анализе данных используются для корректировки распределения данных классов в наборе.

Ресэмплирована была только тренировочная часть выборки, тестовая часть была оставлена без изменений. SMOTE считается одним из наиболее часто используемых методов для решения проблемы дисбаланса и генерирует синтетические обучающие случайного исключения примеров большинства классов. Когда экземпляры двух разных классов очень близки друг к другу, мы удаляем экземпляры мажоритарного класса, чтобы увеличить пробелы между двумя классами. Это помогает в процессе классификации [22].

На рисунках 2-4 приведены графики, иллюстрирующие изменение баланса «фрод - не фрод».

Изначально преобладало количество «не фрода». Использованные методы ресэмплинга выровняли транзакций по типам, и позволили избежать проблемы несбалансированных классов.

В табл. 1 сравниваются количество строк в массиве данных в оригинальном наборе и после применения алгоритмов ресэмплинга.

Таблица 1: Размер тренировочных данных в исходном наборе и после работы алгоритмов.

	Оригинальный датасет	*SMOTE*	*NearMiss*
Количество строк	216837	432910	764

Для обеспечения быстроты работы алгоритмов и с целью избегания переобучения выбран алгоритм ресэмплинга NearMiss.

VI. ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ

Основной метрикой качества решения была выбрана точность (accuracy), по следующим причинам:
- является одной из распространённых метрик оценки классификатора;
- показывает количество правильно проставленных меток класса (истинно положительных и истинно отрицательных) от общего количества данных и вычисляется по формуле (1)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Эта метрика не учитывает соотношения ложных срабатываний.

VI. ОТБОР ЗНАЧИМЫХ ПРИЗНАКОВ

С помощью алгоритмов машинного обучения были выявлены наиболее значимые для программного решения признаки. Графики на рисунках 5-7 демонстрируют предполагаемый отбор признаков разными алгоритмами.

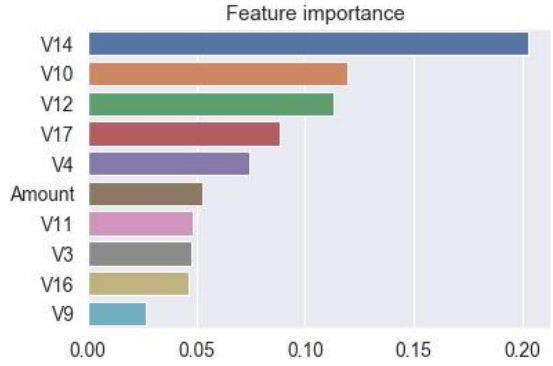

Рисунок 5: Отбор алгоритмом Random Forest.

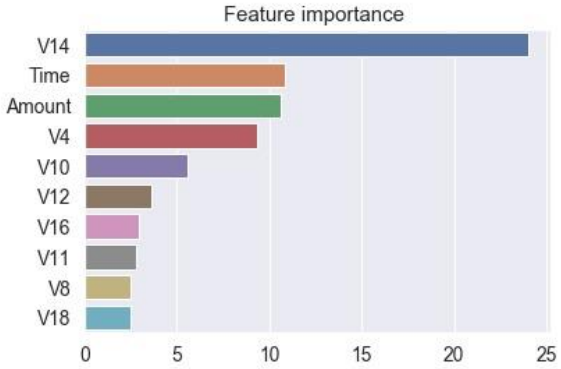

Рисунок 6: Отбор алгоритмом CatBoost.

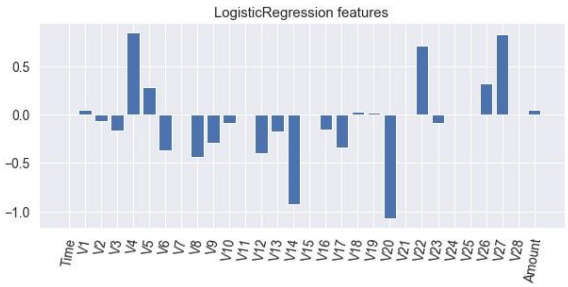

Рисунок 7: Отбор алгоритмом Logistic Regression.

Согласно алгоритмам машинного обучения, наибольшую значимость несут такие признаки, как V14 и V10 (выявляет алгоритм “Случайный лес”); V4, Time, Amount (выявляет алгоритм CatBoost);V4, V22, V27 (выявляет логистическая регрессия). V14, V4, V12 можно назвать весомыми признаками, так как их выделяют 2 алгоритма.

VII. ОТБОР ПРИЗНАКОВ С ПОМОЩЬЮ АЛГОРИТМОВ ГЛУБОКОГО ОБУЧЕНИЯ

Результаты работы методов глубокого обучения представлены на рисунках 8-10 с помощью библиотеки Lime, специализирующейся на интерпетации моделей [19].

На рисунках 8-10 левая часть показывает вероятности предсказанная для классов «0» (нормальные транзакции) и «1» (мошеннические транзакции). Средняя часть выводит 8 наиболее важных признаков. Атрибуты оранжевого цвета принадлежат классу «1», атрибуты синего цвета поддерживают класс «0». В горизонтальных столбцах числа с плавающей точкой демонстрируют относительную важность признаков. Правая часть включает фактические значения для переменных.

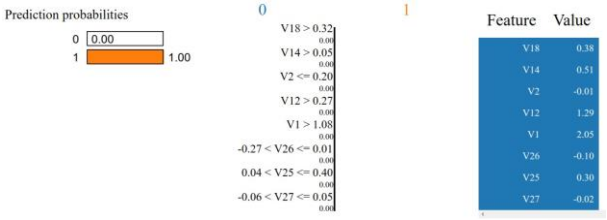

Рисунок 8: Отбор признаков стандартным перцептроном.

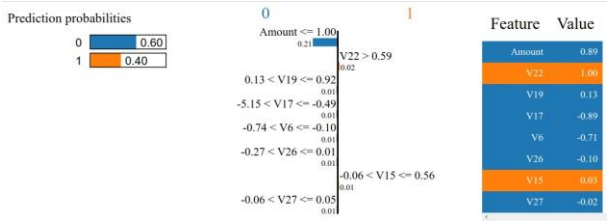

Рисунок 9: Отбор признаков перцептрона Румельхарта L-BFGS.

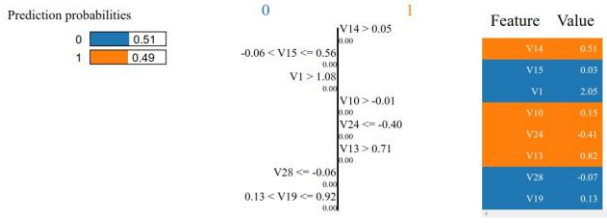

Рисунок 10: Отбор признаков перцептроном Румельхарта SGD.

Исследование показало, что при использовании обыкновенного перцептрона с 2 слоями значимых признаков не обнаружено. Многослойный перцептрон Румельхарта L-BFGS выделяет признаки V2 и V15. Многослойный перцептрон Румельхарта SGD выделяет признаки V10, V13, V14, V24.

Т.к. алгоритмы машинного обучения «Случайный Лес» и CatBoost выделили признак V14 наравне с SGD, то обобщая по всем алгоритмам можно выделить V14 в качестве наиболее значимого признака.

VIII. ИССЛЕДОВАНИЕ РАЗЛИЧИЙ РЕЗУЛЬТАТОВ РАБОТЫ АЛГОРИТМОВ

В результате подробного изучения всех выделенных в работе алгоритмов были получены значения их ключевых метрик - Time и Accuracy (см. табл.2).

Таблица 2: Таблица сравнения алгоритмов.

Метод	*Elapsed Time*	*Accuracy*
RandomForest	38.5	0.8134
CatBoost	38.378	0.6888
Logistic Regression	0.146	0.746
Keras RELU	1.263	0.4686
MLP L-BFGS	0.282	0.5352
MLP SGD	0.41	0.9971

Как видно, нейронные сети отличаются лучшим быстродействием. Наибольшее время исполнения имеет алгоритм «RandomForest», наименьшее - многослойный перцептрон Румельхарта L-BFGS. При этом наиболее высокий показатель точности принадлежит многослойному перцептрону Румельхарта SGD, наименее высокий — полноценному перцептрону Keras RELU.

IX. РЕЗУЛЬТАТЫ ВЫПОЛНЕННОЙ РАБОТЫ

Интересно, что алгоритмы машинного и глубокого обучения со схожей успешностью могут быть применены при выявлении мошенничества с использованием пластиковых карт. Разница коснётся метрик, давших разные результаты, и необходимого времени исполнения. В качестве лучшего следует выбрать алгоритм с наибольшим значением метрики точности.

Дополнительно в рамках исследования рассчитан потенциальный экономический эффект от работы разработанного программного решения (рис. 11), путём подсчёта среднего чека транзакции фрода и процентного отношения суммы фрода к общей сумме транзакций к представленным данным («общая сумма фрода составляет 58316.25 условных единиц или 0.24 % от 24213483.83 условных единиц; средний чек фрода 121.23 условная единица»).

Рисунок 11 демонстрирует количество фрода по отношению к общему количеству транзакций. Логарифмическая шкала представления результатов выбрана с целью улучшения визуализации данных.

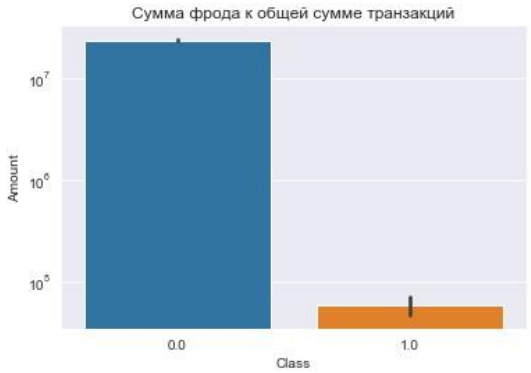

Рисунок 11: Сумма фрода к общей сумме транзакций в логарифмической шкале.

X. ЗАКЛЮЧЕНИЕ

В работе, выполненной в рамках выпускной квалификационной работы магистра, проводилось исследование проблемы мошенничества в использовании пластиковых карт, показавшее, что в настоящее время правонарушения в сфере банковской деятельности являются серьёзной угрозой для мировой экономики, и борьба с ними является очень актуальной.

Была разработана базовая модель обнаружения мошеннических транзакций, включающая в себя алгоритмы машинного обучения RandomForest, CatBoost, LogisticRegression, нейронные сети, и представляющая схему выявления фрода со следующими этапами: подготовка исходных данных, использование технологии ресэмплинга, применение метрик машинного обучения, выявление feature importance, расчёт потенциального экономического эффекта.

Был произведен анализ данных, используемых для выявления мошеннических транзакций, показавший, что для успешного решения поставленной задачи, используемые данные должны быть подготовлены в несколько этапов. Эти этапы - анализ исходных данных, проверка данных на наличие пропусков и ошибок и вывод статистики по ним, деление данных на тестовые и тренировочные, построение тепловой карты для тренировочных данных, ресэмплинг тренировочных данных.

Для реализации предложенного решения в среде Python Jupyter Notebook были разработаны программные модули с соответствующей оценкой качества, которая вычислялась с использованием нескольких метрик. В рамках работы проблема с несбалансированными классами решалась с помощью ресэмплинга. Разработанная программа в настоящее время проходит опытную эксплуатацию на стендах банка.

В качестве алгоритмов глубокого обучения, использованных в работе, были рассмотрены простая полносвязная сеть и полносвязный перцептрон Румельхарта с разными параметрами. Алгоритмы глубокого обучения демонстрируют более высокую скорость работы и более приемлемые значения метрики точности по

сравнению с алгоритмами RandomForest, CatBoost, LogisticRegression.

Оценка работы алгоритмов производилась с помощью метрики точности (accuracy). Выбор алгоритма машинного обучения остановлен на многослойном перцептроне Румельхарта SGD, так как его использование дает максимальное значение.

БЛАГОДАРНОСТИ

Авторы выражают благодарность Высшей инжиниринговой школе НИЯУ МИФИ за помощь в возможности опубликовать результаты выполненной работы.

БИБЛИОГРАФИЯ

[1] Tadviser [online resource] //Bank card fraud [website] URL: https://www.tadviser.ru/index.php//index.php/Статья:Мошенничество_с_банковскими_картами (Дата обращения 23.10.2020)

[2] HABR [электронный ресурс] // Алгоритмы машинного обучения [сайт] URL: https://habr.com/en/company/ods/blog/324402/#2-sluchaynyy-les (Дата обращения 6.12.2020)

[3] KPFU [электронный ресурс] // Искусственные нейронные сети и их приложения [сайт] URL: https://kpfu.ru/staff_files/F1493580427/NejronGafGal.pdf (Дата обращения 4.04.2021)

[4] Apmonitor [электронный ресурс] // Deep Learning [сайт] URL: https://apmonitor.com/do/index.php/Main/DeepLearning (Дата обращения 2.04.2021)

[5] Python-school [электронный ресурс] // TensorFlow vs PyTorch: что и когда выбирать для Machine Learning [сайт] URL: https://python-school.ru/tensorflow-vs-pytorch/ (Дата обращения 20.03.2021)

[6] Kaggle[электронный ресурс] // Kaggle.com [сайт] https://www.kaggle.com/dejavu23/titanic-survival-seaborn-and-ensembles (Дата обращения 17.10.2020)

[7] Kaggle [электронный ресурс] // Kaggle.com [сайт] https://www.kaggle.com/mlg-ulb/creditcardfraud/version/3 (Дата обращения 10.10.2020)

[8] Neural-university [электронный ресурс] // Что такое нейронные сети [сайт] URL: https://neural-university.ru/neural-networks-basics (Дата обращения 10.04.2021)

[9] "A literature survey on Machine Learning Algorithms", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.6, Issue 4, page no.471-474, April-2019, Available :http://www.jetir.org/papers/JETIR1904C77.pdf

[10] Machine learning approach to literature mining for the genetics of complex diseases, Jessica Schuster, Michael Superdock, Anthony Agudelo, Paul Stey, James Padbury, Indra Neil Sarkar, Alper Uzun, Author Notes, Database, Volume 2019, 2019, baz124. URL: https://doi.org/10.1093/database/baz124

[11] Fraud [электронный ресурс] // Fraud Detection with Machine Learning [сайт] URL: https://www.researchgate.net/project/Fraud-detection-with-machine-learning (Дата обращения 21.04.2021)

[12] Fontanka.ru [электронный ресурс] // Хищения с банковскими картами [сайт] URL: https://www.fontanka.ru/2020/11/09/69533506/ (Дата обращения 18.04.2021)

[13] Bazhenov [электронный ресурс] // Оценка классификатора (точность, полнота, F-мера) [сайт] URL: http://bazhenov.me/blog/2012/07/21/classification-performance-evaluation.html (Дата обращения 16.11.2020)

[14] Machinelearningmastery [электронный ресурс] // Работа с несбалансированными данными [сайт] URL: https://www.machinelearningmastery.ru/methods-for-dealing-with-imbalanced-data-5b761be45a18/ (Дата обращения 2.12.2020)

[15] docs.microsoft.com [электронный ресурс] // Предотвращение лжевзаимосвязи и несбалансированных данных посредством автоматизированного машинного обучения [сайт] URL: https://docs.microsoft.com/ru-ru/azure/machine-learning/concept-manage-ml-pitfalls (Дата обращения 2.04.2021)

[16] Reglament.net [электронный ресурс] // Нейронные сети в антифрод-моделировании [сайт] URL: http://www.reglament.net/bank/r/2018_2/get_article.htm?id=5615 (Дата обращения 12.04.2021)

[17] mql5 [электронный ресурс] Градиентный бустинг [сайт] URL: https://www.mql5.com/ru/articles/8642 (Дата обращения 7.12.2020)

[18] AI-news [электронный ресурс] // Большие данные / big data [сайт] URL: https://ai-news.ru/big_data.html (Дата обращения 26.05.2021)

[19] AI-news [электронный pelcypc] // Проблемы и ошибки машинного обучения, нейронных сетей / big data [сайт] URL: https://ai-news.ru/problemy_i_oshibki_mashinnogo_obucheniya.html (Дата обращения 26.05.2021)

[20] AI-news [электронный pelcypc] // Новости нейросетей: кто такой ПРовидец-7 и почему качество выборки важнее, чем её размер [сайт] URL: https://ai-news.ru/2021/04/novosti_nejrosetej_kto_takoj_providec_7_i_pochemu_kachestvo_vybork.html (Дата обращения 27.05.2021)

[21] Medium.com [электронный ресурс] // ReLU: Not a Differentiable Function [сайт] URL: https://medium.com/@kanchansarkar/relu-not-a-differentiable-function-why-used-in-gradient-based-optimization-7fef3a4cecec (Дата обращения 13.04.2021)

[22] Arefyevstudio [электронный ресурс] // Что такое ресэмплинг? [сайт] https://arefyevstudio.com/2019/01/11/chto-takoe-resempling/ (Дата обращения 20.03.2021)

Статья получена 12 июня 2021.

Осипова Татьяна Александровна, Национальный Исследовательский Ядерный Университет МИФИ, магистрант, osipova.tatyana98@gmail.com

Зайцев Константин Сергеевич, Национальный Исследовательский Ядерный Университет МИФИ, профессор, KSZajtsev@mephi.ru

Биферт Виталий Оттович, ПАО Сбербанк, главный технолог, VOBifert@sberbank.ru

The use of ML algorithms in the task of detecting fraud when using plastic cards

T.A. Osipova, K.S. Zaytsev, V.O. Bifert

***Abstract* - Today there is a significant increase in the number of incidents involving the use of plastic cards, as well as a variety of fraudulent methods used by cybercriminals. The presented article is devoted to the application of machine learning methods to counter fraudulent transactions using plastic cards. The aim of the article is to study the effectiveness of various machine learning models in the analysis of transactions with plastic cards to identify various types of fraud. The article sequentially analyzes such machine learning methods as RandomForest, CatBoost, LogisticRegression, 2-layer ordinary perceptron and Rumelhart's multilayer perceptrons L-BFGS and SGD. Considerable attention is paid to the process of preparing data for participation in modeling, which is iterative and includes operations for selecting tables, attributes, records, converting, cleaning data, filtering and combining them in the desired format. The dataset was taken from Worldline and the ULB Machine Learning Group for Big Data Mining and Fraud Detection. The problem with class imbalance was solved using resampling. Rumelhart's L-BFGS and SGD perceptrons showed the best results in experiments when using the metrics "response time" and "accuracy".**

***Keywords* – fraudulent transactions, plastic cards, machine learning, neural networks, data preparation.**

REFERENCES

[1] Tadviser [online resource] //Bank card fraud [website] URL: https://www.tadviser.ru/index.php//index.php/ Article: Bank Card Fraud (Date of request 23.10.2020)

[2] HABR [online resource] // Machine learning algorithms [website] URL: https://habr.com/en/company/ods/blog/324402/#2-sluchaynyy-les (Date of request 6.12.2020)

[3] KPFU [online resource] // Artificial neural networks and their applications [website] URL: https://kpfu.ru/staff_files/F1493580427/NejronGafGal.pdf (Date of request 4.04.2021)

[4] Apmonitor [online resource] // Deep Learning [website] URL: https://apmonitor.com/do/index.php/Main/DeepLearning (Date of request 2.04.2021)

[5] Python-school [online resource] // TensorFlow vs PyTorch: What and When to Choose for Machine Learning [website] URL: https://python-school.ru/tensorflow-vs-pytorch/ (Date of request 20.03.2021)

[6] Kaggle [online resource] // Kaggle.com [website] https://www.kaggle.com/dejavu23/titanic-survival-seaborn-and-ensembles (Date of request 17.10.2020)

[7] Kaggle [online resource] // Kaggle.com [website] https://www.kaggle.com/mlg-ulb/creditcardfraud/version/3 (Date of request 10.10.2020)

[8] Neural-university [online resource] // What are neural networks [website] URL: https://neural-university.ru/neural-networks-basics (Date of request 10.04.2021)

[9] "A literature survey on Machine Learning Algorithms", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.6, Issue 4, page no.471-474, April-2019, Available :http://www.jetir.org/papers/JETIR1904C77.pdf

[10] Machine learning approach to literature mining for the genetics of complex diseases, Jessica Schuster, Michael Superdock, Anthony Agudelo, Paul Stey, James Padbury, Indra Neil Sarkar, Alper Uzun, Author Notes, Database, Volume 2019, 2019, baz124. URL: https://doi.org/10.1093/database/baz124

[11] Fraud [online resource] // Fraud Detection with Machine Learning [website] URL: https://www.researchgate.net/project/Fraud-detection-with-machine-learning (Date of request 21.04.2021)

[12] Fontanka.ru [online resource] // Theft with bank cards [website] URL: https://www.fontanka.ru/2020/11/09/69533506/ (Date of request 18.04.2021)

[13] Bazhenov [online resource] // Classifier score (presicion, recall, F-score) [website] URL: http://bazhenov.me/blog/2012/07/21/classification-performance-evaluation.html (Date of request 16.11.2020)

[14] Machinelearningmastery [online resource] // Dealing with unbalanced data [website] URL: https://www.machinelearningmastery.ru/methods-for-dealing-with-imbalanced-data-5b761be45a18/ (Date of request 2.12.2020)

[15] docs.microsoft.com [online resource] // Prevent false relationships and imbalanced data through automated machine learning [website] URL: https://docs.microsoft.com/ru-ru/azure/machine-learning/concept-manage-ml-pitfalls (Date of request 2.04.2021)

[16] Reglament.net [online resource] // Neural networks in anti-fraud modeling [website] URL: http://www.reglament.net/bank/r/2018_2/get_article.htm?id=5615 (Date of request 12.04.2021)

[17] mql5 [online resource] Gradient boosting [website] URL: https://www.mql5.com/ru/articles/8642 (Date of request 7.12.2020)

[18] AI-news [online resource] // Big data [website] URL: https://ai-news.ru/big_data.html (Date of request 26.05.2021)

[19] AI-news [online resource] // Problems and errors of machine learning, neural networks / big data [website] URL: https://ai-news.ru/problemy_i_oshibki_mashinnogo_obucheniya.html (Date of request 26.05.2021)

[20] AI-news [online resource] // Neural network news: who is Prover-7 and why sample quality is more important than sample size [website] URL: https://ai-news.ru/2021/04/novosti_nejrosetej_kto_takoj_providec_7_i_pochemu_kachestvo_vybork.html (Date of request 27.05.2021)

[21] Medium.com [online resource] // ReLU: Not a Differentiable Function [website] URL: https://medium.com/@kanchansarkar/relu-not-a-differentiable-function-why-used-in-gradient-based-optimization-7fef3a4cecec (Date of request 13.04.2021)

[22] Arefyevstudio [online resource] // What is resampling? [website] https://arefyevstudio.com/2019/01/11/chto-takoe-resempling/ (Date of request 20.03.2021)