ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ОБНАРУЖЕНИЯ МОШЕННИЧЕСТВА С КРЕДИТНЫМИ КАРТАМИ

А. М. Пастушенко

Paltus@mail.ru;
Научный руководитель – Д. В. Щегрикович, кандидат физико-математических наук, доцент

Основная проблема, рассмотренная в данной работе – это интернет мошенничество с кредитными картами банков. В работе представлена разработка полностью интегрированной интеллектуальной системы для обнаружения мошенничества и обеспечения финансовой кибербезопасности. Определены подходящие алгоритмы машинного обучения для решения задачи распознавания мошенничества с банковскими карточками. Предложенная система обнаруживает до 80% мошеннических транзакций.

Ключевые слова: антифрод; полностью интегрированная интеллектуальная система; анализ данных; большие данные; машинное обучение; ансамблевые методы.

ВВЕДЕНИЕ

В постоянно развивающемся современном мире люди всё больше пользуются банковскими картами. Повсеместно в жизнь приходит безналичный расчет за услуги. Большинство транзакций происходит через интернет, но мало кто задумывается о том, как реализована система обеспечения безопасности. В статье представлены этапы разработки интеллектуальной системы для классификации интернет-транзакций на факт мошенничества для обеспечения второго барьера информационной безопасности, когда первый – физическое изъятие банковской карты – уже прорван.

Рассмотрены существующие системы обеспечения безопасности интернет-транзакций, которые используются банками – мониторинг транзакций, 3D-Secure, биометрия, экспертные системы анализа данных, фильтрация данных, блокчейн, правоохранительные органы [1]. Определена ниша, которую может занять разрабатываемая система.

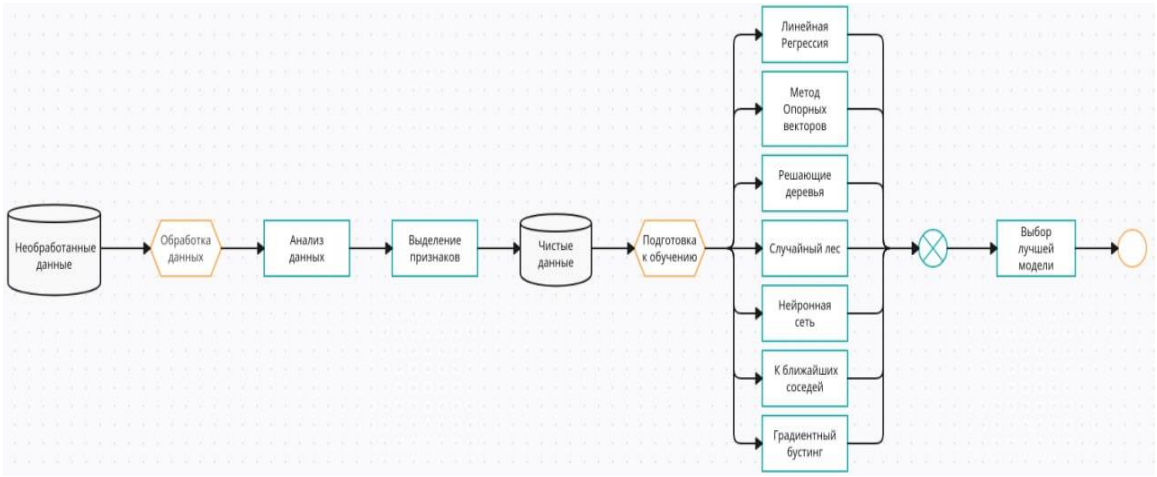
Система занимает нишу на пересечении 3D-Secure, экспертных систем анализа данных, фильтрации данных и мониторинга транзакций.

3D-Secure – дополнительный уровень подтверждения личности пользователя, так же, как и биометрия. Их плюсы в том, что они нетребовательны к вычислительным ресурсам. Анализ и фильтрация данных по сравнению с ними требуют больше вычислений, но позволяют выявлять сложные паттерны поведения данных, скрытые значения.

СОЗДАНИЕ СИСТЕМЫ

Разрабатываемая система должна удовлетворять перечисленным далее критериям. Обеспечение безопасности данных – хранение и обработка персональных данных в соответствии с действующими политиками безопасности. Высокое качество классификации – система должна иметь высокое качество классификации, чтобы ей было можно доверять и пользоваться ею. Простота вычислений – система должна отрабатывать достаточно быстро отрабатывать, чтобы пользователь меньше ожидал разрешения на совершение транзакции. Масштабируемость – удобство расширения системы на новые базы знаний [2].

Алгоритм работы системы представлен на рисунке ниже. Необработанные данные поступают в блок обработки, где они преобразуются и структурируются, производится их очистка. Затем над обработанными данными производится разведывательный анализ и выделение значащих признаков. Для упрощения моделирования выделенные по признакам данные записываются в хранилище. Следующим шагом является подготовка данных к моделированию, а затем само обучение алгоритмов.



Алгоритм системы

По итогу проведения разведывательного анализа данных были выделены следующие ключевые признаки транзакций, которые имеют наибольшее влияние на целевую переменную – факта мошенничества:

- Возраст.
- Сумма транзакции.
- Время проведения транзакции.

- Категория товара.
- Магазин.
- Население города держателя карты.
- Профессия держателя карты.

РЕЗУЛЬТАТЫ

Изучались такие алгоритмы, как линейная регрессия, метод опорных векторов, деревья решений, случайный лес, нейронная сеть прямого распространения, метод K-ближайших соседей, градиентный бустинг. Заключительным этапом разработки является выбор наилучшей модели и ее использование [3, 4].

Оценка обученности алгоритмов проводилась по следующим критериям: точность, полнота, их гармоническое среднее, качество и ROC-AUC-метрика. Были изучены методы борьбы с несбалансированностью данных такие, как SMOTE и ADASYN, undersampling. Oversampling. Были изучены следующие способы кодирования текстовых полей: Ordinal Encoding, Label Encoding, Count Encoding, One-Hote Encoding [5].

Рассмотрены алгоритмы и все комбинации данных методов, и лучший результат показал алгоритм случайного леса с методами предобработки Label Encoding и SMOTE при k = 5. Остальные алгоритмы мало подходят для решения поставленной задачи.

По итогу проведенных исследований можно сделать следующие выводы:

- Модель случайного леса имеет точность определения мошенничества 79%;
- Система корректно обрабатывает личные данные пользователей;
- Система имеет возможность горизонтальной масштабируемости;
- Система использует относительно нетребовательный алгоритм.

Данная система отличается от существующих тем, что ее легко внедрить в систему безопасности транзакций, как и 3D-Secure, но в то же время она является интеллектуальной с гибкой настройкой параметров и имеет возможность масштабирования.

В перспективе можно использовать вычисления на графических процессорах. Проводить более глубокую аналитику и выявлять мультиколлинеарные зависимости. Использовать Apache Spark для горизонтальной масштабируемости системы и многопотоковой обработки информации. Использовать Apache Airflow для приоретизации

выполнения задач. Целесообразным представляется использование облачной архитектуры для горизонтальной масштабируемости.

Библиографические ссылки

1. *Белянина Н. В., к.т.н., доцент, Кожин Е. В.,* аспирант. Информационная система определения мошенничества по платежным картам в режиме реального времени [Электронный ресурс] // Материалы публикации. Современная гуманитарная академия / Москва, 2018. URL: https://cyberleninka.ru/article/n/informatsionnaya-sistema-opredeleniya-moshennichestva-po-platezhnym-kartam-v-rezhime-realnogo-vremeni/viewer (дата обращения: 01.06.2023).
2. *Hastie, T., Tibshirani, R.,* & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. URL: https://hastie.su.domains/Papers/ESLII.pdf (дата обращения: 01.06.2023).
3. Яндекс. Учебник по машинному обучению [Электронный ресурс]. URL: https://academy.yandex.ru/handbook/ml (дата обращения: 01.06.2023).
4. *Swalin, A.* (2018). Choosing the Right Metric for Evaluating Machine Learning Models Part 2. [Электронный ресурс]. URL: https://www.kdnuggets.com/2018/06/right-metric-evaluating-machine-learning-models-2.html (дата обращения: 01.06.2023).
5. *Воронцов К. В.* Курс лекций «Машинное обучение». URL: http://www.machinelearning.ru/ (дата обращения: 01.06.2023).