УДК 336.71 JEL G21; C55 DOI 10.26425/1816-4277-2022-10-193-204

Оценка технологических возможностей противодействия мошенническим практикам в банковском секторе

Бердышев Александр Валентинович

Канд. экон. наук, доц. департамента банковского дела и монетарного регулирования
ORCID: 0000-0002-0634-9321, e-mail: AVBerdyshev@fa.ru

Зархин Иван Евгеньевич

Студент, ORCID: 0000-0002-5335-3470, e-mail: zarhin2002@gmail.com

Катышева Арина Алексеевна

Студент, ORCID: 0000-0002-7421-6272, e-mail: arinakatysheva@gmail.com

Финансовый университет при Правительстве Российской Федерации, г. Москва, Россия

Аннотация

В связи с усилением западных санкций на российский банковский сектор резко выросло число злоумышленников, пользующихся доверием паникующих вкладчиков и нестабильной обстановкой на банковском рынке. В статье рассматриваются ключевые вопросы применения анализа больших данных как технологической основы противодействия мошенничеству в практической деятельности банков. Задачи такой борьбы – определить операции злоумышленников в потоке больших объемов статистической информации с наибольшей точностью и принять превентивные меры для минимизации ущерба. Целью статьи является оценка возможности использования банками технологии машинного обучения и разработка алгоритма выявления мошеннических операций на основе программирования. Особое внимание уделяется текущей экономической обстановке, ее влиянию на финансовую систему в целом, и особенно на переориентацию деятельности банковского сектора на борьбу с мошенническими действиями в условиях активизации фрод-деятельности.

Ключевые слова

Банки, транзакции, мошеннические операции, большие данные, искусственный интеллект, машинное обучение, антифрод-системы

Для цитирования: Бердышев А.В., Зархин И.Е., Катышева А.А. Оценка технологических возможностей противодействия мошенническим практикам в банковском секторе// Вестник университета. 2022. № 10. С. 193–204.

Assessment of technological capabilities to counter fraudulent practices in the banking sector

Aleksandr V. Berdyshev
Cand. Sci. (Econ.), Assoc. Prof. at the Banking and Monetary Regulation Department
ORCID: 0000-0002-0634-9321, e-mail: AVBerdyshev@fa.ru

Ivan E. Zarkhin
Student, ORCID: 0000-0002-5335-3470, e-mail: zarhin2002@gmail.com

Arina A. Katysheva
Student, ORCID: 0000-0002-7421-6272, e-mail: arinakatysheva@gmail.com

Financial University under the Government of the Russian Federation, Moscow, Russia

Abstract

In connection with the strengthening of Western sanctions on the Russian banking sector, the number of malefactors, who enjoy the confidence of panicking depositors and the unstable situation in the banking market, has increased dramatically. The article discusses the key issues of the application of big data analysis as a technological basis for countering fraud in the practical activities of banks. The objectives of such a struggle are to determine the operations of intruders in the flow of large volumes of statistical information with the greatest accuracy and to take preventive measures to minimize damage. The purpose of the article is to assess the possibility of using machine learning technology by banks and develop an algorithm for detecting fraudulent transactions based on programming. Particular attention is paid to the current economic environment, its impact on the financial system as a whole, and in particular, on the reorientation of the banking sector to combat fraud in the context of increased fraud activity.

Keywords

Banks, transactions, fraudulent transactions, big data, artificial intelligence, machine learning, anti-fraud systems

For citation: Berdyshev A.V., Zarkhin I.E., Katysheva A.A. (2022) Assessment of technological capabilities to counter fraudulent practices in the banking sector. *Vestnik universiteta*, no. 10, pp. 193–204.

ВВЕДЕНИЕ

В практической деятельности банк получает и хранит в своих базах огромный объем информации о клиенте: от уровня его дохода до суммы, локации и времени совершения каждой операций. Все доступные банку данные представляют собой состоящий из множества строк массив и формируют поведенческую картину клиента, анализировать которую крайне важно для того, чтобы банк мог составлять индивидуальные предложения и снижать собственные затраты и риски. Разумеется, вручную обработать огромный объем информации не представляется возможным, что предполагает применение машинного обучения в банковской деятельности. Однако в современных условиях далеко не каждый банк обладает развитой цифровой системой и может моментально отслеживать проводимые операции, что определяет необходимость разработки удобного и понятного алгоритма работы с большими данными (англ. big data).

Современная геополитическая ситуация серьезно повлияла на стабильность российской экономики, в том числе на работу банковского сектора. Обострившаяся ситуация сказалась и на психологическом состоянии россиян. Статистика показывает, что 70 % населения испытывают повышенную тревогу относительно текущей экономической ситуации. 60 % россиян заявляют, что санкции во многом повлияли на характер их потребления [1]. Следует отметить и тот факт, что повышение уровня цен при неизменном уровне зарплат является основным страхом граждан в 2022 г., по версии Всероссийского центра изучения общественного мнения [2]. Нестабильная макроэкономическая ситуация в сочетании с тревожными ожиданиями населения усиливают панические настроения, в особенности среди пользователей банковских услуг, что может спровоцировать массовый отток вкладов и способствовать обострению проблем банковского сектора с ликвидностью.

ПОСТАНОВКА ПРОБЛЕМЫ

В современных условиях банкам как никогда важно сохранить доверие клиентов. Из-за ужесточения коллективных санкций Запада сделать это становится все труднее. Основные системно значимые банки, на долю которых приходится почти 75 % всех сбережений россиян, попали в список CAPTA (англ. Correspondent Account or Payable-Through Account Sanctions – санкции в отношении корреспондентского счета или счета с оплатой через счет) или SDN (англ. Specially Designated Nationals and Blocked Persons – специально означенные граждане и заблокированные лица) – еще более жесткий пакет, введенный США в начале марта 2022 г. Многие не только государственные, но и частные банки подверглись жестким санкционным ограничениям. Так, например, введенными Великобританией в отношении Банка ВТБ санкциями была блокирована его дочерняя компания VTB Capital PLS, стоимость которой оценивается в 416 млн долл. США.

В складывающихся обстоятельствах банкам необходимо максимально сохранять лояльность клиентов и избежать банковской паники. В этой ситуации они все более активно прибегают к детальному анализу поведения клиентов [3]. Эта возможность обеспечивается на основе использования методов машинного обучения, способных за считанные секунды обработать объемный массив данных по проведенным транзакциям и вывести требуемые результаты.

Анализ больших данных в банковском секторе способствует снижению вероятности возникновения банковской паники посредством определения структуры расходов клиентов, выяснения транзакционных каналов, обнаружения и предотвращения мошенничества, а также оценки рисков и безопасности личных данных [4]. Анализ данных по клиентопотоку является распространенной банковской практикой, в то время как оценка финансовой безопасности – инновационное направление для России. Как указывается в отчете McKinsey, в США порядка 76 % крупнейших банков уже активно применяют технологии анализа больших данных с целью увеличения числа транзакций, повышения лояльности и улучшения качества взаимодействия с клиентами. Консалтинговая компания Gartner оценивает вовлеченность компаний по всему миру в развитие этих инноваций на уровне 34 % [5]. Примечательно, что четверть этого показателя формируется банковским сектором, что подтверждает актуальность анализа в банковской сфере. Действительно, эта технология способствует решению значительного количества существенных проблем и укреплению доверия клиентов к финансовому институту. Одной из наиболее актуальных задач современных российских банков, которая может быть решена с помощью больших данных, является противодействие финансовому мошенничеству [6].

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Турбулентная макроэкономическая ситуация и усиление панических настроений клиентов способствуют активизации мошенничества с банковскими картами и денежными переводами. Только за три месяца 2022 г. кража данных клиентов и количество незаконных транзакций возросли на 40 % [7]. Основными типами мошенничества в банковской сфере являются CNP-атаки (англ. Card Not Present Transaction – операции без присутствия карты), когда для совершения противоправного действия не требуется банковская карта, Scam–кража (англ. scam – мошенничество), основанная на одобрении клиентом незаконного платежа и кибер-мошенничество – наиболее распространенный в современном цифровом мире вид атаки [8; 9]. Разумеется, каждый тип требует отдельного изучения и индивидуальной разработки подходов к его предотвращению, однако существуют единые каналы и методы кражи данных и денежных средств, отследить которые возможно при помощи искуственного интеллекта.

Ручная обработка массивов данных представляет собой сложный процесс по причине ограниченного доступа к информации, ее большого объема, отсутствия единой формализации и медленных процессов обработки огромных баз данных. Именно поэтому большинство антифрод-систем построены на моделях машинного обучения, автоматически определяющих подозрительные транзакции по определенным патернам. Так, 82 % мошеннических действий может быть выявлено с помощью автоматизированных систем отслеживания транзакций, что превышает показатели других каналов. Для сравнения, системы ручного управления обеспечивают возможность отследить лишь 71 % противоправных транзакций; отдел аудита способен выявить только 58 % от их общего числа; на основе информации, полученной банком от третьих сторон (другие банки, мобильные операторы, инвестиционные платформы и т.п.), которым клиенты предоставляют свои данные, может быть выявлено 55 % мошеннических транзакций; по информации, поступающей на горячую линию банка, идентифицируется 68 % мошеннических операций [7].

Большинство розничных банков уже способны предотвращать наиболее распространенные формы мошенничества, отслеживая сомнительные транзакции по нетипичному IP-адресу (англ. Internet Protocol) клиента, времени или локации совершенной транзакции [10]. Основная задача современного банкинга – постоянное совершенствование систем противодействия финансовому мошенничеству. Далее, с целью оценки возможностей технологии больших данных в процессе противодействия мошенничеству в банковской сфере, будет предложена аналитическая модель антифрод-системы (англ. fraud – мошенничество), построенная авторами на языке программирования Python, которая на основе обработки загруженного массива данных, автоматически прогнозирует, какая транзакция будет с наибольшей вероятностью мошеннической.

Отметим, что принципы работы всех антифрод-систем однотипны. Чтобы валидировать транзакции, алгоритм сравнивает их в соответствии с определенными условиями и правилами, что позволяет отнести операцию к категории мошеннических. Если операция не проходит по определнному критерию, то процесс валидации завершается, а операция блокируется для последующего разбирательства [11].

В качестве входных данных будет использоваться информация о проведенных за сутки транзакциях условным банком «Омега» (табл. 1).

Таблица 1

База данных о транзакциях банка «Омега»

Машинный код:
In [190]: database
Out [190]:

Тип транзакции	**Число транзакций**	**Сумма транзакции**	**Валюта**	**Разные получатели**	**Единственный IP-адрес аккаунта**	**Комментарий к транзакции**	**Мошенничество**
Элитный бутик	1	110	руб.	Нет	Да	Нет	Нет
Электронная коммерция	6	30	евро	Да	Нет	Да	Нет
Электронная коммерция	1	15	долл. США	Нет	Нет	Да	Нет
Снятие наличных	3	185	долл. США	Нет	Да	Да	Да
Перевод	15	10 300	руб.	Да	Да	Нет	Нет
Перевод	2	19 800	руб.	Нет	Нет	Нет	Нет
Снятие наличных	4	60 50	руб.	Да	Да	Да	Да
Электронная коммерция	2	2 100	долл. США	Нет	Нет	Да	Да
Элитный бутик	1	1 350	руб.	Да	Да	Нет	Нет
Снятие наличных	1	5 100	руб.	Нет	Да	Да	Нет
Перевод	3	75 000	руб.	Нет	Нет	Нет	Да

Окончание табл. 1

Тип транзакции	Число транзакций	Сумма транзакции	Валюта	Разные получатели	Единственный IP-адрес аккаунта	Комментарий к транзакции	Мошенничество
Электронная коммерция	1	990	евро	Нет	Нет	Да	Да
Элитный бутик	1	75	евро	Нет	Нет	Нет	Нет
Перевод	2	110	долл. США	Да	Нет	Нет	Да
Элитный бутик	1	21 000	руб.	Нет	Да	Нет	Нет
Снятие наличных	4	65 000	руб.	Да	Да	Нет	Да
Элитный бутик	2	20	евро	Нет	Нет	Да	Да
Электронная коммерция	1	145	долл. США	Нет	Нет	Нет	Нет
Перевод	4	310	евро	Да	Нет	Да	Да
Снятие наличных	2	820	долл. США	Нет	Нет	Нет	Да
Элитный бутик	1	690	руб.	Нет	Да	Да	Нет
Электронная коммерция	3	85	евро	Нет	Нет	Нет	Да
Снятие наличных	4	9 700	руб.	Да	Да	Нет	Да
Перевод	2	18 500	руб.	Да	Да	Нет	Нет
Перевод	7	73 500	руб.	Нет	Да	Да	Нет

Составлено авторами по материалам источника [12]

База данных включает 25 транзакций, в которых уже содержится информация о том, совершена ли операция мошенником или нет. Каждая мошенническая атака включает определенные паттерны, характеризующие ее как подозрительную. Главная задача банка – определить эти признаки и обучить будущую модель находить их.

На начальном этапе данные о транзакциях разделяются на две выборки – обучающую, на которой модель будет учиться определять закономерности, и контрольную, с помощью которой модель проверяется на точность. Пусть в тренировочную выборку войдут первые 20 значений, а в тестовую – последние 5. Они будут содержаться в двух файлах формата Excel под названиями «database» для тренировочных данных и «control» – для контрольных. Важно, чтобы данные были представлены в формате латинского алфавита, иначе Python не сможет с ними работать.

После подготовки данных в Python импортируются основные статистические и математические пакеты и загружаются базы данных в программу. Прежде всего, будет использоваться статистическая библиотека numpy, библиотека pandas как надстройка к numpy с мощными инструментами прогнозирования и библиотека для визуализации matplotlib. Библиотека sklearn понадобится для будущего построения дерева решений.

На начальном этапе подготовки программы к работе импортируем необходимые для анализа библиотеки:

```
#импортируем необходимые библиотеки
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import os import pydotplus
import seaborn as sns
from sklearn.metrics import accuracy_score
from sklearn.tree import DecisionTreeClassifier
```

Называем переменные путем файла к нему:

database = pd.read_csv(r"C:/Users/UserName/Desktop/Python/database.csv", delimiter = ";")

control = pd.read_csv(r"C:/Users/UserName/Desktop/Python/control.csv", delimiter = ";")

На следующем этапе необходимо определить, какие признаки являются статистически значимыми. Для этого анализируется отдельно каждый параметр методами статистики. Попробуем выявить закономерность для числа транзакций (number of transactions). Существует гипотеза, что мошенники будут совершать большее число транзакций, так как банковские переводы – основная сфера их деятельности [13]. Найдем медиану от количества транзакций:

In [136]:

#Мошеннические операции характеризуются большим числом транзакций

database.groupby("Fraud")["Number of transactions"].median()

Out [136]:

Fraud

No 1.0

Yes 2.5

Name: Number of transactions, dtype: float64

Гипотеза подтвердилась: мошенники действительно совершают большее число транзакций.

Теперь проанализируем бинарные признаки. Начнем со столбца «Разные получатели» (Dif receivers), который показывает, разным ли получателям перечислена сумма. Как правило, мошенники совершают переводы на различные карты или тратят деньги в разных местах, чтобы замести цифровой след, поэтому у фрод-операции количество переводов нескольким лицам должно встречаться чаще. Гипотеза была подтверждена с помощью гистограммы (рис. 1).

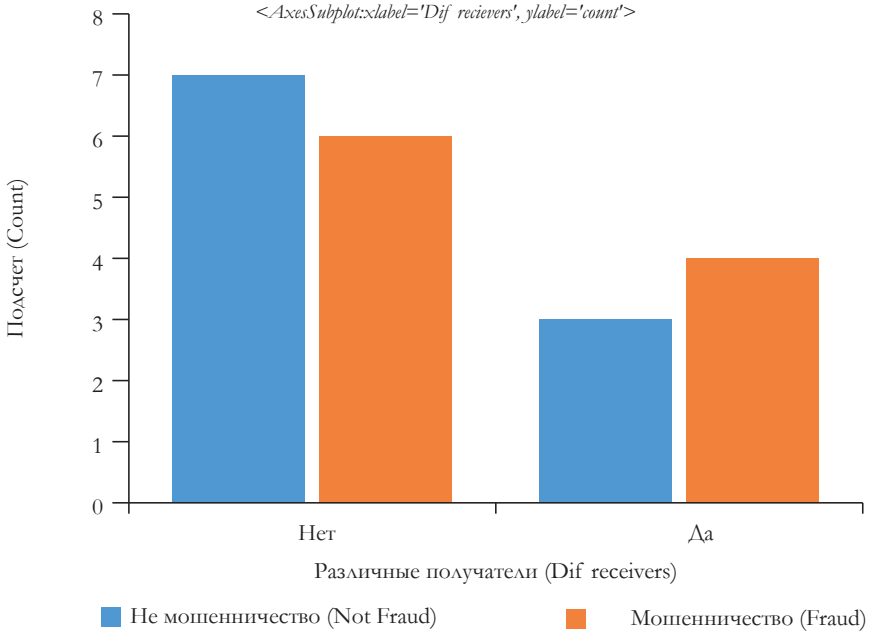
Машинный код:

In [140]:

```
#Анализ бинарных признаков
countplot, ax = plt.subplots(figsize = (10,5))
sns.countplot(x = "Dif recievers", hue = "Fraud", data = database)
```

Out [140]:

<AxesSubplot:xlabel='Dif recievers', ylabel='count'>



Составлено авторами по материалам исследования

Рис. 1. Гистограмма соотношения фрод-транзакции и количества получателей перевода

Основываясь на данных гистограммы, можно сделать вывод о том, что при подозрительной транзакции перевод действительно чаще осуществляется на несколько карт или счетов. Тем не менее, транзакция может являться корректной при переводе в пользу разных людей, например, если владелец карты совершил покупку в разных магазинах в период распродажи. Поэтому по одному паттерну нельзя делать какие-либо выводы. Это будет относиться и к последующим анализируемым параметрам.

Аналогично проверяются два других бинарных признака: «Единственный IP-адрес аккаунта» (One IP for acc) и «Комментарий к транзакции» (Without comment). «Единственный IP-адрес аккаунта» проверяет, со скольких IP-адресов был совершен вход в аккаунт владельца карты: с одного (Да) или с разных (Нет) (рис. 2). Предположение авторов заключается в том, что при подозрительной операции вход будет чаще совершен с компьютера мошенника или даже нескольких компьютеров группы мошенников, поэтому IP за день сменится минимум дважды.

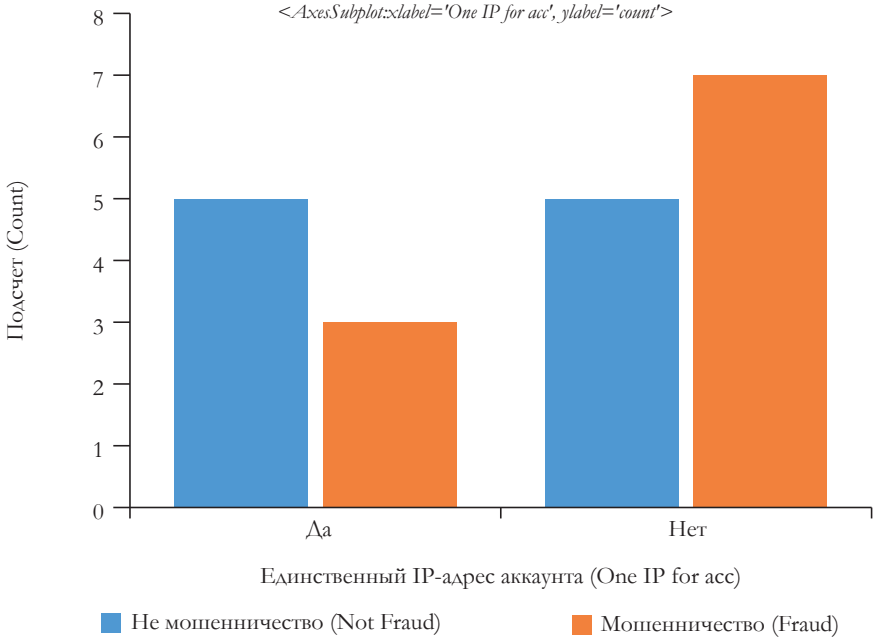
«Комментарий к транзакции» показывает, указал ли отправитель назначение транзакции. Гипотеза указывает на то, что мошенник маловероятно будет оставлять комментарии о платеже, он просто переведет деньги. Для данных признаков также построена гистограмма (рис. 3).

Основываясь на представленных данных, отметим, что чаще при фрод-операции отсутствует комментарий к платежу и вход в банковский аккаунт осуществляется с разных IP-адресов, что подтверждает справедливость гипотезы.

Далее проверим категориальные признаки, где результатов может быть несколько. Проанализируем виды транзакций (type of transaction). Всего за день были совершены четыре типа операций: – покупки в элитных бутиках (luxury store), например, в ювелирных или бренд-бутиках, электронная коммерция (e-commerce) – покупки на онлайн-площадках и в интернет-магазинах, перевод денег

Машинный код:

```
In [141]: #Анализ бинарных признаков
          fig, ax =plt.subplots(1,2)
          sns.countplot(x="Without comment", hue="Fraud", data=database, ax = ax[0])
          sns.countplot(x="One IP for acc", hue="Fraud", data=database, ax = ax[1])
Out [141]:
          <AxesSubplot:xlabel='One IP for acc', ylabel='count'>
```

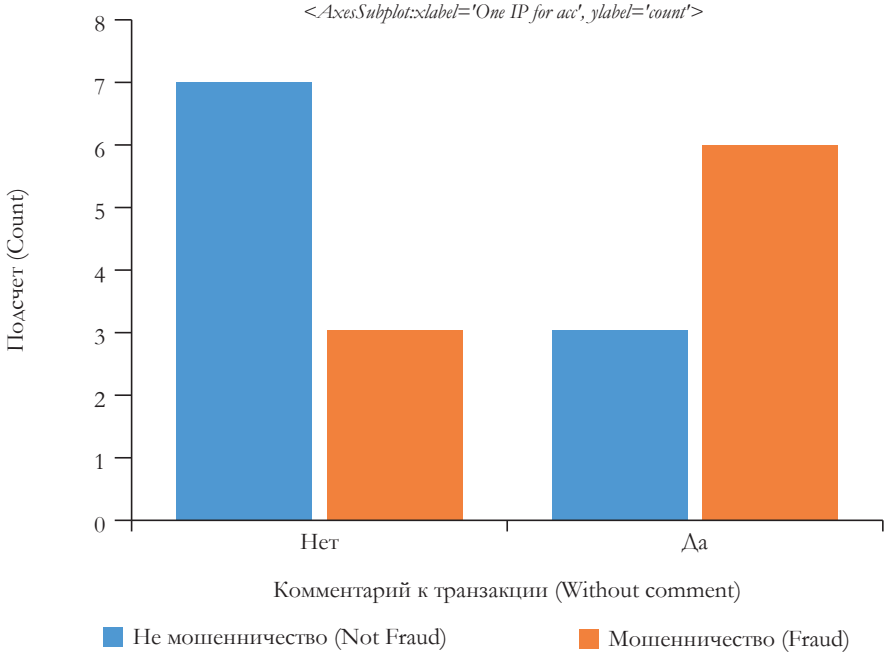


Составлено авторами по материалам исследования

Рис. 2. Гистограмма отнесения транзакции к мошеннической по признаку «Единственный IP-адрес аккаунта»

Машинный код:

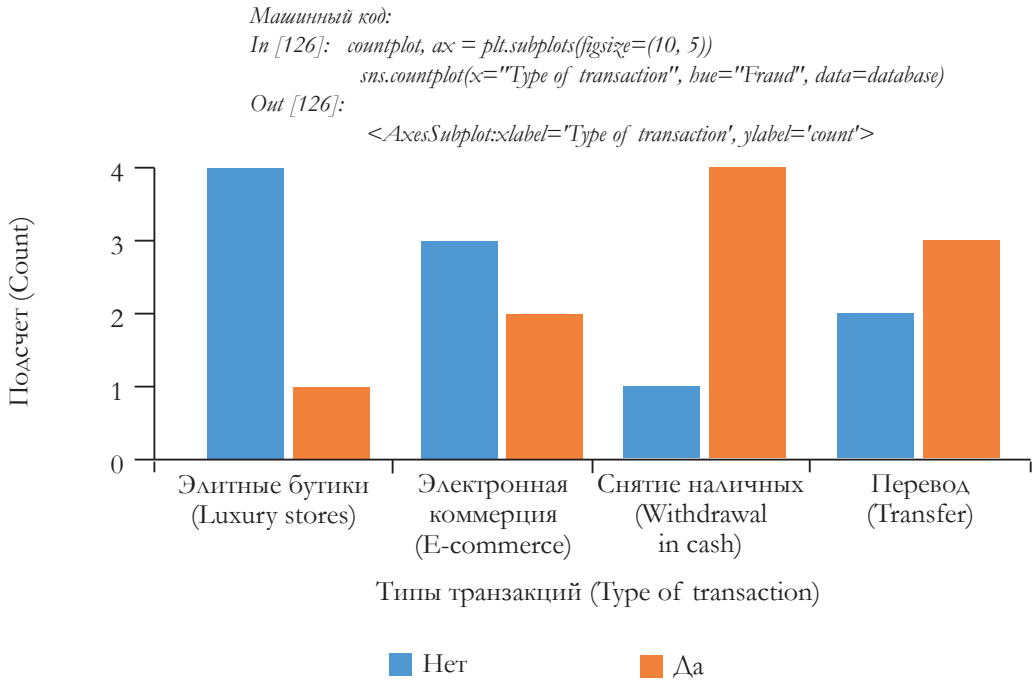
```
In [141]: #Анализ бинарных признаков
          fig, ax =plt.subplots(1,2)
          sns.countplot(x="Without comment", hue="Fraud", data=database, ax = ax[0])
          sns.countplot(x="One IP for acc", hue="Fraud", data=database, ax = ax[1])
Out [141]:
          <AxesSubplot:xlabel='One IP for acc', ylabel='count'>
```



Составлено авторами по материалам исследования

Рис. 3. Гистограмма отнесения транзакции к мошеннической по признаку «Комментарий к транзакции»

с карты на карту (transfer) и снятие наличных денег (withdrawal in cash). Посмотрим на диаграмме, как соотносятся вид транзакции и факт мошенничества (рис. 4).

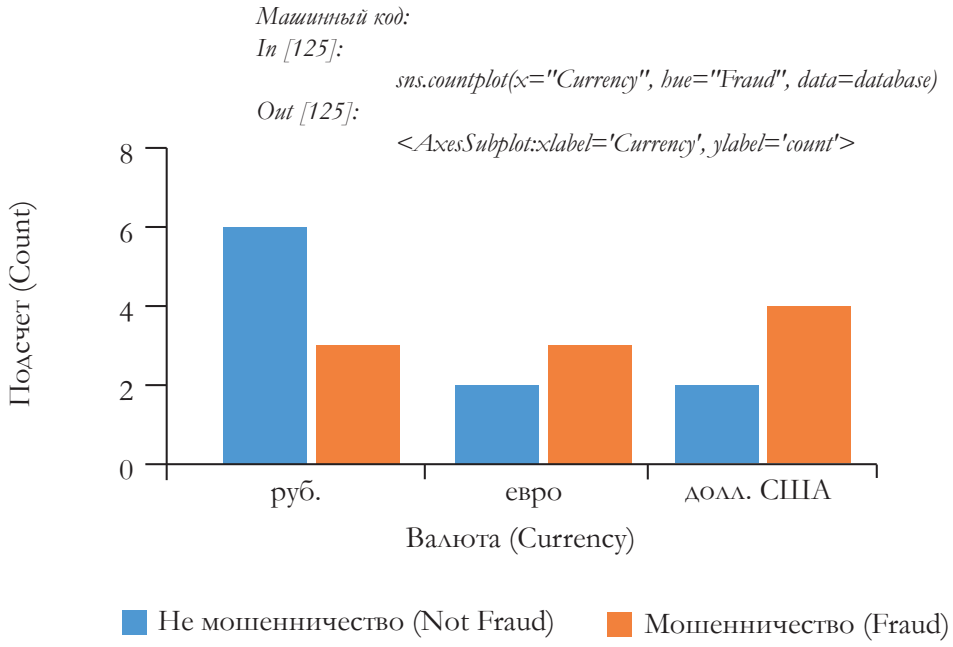


Составлено авторами по материалам исследования

Рис. 4. Гистограммы отнесения транзакции к мошеннической по типу операции

Представленные данные свидетельствуют о том, что наиболее часто мошенничество совершается в процессе снятия наличных. На втором месте находятся денежные переводы. Покупки в онлайн-магазинах, а также в элитных бутиках намного реже совершаются мошенниками, так как покупки за чужой счет могут вызвать подозрение со стороны банка и раскрыть преступника, поэтому злоумышленники будут чаще стремиться заполучить именно денежные средства.

Анализ транзакций по валюте платежа (currency) также может быть полезен в определении факта кражи денежных средств с карты. Обратим внимание на построенную на основе этого предположения гистограмму (рис. 5).



Составлено авторами по материалам исследования

Рис. 5. Гистограммы отнесения транзакции к мошеннической по валюте платежа

Представленные данные свидетельствуют о том, что мошенничество чаще совершается не в рублях, а в евро и долларах. Трансграничность этих валют позволяет быстро вывести средства за границу и разместить их, например, в иностранном банке или на офшорном счете, что существенно осложняет их возврат. Такое распределение валют указывает на реальный факт международных мошеннических сетей, которые занимаются незаконной деятельностью за пределами нашей страны, поэтому им выгоднее использовать мировые валюты.

В целях дальнейшего анализа необходимо рассмотреть распределение суммы одной транзакции (Sum of one transaction) при совершении операции. Создадим новый столбец «Сумма одной транзакции», который мы получим, поделив сумму всех транзакций на их количество. Столбец «Сумма» (Sum), показывающий сумму всех транзакций для одной операции, удалим, так как он малоинформативен и впоследствии использоваться не будет (табл. 2).

Таблица 2

База данных транзакций с новой колонкой о сумме одной транзакции

Машинный код:
In [145]: #Добавим дополнительное поле "Sum of one transaction"
database.insert(3, "Sum of one transaction", database["Sum"] / database["Number of transactions"])
del database["Sum"]
database.head()
Out [145]:

Тип транзакции	Число транзакций	Сумма одной транзакции	Валюта	Разные получатели	Единственный IP-адрес аккаунта	Комментарий к транзакции	Мошенничество
Элитный бутик	1	110,000 000	руб.	Нет	Да	Нет	Нет
Электронная коммерция	6	5,000 000	евро	Да	Нет	Да	Нет
Электронная коммерция	1	15,000 000	долл. США	Нет	Нет	Да	Нет
Снятие наличных	3	61,6 666 667	долл. США	Нет	Да	Да	Да
Перевод	15	686,666 667	руб.	Да	Да	Нет	Нет

Составлено авторами по материалам исследования

Сумма одной транзакции может быть полезна при определении того, является ли операция подозрительной или нет. Как правило, мошенники стремятся совершать небольшие переводы, так как крупные транзакции привлекают внимание не только банка и клиента, но и налоговых органов.

Также размер транзакции может зависеть от валюты перевода. Эту зависимость, если она существует, нам необходимо найти. С целью определения разброса сумм для трех разных валют при мошеннических и правомерных операциях был запущен код и получены данные, представленные в табл. 3.

Таблица 3

Распределение суммы одной транзакции в зависимости от валюты при мошеннических и правомерных операциях

Машинный код:
In [147]:
_, axes = plt.subplots(nrows=1, ncols=3, figsize=(15, 5)) for money, curr in enumerate(["RUB", "DOL", "EUR"]):
ax = axes[money % 3]
sns.boxplot(x="Fraud", y="Sum of one transaction", data=database[database["Currency"] == curr], ax=ax)
ax.set_ylabel(curr)
Out [147]:

Валюта транзакции	Сумма правомерной операции	Сумма мошеннической операции
руб.	4 000	16 000
долл. США	90	210

Окончание табл. 3

Валюта транзакции	Сумма правомерной операции	Сумма мошеннической операции
евро	50	80

Составлено авторами по материалам исследования

Представленные данные свидетельствуют о том, что средняя сумма одной транзакции в рублях и долларах США значительно выше, чем сумма одного мошеннического перевода в евро. Причем заметно, что средняя сумма перевода в международных валютах меньше рублевой минимум в два раза. Обычно трансграничные платежи проверяются тщательнее, поэтому мошенники стремятся быть как можно более осторожными, что определяет существенное превышение рублевых мошеннических операций над валютными.

Проверив все факторы и определив, что каждый из них в той или иной степени может указывать на то, является ли операция мошеннической или нет, можем перейти к разработке модели. Для начала необходимо закодировать тренировочные данные. Отметим, что модели, анализирующие данные, могут работать только с числовыми значениями.

В задаче будет использоваться метод машинного обучения – построение дерева решений. Мы должны разбить обе базы данных (обучающую и тренировочную) на две части – целевое значение (target_database и target_control) и описательные значения (var_database и var_control), от которых зависит целевое. Целевые значения равны показателям мошенничества в первоначальных файлах (табл. 4).

Таблица 4

Часть новой базы данных транзакций

Машинный код:

```
In [198]:
to_code = database.columns.drop([«Number of transactions», «Sum of one transaction»]) for df in [database, control]:
    df[to_code]=df[to_code].apply(lambda column: pd.Categorical(column).codes)
var_database = database[database.columns.drop(«Fraud»)]
target_database = database[«Fraud»]
var_control = control[control.columns.drop(«Fraud»)]
target_control = control[«Fraud»]
var_database.head()
Out [198]:
```

Тип транзакции	Число транзакций	Сумма одной транзакции	Валюта	Разные получатели	Единственный IP-адрес аккаунта	Комментарий к транзакции
1	1	110,000 000	2	0	1	0
0	6	5,000 000	1	1	0	1
0	1	15,000 000	0	0	0	1
3	3	61,6 666 667	0	0	1	1
2	15	686,666 667	2	1	1	0

Составлено авторами по материалам исследования

На следующем этапе необходимо построить дерево решений, которое будет прогонять все значения по блокам условий. Если условие выполняется (или не выполняется), значение будет сдвинуто на блок ниже до тех пор, пока программа не дойдет до конца ветви и не получит итоговое значение. Дерево будет построено на основе целевых и вариативных значений тренировочной выборки. Затем по полученной древовидной структуре будут пропущены контрольные данные для проверки работы программы. В процессе построения дерева на основе использования функции семейства DecisionTreeClassifier из библиотеки sklearn был получен следующий результат:

```
In [200]:
decide = DecisionTreeClassifier(random_state=17)
decide.fit(var_database, target_database)
predictions = decide.predict(var_control)
```

print(predictions)
[0 1 1 0 0]

Построив дерево, мы пропустили через него данные из выборки var_control. Получили результаты в квадратных скобках, где 0 – «чистая операция», а 1 – «мошенничество». Ниже представлены результаты оценки результативности построенной модели.

In [201]:
*print(accuracy_score(target_control, predictions) * 100)*
print(predictions)
print(target_control.values)
100.0
[0 1 1 0 0]
[0 1 1 0 0]

Как видим, модель благополучно предсказала 100 % мошеннических операций. Отметим, что в реальности точность прогнозирования будет снижаться в условиях большего количества данных и неопределенности, но в целом антифрод-системы работают схожим образом. На практике данные автоматически обновляются на основе информации о новых транзакциях, поэтому каждая последующая версия модели будет «умнее» предыдущей, обучаясь на все новых и новых входных данных.

ЗАКЛЮЧЕНИЕ

Подводя итог, можно сделать вывод о том, что применение машинного обучения в банковском секторе призвано способствовать противодействию мошенническим практикам и повышению удовлетворенности клиентов качеством работы банков. Использование анализа больших данных на основе искусственного интеллекта обеспечивает возможность получения практических результатов. Примером использования данной инновации является создание антифрод-системы, позволяющей отслеживать каждую проводимую банком транзакцию, и на основе загруженных в модель паттернов определять, является ли конкретная транзакция сомнительной. Преимуществом разработанной авторами модели является ее относительная простота и понятность для неподготовленного с технической точки зрения персонала, а также подтвержденная на реальном примере точность работы и относительно низкие затраты на запуск и поддержку ее функционирования. В целом автоматизация монотонных проверок и мониторинга операций позволит банкам избежать связанных с человеческим фактором ошибок и обеспечит высвобождение финансовых и трудовых ресурсов.

Библиографический список

1. Мамиконян О. 70 % россиян испытывают тревогу из-за сложившейся социально-экономической ситуации. *Forbes*. Среда 16 марта 2022. https://www.forbes.ru/forbeslife/459195-70-rossian-ispytyvaut-trevogu-iz-za-slozivsejsa-social-no-ekonomiceskoj-situacii (дата обращения: 18.08.2022).
2. Калюков Е. ВЦИОМ выявил рост тревоги у россиян из-за экономического кризиса. *РБК*. Среда 22 апреля 2020. https://www.rbc.ru/society/22/04/2020/5ea02d1c9a794704b82df2aa (дата обращения: 19.08.2022).
3. Банки.ру. *Методы борьбы с банковской паникой*. https://www.banki.ru/wikibank/metodyi_borbyi_s_bankovskoy_panikoy/ (дата обращения: 25.08.2022).
4. Ostapchenya D. The role of Big data in Banking: How do modern Banks use Big Data? *Finextra*. Friday 11 June 2021. https://www.finextra.com/blogposting/20446/the-role-of-big-data-in-banking--how-do-modern-banks-use-big-data (accessed 25.08.2022).
5. Филина Ф. Big Data для банкира. *Ведомости*. Среда 25 октября 2017. https://www.vedomosti.ru/partner/articles/2017/10/23/739068-uznat-vse (дата обращения: 27.08.2022).
6. Бердышев А.В. Искусственный интеллект как технологическая основа развития банков. *Вестник университета*. 2018;(5):91–94. https://doi.org/10.26425/1816-4277-2018-5-91-94
7. KPMG. *Global Banking Fraud Survey*; May 2019. https://assets.kpmg/content/dam/kpmg/xx/pdf/2019/05/global-banking-fraud-survey.pdf (accessed 28.08.2022).
8. Гришина Е.А. Риски в платежных системах: мошеннические схемы в мире банковских карт. *Финансы и кредит*. 2018;24(6):1280–1291. https://doi.org/10.24891/fc.24.6.1280
9. Трифонов Д.А. Банковские карты в России: стоит ли пользоваться. *Финансовая экономика*. 2018;(9):287–291.

10. Лунина Е. Как устроен антифрод и почему с мошенниками так сложно бороться. *РБК Тренды*. Пятница 15 октября 2021. https://trends.rbc.ru/trends/industry/6167ff259a7947f4c6908e46 (дата обращения: 05.08.2022).
11. Ярцев А.Г. Алгоритмы оценки рисков и принятия решений по определению мошенничества в современных антифрод-системах. В кн. *Теплотехника и информатика в образовании, науке и производстве: Сборник докладов IX Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых (ТИМ'2021) с международным участием, Екатеринбург, 13–14 мая 2021 г*. Екатеринбург: УрФУ; 2021. С. 331–335.
12. Kaggle. *Fraud detection bank dataset 20K records binary*. https://www.kaggle.com/datasets/volodymyrgavrysh/fraud-detection-bank-dataset-20k-records-binary (accessed 10.08.2022).
13. Сенцова А.Ю., Тимергазин В.Э., Ильясова Р.И. Антифрод-система как инструмент предотвращения мошенничества. *Информационные технологии. Проблемы и решения*. 2021;4(17):101–107.

References

1. Mamikonyan O. 70 % of Russians are worried about the current socio-economic situation. *Forbes*. Wednesday 16 March 2022. https://www.forbes.ru/forbeslife/459195-70-rossian-ispytyvaut-trevogu-iz-za-slozivsejsa-social-no-ekonomiceskoj-situacii (accessed 18.08.2022).
2. Kalyukov Ye. VTsIOM has revealed an increase in anxiety among Russians due to the economic crisis. *RBC*. Wednesday 22 April 2020. https://www.rbc.ru/society/22/04/2020/5ea02d1c9a794704b82df2aa (accessed 19.08.2022).
3. Banki.ru *Methods for dealing with banking panic*. https://www.banki.ru/wikibank/metodyi_borbyi_s_bankovskoy_panikoy/ (accessed 25.08.2022).
4. Ostapchenya D. The role of Big data in Banking: How do modern Banks use Big Data? *Finextra*. Friday 11 June 2021. https://www.finextra.com/blogposting/20446/the-role-of-big-data-in-banking--how-do-modern-banks-use-big-data (accessed 25.08.2022).
5. Filina F. Big Data for a banker. *Vedomosti*. Wednesday 25 October 2017. https://www.vedomosti.ru/partner/articles/2017/10/23/739068-uznat-vse (accessed 27.08.2022).
6. Berdyshev A.V. Artificial intelligence as a technological basis of the development of banks. *Vestnik Universiteta*. 2018;(5):91–94. https://doi.org/10.26425/1816-4277-2018-5-91-94
7. KPMG. *Global Banking Fraud Survey*; May 2019. https://assets.kpmg/content/dam/kpmg/xx/pdf/2019/05/global-banking-fraud-survey.pdf (accessed 28.08.2022).
8. Grishina E.A. Risks in payment systems: fraudulent schemes in the world of bank cards. *Finance and credit*. 2018;24(6):1280–1291. https://doi.org/10.24891/fc.24.6.1280
9. Trifonov D.A. Bank cards in Russia: is it worth using. *Financial economics*. 2018;(9):287–291.
10. Lunina Ye. How antifraud works and why it is so difficult to deal with scammers. *RBC Trends*. Friday 15 October 2021. https://trends.rbc.ru/trends/industry/6167ff259a7947f4c6908e46 (accessed 05.08.2022).
11. Yartsev A.G. Algorithms for risk assessment and decision-making to detect fraud in modern anti-fraud systems. In: *Heat engineering and informatics in education, science and production: Proceedings of the IX All-Russian scientific and practical conference of students, graduate students and young scientists (TIM'2021) with international participation, Yekaterinburg, 13–14 May 2021*. Yekaterinburg: Ural Federal University; 2021. Pp. 331–335.
12. Kaggle. *Fraud detection bank dataset 20K records binary*. https://www.kaggle.com/datasets/volodymyrgavrysh/fraud-detection-bank-dataset-20k-records-binary (accessed 10.08.2022).
13. Sentsova A.Yu., Timergazin V.E., Ilyasova R.I. Anti-fraud system as a fraud prevention tool. *Information Technology. Problems and solutions*. 2021;4(17):101–107.