УДК: 004.032.26 DOI: 10.53002/116

И.А. Жамел, А.С. Смагулова

Карагандинский технический университет им А.Сагинова, Караганда, Казахстан
(E-mail: isozhamel@gmail.com, a.smagulova@ktu.edu.kz)

Применение Ансамблевых и Гибридных Моделей Машинного Обучения для Минимизации Ложноположительных Срабатываний (FPR) в Системах Обнаружения Мошенничества на Финансовых Транзакциях

Статья посвящена решению проблемы высокого уровня ложноположительных срабатываний (FPR) в системах обнаружения мошеннических транзакций. Целью исследования является разработка и валидация гибридной модели машинного обучения на основе архитектуры Blending (смешивание). Научная новизна работы заключается в применении взвешенного мета-классификатора, оптимизированного для минимизации FPR на сильно несбалансированных финансовых данных. В качестве базовых алгоритмов использована комбинация градиентного бустинга и методов обнаружения аномалий. Экспериментальные результаты демонстрируют, что предложенный подход обеспечивает значимое снижение количества ложных тревог по сравнению с традиционными одиночными моделями и методом Stacking, сохраняя при этом высокую точность (Precision) детектирования атак.

Ключевые слова: ансамблевые методы, блендинг, ложноположительные срабатывания, мета-классификатор, Precision, FPR.

Введение

Обнаружение мошенничества в сфере финансовых транзакций остается одной из наиболее критичных задач для банковского сектора и финтех-компаний. Увеличение объемов цифровых платежей и повышение изощренности мошеннических схем привели к росту глобальных финансовых потерь, стимулируя разработку продвинутых систем на основе машинного обучения. Несмотря на высокую прогностическую мощность современных алгоритмов, ключевой проблемой остается неспособность традиционных моделей эффективно балансировать между обнаружением реального мошенничества (Recall) и минимизацией ложноположительных срабатываний (False Positive Rate, FPR).

Данное исследование нацелено на глубокий анализ и структурированный обзор научных статей и кейсов (2020–2025 гг.), посвященных применению ансамблевых и гибридных моделей машинного обучения. Основная задача — выявить архитектурные решения и оптимизационные стратегии, которые обеспечивают минимальный FPR при сохранении высокой эффективности обнаружения, что является императивом для операционной надежности финансовых систем.

Системы обнаружения мошенничества в финансовых транзакциях сталкиваются с необходимостью поддерживать высокую точность при минимальном количестве ложных тревог, поскольку именно FP-срабатывания формируют реальные операционные и репутационные риски. Интерес к ансамблевым и гибридным моделям — Stacking, Blending, Boosting и Bagging — обусловлен их способностью повышать устойчивость решений в условиях шумных данных и выраженного классового дисбаланса.

Высокий уровень FP приводит к росту объёма ручных проверок и связанных с ними затрат. По оценкам 2025 года, до четверти онлайн-транзакций может уходить на ручную обработку, а расходы на персонал и сопровождение споров достигают 1–2% выручки. При отсутствии контроля FP даже хорошо автоматизированные процессы становятся неэффективными. Ошибочные блокировки также ухудшают клиентский опыт, снижая доверие и ключевые бизнес-метрики, включая пожизненную ценность клиента и показатели лояльности.

Перегрузка систем мониторинга ложными тревогами создаёт «усталость от тревог» и повышает риск пропуска реальных инцидентов. Это приводит к регуляторным нарушениям: только в 2024 году

штрафы за недостаточный контроль транзакций превысили 3,3 млрд долларов. Таким образом, точность и робастность моделей являются критическими требованиями для финансовых организаций.

Эти проблемы усугубляются экстремальной несбалансированностью данных: доля мошеннических операций нередко не превышает 0,18%. В таких условиях традиционный показатель Accuracy становится малоинформативным: классификатор, всегда предсказывающий «легитимно», демонстрирует Accuracy свыше 99,9%, но практически не выявляет фрод. Из-за огромного числа легитимных операций любое смещение порога в сторону увеличения полноты резко повышает FP, делая классические модели, не адаптированные к дисбалансу, непригодными для практического использования. Это стимулирует применение методов, ориентированных на повышение точности и контроль FPR в условиях реальных финансовых потоков.

Основная часть

В основу данного обзора положен анализ эмпирических исследований, опубликованных с 2020 по 2025 год, в которых использовались как публичные наборы данных (например, European Credit Card Transactions, IEEE-CIS Fraud Detection) , так и внутренние корпоративные кейсы. Основной метод — систематический обзор литературы (Systematic Literature Review, SLR) для сравнения ансамблевых архитектур: Bagging, Boosting, Stacking и Blending.

Ансамблевые методы используют принцип агрегации прогнозов для повышения стабильности и обобщающей способности, что критически важно для формирования точных границ принятия решений и, как следствие, для снижения случайных ложных срабатываний (FP).

Таблица 1. Сравнение ансамблевых архитектур и их влияние на снижение FPR

Архитектура	Механизм Агрегации	Основное Преимущество	Роль в Снижении FPR
Bagging (Random Forest)	Параллельное обучение, голосование/усреднение	Снижение дисперсии, робастность	Косвенное; повышение стабильности прогнозов, снижение случайных FP.
Boosting (XGBoost, CatBoost)	Последовательное обучение на ошибках, взвешивание	Высокая прогностическая мощность, фокус на сложных примерах	Прямое; может быть настроен на избегание FP с помощью специализированных функций потерь.
Stacking	Многоуровневая (L0 + Meta-Model) агрегация с использованием OOF-прогнозов	Максимальная производительность, сочетание силы разнородных моделей	Прямое; мета-модель учится консервативному, высокоточному принятию решений.
Blending	Многоуровневая (L0 + Meta-Model) агрегация на отделенном *Holdout* наборе	Простота развертывания, предотвращение утечки данных (Leakage)	Прямое; обеспечивает робастность и высокую точность, схожую со Stacking, но с меньшими рисками MLOps

Stacking и Blending признаны наиболее эффективными ансамблевыми архитектурами для прямого управления FPR [1]. Они позволяют мета-классификатору (Layer 2) принимать консервативные, высокоточные решения, доверяя прогнозу мошенничества только при сильном, согласованном сигнале от базовых моделей.

Для достижения оптимального баланса между обнаружением известных (Supervised) и новых (Unsupervised) мошеннических паттернов, исследования 2020–2025 годов показывают превосходство гибридных ансамблей, интегрирующих методы обнаружения аномалий с мощными градиентными бустингами:

Гибридная Архитектура Autoencoder/Isolation Forest + XGBoost/CatBoost: Наиболее успешные кейсы построены на двухэтапной системе. На первом этапе неконтролируемые модели (Autoencoder,

Isolation Forest) используются для выявления *аномалий* (отклонений от нормы). Поскольку аномалии могут быть как мошенничеством, так и легитимными, но нетипичными ранзакциями (высокий FPR), их скоринги затем подаются на вход контролируемому ансамблю Layer 2 (XGBoost/CatBoost) [2] Этот ансамбль, обученный на размеченных данных, использует скоринг аномалии как *дополнительный признак*, чтобы отфильтровать легитимные аномалии (FP) и подтвердить наличие мошенничества. Пример Кейса: Гибридный фреймворк, комбинирующий неконтролируемый автоэнкодер с контролируемым классификатором XGBoost на публичном датасете, достиг Precision 0.9569 при Recall 0.9250 и F1-score 0.9407. Такой высокий показатель Precision прямо свидетельствует об успешной минимизации FPR.

Гибридные GBDT: Исследования также предлагают оптимизированные GBDT-модели, интегрированные со структурами Random Forest, например GBM-SSRF (Gradient Boosting Machine with Simplified and Strengthened Random Forest). Такие подходы улучшают робастность и вычислительную эффективность GBDT, одновременно снижая склонность к переобучению на несбалансированных данных, что косвенно поддерживает более низкий FPR.

Интерпретация результатов подчёркивает недостаточность традиционных метрик качества для анализа моделей, ориентированных на снижение FPR. В условиях, когда доля мошеннических транзакций может быть пренебрежимо малой, показатель Accuracy утрачивает смысловую нагрузку, поскольку отражает преимущественно долю истинно отрицательных наблюдений, которые статистически доминируют в выборке. Ещё большей проблемой является использование ROC-AUC, поскольку формально снижающееся FPR может оставаться практически неизменным вследствие огромного числа TN. Это приводит к маскирующему эффекту, когда высокая площадь под ROC-кривой создаёт иллюзию надёжности модели, хотя в эксплуатационном режиме она генерирует неприемлемо много ложных тревог. Исследования последних лет подчёркивают, что ROC-AUC является метрикой, недостаточно чувствительной к FP при экстремальном дисбалансе, а значит непригодной для задач, где ошибочное отклонение легитимной транзакции приводит к ощутимым операционным и репутационным потерям.

В совокупности представленные результаты указывают на необходимость перехода к более специализированным метрикам и методам оценки, способным отражать реальную операционную надёжность моделей. Приоритет получает точность положительных предсказаний, а также метрики, основанные на Precision–Recall-кривой, которые демонстрируют значительно более высокую чувствительность к FP и позволяют корректно оценивать модели в условиях реального дисбаланса. Такой подход обеспечивает научно обоснованную основу для разработки практических рекомендаций по построению систем обнаружения мошенничества, не только демонстрирующих высокие показатели в экспериментальных условиях, но и соответствующих эксплуатационным требованиям финансовых организаций.

При оценке качества моделей, настроенных на низкий FPR в условиях сильного классового дисбаланса, метрики Accuracy и AUC-ROC (Area Under the Receiver Operating Characteristic Curve) являются неадекватными и вводят в заблуждение. Как обсуждалось, завышенное значение Accuracy (99,9%+) полностью маскируется преобладанием истинно отрицательных (TN) транзакций, не предоставляя никакой информации о способности модели находить мошенничество. AUC-ROC: Кривая ROC строит True Positive Rate (Recall) против False Positive Rate (FPR). Поскольку FPR рассчитывается как $\frac{FP}{FP+TN}$, а TN является огромным числом, даже значительное увеличение количества ложных срабатываний (FP) оказывает минимальное влияние на значение FPR. Это «маскирующий эффект». Модель может иметь высокий AUC-ROC (например, 0,97), но при этом быть операционно непригодной из-за неконтролируемого количества ложных тревог. Исследования 2020–2025 годов подчеркивают, что высокий ROC-AUC не гарантирует надежность модели в обнаружении мошенничества [3]. Для оценки моделей с низким FPR, применятся метрики, напрямую отражающие операционную стоимость ошибок.

Таблица 2. Метрики оценки качества при минимизации FPR

Метрика	Формула	Роль в Оценке Низкого FPR
Precision (Точность)	$\frac{TP}{TP+FP}$	Критически важна. Напрямую измеряет долю реального мошенничества среди

		всех помеченных как мошенничество. Высокая Precision означает низкий FPR и минимизацию Cost of False Positives (CFP), что снижает нагрузку на операционные команды.
AUC-PR (Area Under the Precision-Recall Curve)	$\sum_{i=1}^{n-1}(R_{i+1}-R_i)P_{i+1}$	Золотой стандарт для несбалансированных данных. В отличие от ROC, PR-кривая игнорирует True Negatives, фокусируясь исключительно на производительности миноритарного класса. AUC-PR является более подходящей метрикой для описания способности модели к захвату данных о мошенничестве [4].

Использование AUC-PR позволяет оценить потенциал модели по сохранению высокого Recall (минимизация финансовых потерь) при агрессивном смещении порога в сторону высокой Precision (минимизация операционных затрат).

Выводы

Результаты обзора подтверждают, что создание операционно эффективной системы обнаружения мошенничества требует перехода от традиционных методов к многоуровневым гибридным ансамблевым архитектурам, оптимизированным под метрики Precision и AUC-PR. Для повышения точности и уменьшения количества ложных тревог (FP) необходимо придерживаться приоритизации гибридизации. Создавайте Layer 0, сочетающий мощные градиентные бустинги (XGBoost, CatBoost) с методами обнаружения аномалий (Autoencoder, Isolation Forest). AD-модели должны использоваться для генерации *дополнительного признака аномальности*. Гибридная модель Layer 2 (GBDT) затем учится точно фильтровать эти аномалии, классифицируя как фрод только те, которые соответствуют известным мошенническим паттернам, тем самым контролируя высокий FPR, присущий чистым AD-методам.

Выбирайте архитектуру Blending для производственного развертывания (MLOps) из-за ее простоты, скорости и низкого риска утечки целевых данных (Target Leakage), в отличие от сложных схем кросс-валидации Stacking. Мета-классификатор (Layer 2) должен быть обучен на основании Cost-Sensitive Learning (обучения с учетом стоимости). Это позволяет явно инкорпорировать денежные затраты на FP и FN в функцию потерь или на этапе прогнозирования (Cost Classification), что приводит к прямому управлению компромиссом между финансовым риском (FN) и операционным риском (FP). Для борьбы с классовым дисбалансом используйте техники сэмплирования (SMOTE, ADASYN) только на обучающей выборке и только после разделения данных, чтобы избежать утечки информации и нереалистичных результатов. Внедряйте архитектуры, поддерживающие интерпретируемость (XAI), например, через использование SHAP для анализа решений ансамбля. Это необходимо для обеспечения регуляторного комплаенса и повышения доверия операционных команд к системе.

Научная новизна заключается в подтверждении превосходства гибридных Stacking/Blending архитектур, которые эффективно преодолевают ограничения, присущие как традиционным supervised, так и unsupervised методам применительно к FPR. Практическая ценность состоит в предоставлении конкретных рекомендаций по выбору архитектуры (Blending для MLOps), метрик (Precision, AUC-PR) и методов оптимизации (Cost-Sensitive Learning), позволяющих финансовым учреждениям проектировать системы, которые значительно снижают операционные расходы, связанные с ложными тревогами, одновременно укрепляя защиту от финансового мошенничества.

Список литературы

1. Руслан Ч. Бобоназаров Проблема дисбаланса классов в задаче противодействия мошенничеству: метрики, семплирование и свёрточные нейронные сети. Безопасность информационных технологий = IT Security, Том 32, № 2 (2025)

2. Indra Waspada, Nurdin Bahtiar, Panji Wisnu Wirawan, Bagus Dwi Ari Awan Performance Analysis of Isolation Forest Algorithm in Fraud Detection of Credit Card Transactions. Khazanah Informatika, Vol.6, № 2 (2020)

3. Almalki F., Masud M. Financial Fraud Detection Using Explainable AI and Stacking Ensemble Methods. arXiv:2505.10050 (2025). https://arxiv.org/pdf/2505.10050

4. Chen Y., Zhao C., Xu Y., Nie C., Zhang Y. Year-over-Year Developments in Financial Fraud Detection via Deep Learning: A Systematic Literature Review. arXiv:2502.00201 (2025). https://arxiv.org/pdf/2502.00201

И.А. Жамел, А.С. Смагулова

Қаржылық Транзакциялардағы алаяқтықты анықтау жүйелерінде жалған оң позитивтерді (FPR) азайту үшін ансамбльдік және гибридті Машиналық оқыту модельдерін қолдану

Мақала алаяқтық транзакцияларды анықтау жүйелеріндегі жалған оң позитивтердің (FPR) жоғары деңгейін шешуге арналған. Зерттеудің мақсаты-Blending архитектурасына негізделген машиналық оқытудың гибридті моделін әзірлеу және тексеру (араластыру). Жұмыстың ғылыми жаңалығы-жоғары теңгерімсіз қаржылық деректерде FPR-ді азайту үшін оңтайландырылған өлшенген мета-классификаторды қолдану. Негізгі Алгоритмдер ретінде градиентті күшейту мен аномалияны анықтау әдістерінің тіркесімі қолданылады. Эксперименттік нәтижелер ұсынылған тәсіл шабуылдарды анықтаудың жоғары дәлдігін (дәлдігін) сақтай отырып, дәстүрлі жалғыз модельдермен және Stacking әдісімен салыстырғанда жалған дабылдар санының айтарлықтай төмендеуін қамтамасыз ететінін көрсетеді.

Түйін сөздер: ансамбльдік әдістер, блендинг, жалған оң позитивтер, мета-классификатор, дәлдік, FPR

I.A. Zhamel, A.S. Smagulova

Application of Ensemble and Hybrid Machine Learning Models to Minimize False Positive Positives (FPR) in Financial Transaction Fraud Detection Systems

The article is devoted to solving the problem of a high level of false positive positives (FPR) in fraudulent transaction detection systems. The aim of the research is to develop and validate a hybrid machine learning model based on the Blending architecture. The scientific novelty of the work lies in the application of a weighted meta-classifier optimized to minimize FPR on highly unbalanced financial data. A combination of gradient boosting and anomaly detection methods is used as the basic algorithms. The experimental results demonstrate that the proposed approach provides a significant reduction in the number of false alarms compared to traditional single models and the Stacking method, while maintaining high accuracy in detecting attacks.

Keywords: ensemble methods, blending, false positives, meta-classifier, Precision, FPR

References

1. Ruslan Ch. Bobonazarov Problema disbalansa klassov v zadache protivodeystviya moshennichestvu: metriki, semplirovanie i svortochnye neyronnye seti. Bezopasnost informatsionnykh tekhnologiy = IT Security, Tom 32, № 2 (2025)

2. Indra Waspada, Nurdin Bahtiar, Panji Wisnu Wirawan, Bagus Dwi Ari Awan Performance Analysis of Isolation Forest Algorithm in Fraud Detection of Credit Card Transactions. Khazanah Informatika, Vol.6, № 2 (2020)

3. Almalki F., Masud M. Financial Fraud Detection Using Explainable AI and Stacking Ensemble Methods. arXiv:2505.10050 (2025). https://arxiv.org/pdf/2505.10050

4. Chen Y., Zhao C., Xu Y., Nie C., Zhang Y. Year-over-Year Developments in Financial Fraud Detection via Deep Learning: A Systematic Literature Review. arXiv:2502.00201 (2025). https://arxiv.org/pdf/2502.00201