УДК 004.891

ИНТЕРПРЕТИРУЕМОСТЬ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ И ПРОБЛЕМА ДИСБАЛАНСА КЛАССОВ В ЗАДАЧАХ СНИЖЕНИЯ РИСКОВ КРЕДИТНО-ФИНАНСОВЫХ ОРГАНИЗАЦИЙ

Феста Ю.Ю. (Национальный исследовательский университет «Высшая школа экономики»),
Воробьев И.А. (Национальный исследовательский университет «Высшая школа экономики»)
Научный руководитель – доктор экономических наук Пеникас Г.И.
(Национальный исследовательский университет «Высшая школа экономики»)

Введение.

Рост рынка электронной коммерции и цифровизация финансовой отрасли привели к тому, что машинное обучение стало доминирующим инструментом для решения задач снижения рисков в этой сфере. Отличительной особенностью процесса построения моделей для выявления рисков является малое количество целевых объектов в обучающих наборах данных, так называемый дисбаланс классов [1]. Для улучшения результирующих показателей модели исследователи применяют различные техники балансировки классов (например, SMOTE) или используют алгоритмы глубокого обучения. Вследствие такого усложнения, полученная модель теряет интерпретируемость, а эксперты, работающие с результатами классификации, не получают детальную информации о случившемся риске [2]. Также алгоритмы искусственного увеличения выборки имеют свои недоставки, например, отсутвие учета совместного распределения объясняющих переменных [3]. Данная проблема может привести к росту недоверия у регуляторов и экспертов к модели или даже полному отказу от ее использования при анализе рисков. В данном исследовании предлагается рассмотреть подходы к улучшению интерпретируемости результатов классификации в задачах противодействия мошенничеству в банке и страховой компании, а также в оценки вероятности дефолта банка.

Основная часть.

Немногие финансовые компании готовы публиковать наборы с размеченными данными для исследования и применения методов машинного обучения. В области оценки рисков такие данные вообще единичны по причинам, связанным с конфиденциальностью [4]. В данном исследовании использован уникальный частный набор данных крупного банка, а также собранная из открытых источников информация о дефолтах зарубежных банков. Вместе с тем исследован единственный открытый набор данных о мошенничестве в страховании. Для каждого набора сформирован уникальный процесс обучения модели повышающий интерпретируемость и одновременно решающий проблему дисбаланса классов. С этой целью применены различные подходы отбора признаков и выделения из них наиболее значимых. Также использована теория графов для создание новых признаков, которые усиливают разделяющую способность модели классификации. Рассмотрены методы индукции правил для поддержки принятия решения экспертами. Проведены численные эксперименты, а для оценки результатов выбраны традиционно используемые в задачах выявления рисков метрики оценки классификации – точность, полнота и AUC под кривыми ROC и PR [5]. Положительный эффект, заключающийся в улучшении качества классификации рисковых событий, достигается благодаря включению в обучение модели признаков, сформулированных экспертами. Полученные на их основе правила интерпретируются экспертами и, в отличие от более сложных моделей машинного обучения, могут быть использованы при составлении описательной части события.

Выводы.

Исследована возможность повышения интерпретируемости и качества классификации рисковых событий в условиях дисбаланса классов. Прирост эффективности предложенного подхода подтвержден приростом основных метрик моделей. При использовании

предложенного подхода у финансовых компаний появляется возможность не только выносить вердикты по рисковым событиям, но и составлять описательную часть рискового события с помощью методов машинного обучения.

Список использованных источников:

1. Gupta P. и др. Unbalanced Credit Card Fraud Detection Data: A Machine Learning-Oriented Comparative Study of Balancing Techniques // Procedia Computer Science. 2023. Т. 218. С. 2575–2584.
2. Vorobyev I., Krivitskaya A. Reducing false positives in bank anti-fraud systems based on rule induction in distributed tree-based models // Computers & Security. 2022. Т. 120. С. 102786.
3. Blagus R., Lusa L. SMOTE for High-Dimensional Class-Imbalanced Data // BMC bioinformatics. 2013. Т. 14. С. 106.
4. Subudhi S., Panigrahi S. Use of optimized Fuzzy C-Means clustering and supervised classifiers for automobile insurance fraud detection // Journal of King Saud University - Computer and Information Sciences. 2020. Т. 32. № 5. С. 568–575.
5. Baesens B., Höppner S., Verdonck T. Data engineering for fraud detection // Decision Support Systems. 2021. Т. 150. С. 113492.