112. ОБНАРУЖЕНИЕ И ОБЪЯСНЕНИЕ АНОМАЛИЙ В ТРАНЗАКЦИОННЫХ ДАННЫХ: СРАВНЕНИЕ МЕТОДОВ И ИНТЕРПРЕТИРУЕМОСТЬ РЕЗУЛЬТАТОВ

Фомин А.В., Ременчик Е.К., студенты группы 377901

Белорусский государственный университет информатики и радиоэлектроники

г. Минск, Республика Беларусь

Федосенко В.А.– кандидат технических наук, доцент

Аннотация. В работе выполнено сравнение подходов к обнаружению аномалий в транзакционных данных на наборах PaySim и Credit Card Fraud Detection. Исследованы supervised-модель XGBoost и unsupervised-модель Isolation Forest, а также интерпретация решений с помощью SHAP и permutation importance. Показано, что supervised-подход обеспечивает более высокое качество, однако содержательная интерпретируемость зависит не только от алгоритма, но и от природы признаков. Для PaySim дополнительно проведен anti-leakage эксперимент, позволивший отделить максимальное качество от практически предпочтительной модели.

Финансовые транзакции относятся к данным с высокой скоростью поступления, значительным дисбалансом классов и постоянным изменением шаблонов поведения. По этой причине задача обнаружения аномалий в таких данных требует одновременного учета качества классификации, устойчивости к редким событиям и возможности интерпретации результата [4; 6]. Особую ценность представляет сравнение методов, способных выявлять подозрительные операции как в режиме обучения с учителем, так и без разметки.

Целью работы является сравнение методов обнаружения аномалий в транзакционных данных и анализ интерпретируемости результатов. Для этого были использованы два открытых набора данных. PaySim содержит синтетические транзакции мобильных платежей и хорошо подходит для предметной интерпретации признаков [5]. Набор Credit Card Fraud Detection включает анонимизированные операции по банковским картам и широко используется как benchmark для оценки алгоритмов при сильном дисбалансе классов [7; 8].

Вычислительные эксперименты выполнялись в среде Kaggle Notebook. Для PaySim использовалось 6 362 620 транзакций, включая 8 213 мошеннических операций, а для Credit Card Fraud Detection – 284 807 транзакций, из которых 492 являлись мошенническими [5; 7]. Во всех экспериментах применялось хронологическое разбиение на обучающую, валидационную и тестовую выборки. Такой подход лучше имитирует реальный сценарий, при котором модель обучается на прошлых операциях и проверяется на более новых наблюдениях.

В качестве supervised-подхода была выбрана модель XGBoost, хорошо зарекомендовавшая себя на табличных данных и позволяющая учитывать сложные нелинейные зависимости [3]. В качестве unsupervised-базовой модели использовался Isolation Forest, основанный на идее изоляции аномальных наблюдений [2]. Для объяснения результатов применялись SHAP-значения, обеспечивающие локальную и глобальную интерпретацию предсказаний [1], а также permutation importance, измеряющий снижение качества модели после случайной перестановки значений признака.

Сначала для каждого датасета были обучены XGBoost и Isolation Forest, после чего порог принятия решения подбирался по валидационной выборке на основе максимизации F1-мера. Для PaySim дополнительно выполнен anti-leakage эксперимент: из модели были исключены признаки, непосредственно связанные с балансами после проведения транзакции. Это позволило разделить два сценария: модель максимального качества и модель, более предпочтительную с практической точки зрения.

Таблица 1 – Основные результаты на тестовых выборках

Набор данных	Модель	PR-AUC	Recall	F1
PaySim	XGBoost_Full	0,999409	0,997754	0,993294
PaySim	XGBoost_Restricted	0,973805	0,766218	0,860465
PaySim	Isolation Forest	0,549569	0,241517	0,386350
Credit Card Fraud Detection	XGBoost	0,757670	0,750000	0,795918
Credit Card Fraud Detection	Isolation Forest	0,040380	0,519231	0,116631

Основные результаты экспериментов приведены в таблице 1. На PaySim полная модель XGBoost показала почти идеальные значения PR-AUC = 0,999409 и F1 = 0,993294. Однако после исключения наиболее сильных post-transaction признаков качество снизилось до PR-AUC = 0,973805 и F1 = 0,860465. Несмотря на снижение, restricted-модель сохранила высокий уровень качества и была выбрана как practically preferred вариант. Подбор гиперпараметров для restricted-модели был выполнен ограниченным перебором, однако на тестовой выборке он не дал улучшения: tuned-вариант не превзошел базовую restricted-модель по PR-AUC и F1.

На наборе Credit Card Fraud Detection модель XGBoost достигла PR-AUC = 0,757670 и F1 = 0,795918, тогда как Isolation Forest показал PR-AUC = 0,040380 и F1 = 0,116631. Следовательно, наличие размеченных данных дает supervised-подходу заметное преимущество. При этом даже на более сложном benchmark-наборе модель XGBoost обнаруживала 39 из 52 мошеннических транзакций тестовой выборки при семи ложных срабатываниях, что можно считать хорошим практическим результатом.

Сравнение по PR-AUC и F1 для ключевых моделей приведено на рисунках 1 и 2. Графики подтверждают два устойчивых эффекта. Во-первых, на обоих наборах данных supervised-модели значительно превосходят unsupervised baseline. Во-вторых, различие между полной и restricted-моделью на PaySim показывает, что часть максимального качества обеспечивается сверхсильными признаками, отражающими состояние системы после транзакции. Это важно учитывать при построении решений для раннего обнаружения подозрительных операций.

Для наглядного сопоставления качества по PR-AUC на рисунке 1 представлено сравнение ключевых моделей.

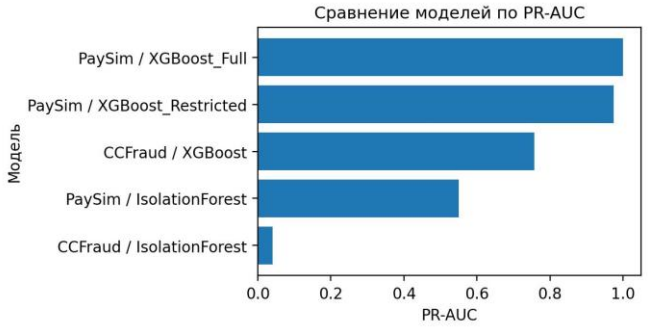

Рисунок 1 – Сравнение моделей по PR-AUC

Аналогичное сравнение по F1-мере приведено на рисунке 2.

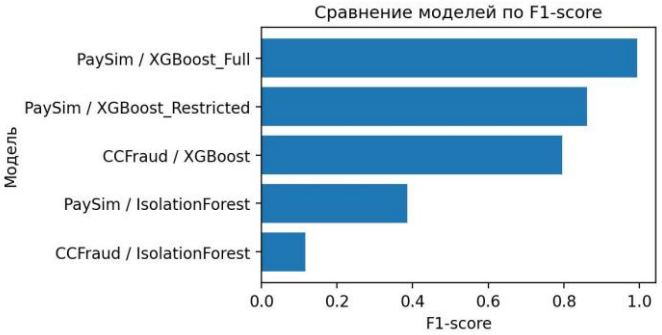

Рисунок 2 – Сравнение моделей по F1-score

Интерпретация результатов также различалась между датасетами. Для PaySim наиболее значимыми признаками стали характеристики суммы операции, исходных и новых балансов, а также тип транзакции, что позволяет объяснять предсказания в прикладных терминах. Для Credit Card Fraud Detection верхние позиции заняли компоненты V4, V14, V12, V10 и V11. Хотя SHAP и permutation importance согласованно выделяли эти признаки, их предметный смысл остается ограниченным, поскольку они являются результатом преобразования пространства признаков. Следовательно, интерпретируемость зависит не только от метода объяснения, но и от смысловой прозрачности исходных данных.

Таким образом, проведенное исследование показало, что в задачах поиска мошеннических транзакций supervised-подход на основе XGBoost обеспечивает более высокое качество по сравнению с Isolation Forest. При этом PaySim лучше подходит для содержательной интерпретации аномалий, а Credit Card Fraud Detection – для сравнительной оценки качества методов на классическом benchmark-наборе. Практически предпочтительной моделью для PaySim следует считать XGBoost_Restricted, так как она сочетает высокое качество с более корректной прикладной интерпретацией.

Список использованных источников:

1. A Unified Approach to Interpreting Model Predictions / S. M. Lundberg, S.-I. Lee // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 4765–4774.

2. Isolation Forest / F. T. Liu, K. M. Ting, Z.-H. Zhou // Proceedings of the 8th IEEE International Conference on Data Mining. – 2008. – P. 413–422.

3. XGBoost: A Scalable Tree Boosting System / T. Chen, C. Guestrin // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – 2016. – P. 785–794.

4. Learned lessons in credit card fraud detection from a practitioner perspective / A. Dal Pozzolo [et al.] // Expert Systems with Applications. – 2014. – Vol. 41, No. 10. – P. 4915–4928.

5. PaySim: A financial mobile money simulator for fraud detection / E. A. Lopez-Rojas, S. Axelsson // 28th European Modeling and Simulation Symposium. – 2016. – P. 249–255.

6. Random Forests / L. Breiman // Machine Learning. – 2001. – Vol. 45, No. 1. – P. 5–32.

7. Credit Card Fraud Detection [Electronic resource] / Kaggle. – 2018. – Mode of access: https://www.kaggle.com/mlg-ulb/creditcardfraud. – Date of access: 26.03.2026.

8. All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously / A. Fisher, C. Rudin, F. Dominici // Journal of Machine Learning Research. – 2019. – Vol. 20, No. 177. – P. 1–81.