ИНФОРМАТИКА

Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. 2024. Т. 24, вып. 4. С. 619–628
Izvestiya of Saratov University. Mathematics. Mechanics. Informatics, 2024, vol. 24, iss. 4, pp. 619–628
https://mmi.sgu.ru
https://doi.org/10.18500/1816-9791-2024-24-4-619-628
EDN: HLJKIS

Научная статья
УДК 004.891

Методы машинного обучения в задаче оценки риска мошенничества в автостраховании

И. А. Воробьев

Московский институт электроники и математики им. А. Н. Тихонова НИУ «Высшая школа экономики», Россия, 123458, г. Москва, ул. Таллинская, д. 34

Воробьев Иван Александрович, аспирант кафедры компьютерной безопасности, vorobyev-ivan@yandex.ru, https://orcid.org/0000-0002-2886-6813

Аннотация. Оценка уровня мошенничества в автостраховании представляет собой актуальную и сложную задачу, что обусловлено деятельностью мошеннических групп. Для уверенности менеджмента страховых компаний в стратегии противодействия мошенничеству необходим инструмент, позволяющий оценить текущее состояние портфеля претензий. Современные методы машинного обучения позволяют проводить такую оценку, используя данные о страхователях и страховых случаях. При применении данных подходов возникает ряд проблем, не позволяющих достичь необходимого качества выявления мошенничества. К ним можно отнести дисбаланс классов и так называемый дрейф концепции (concept drift), возникающий вследствие изменения сценариев схем мошенников и субъективности экспертной оценки конкретного страхового случая. В настоящем исследовании предлагается подход, позволяющий улучшить метрики моделей для выявления мошенничества в портфеле претензий. Численный эксперимент на двух открытых наборах данных показал прирост полноты выявления страхового мошенничества на 49 п.п. и 19 п.п. в сравнении с классическим моделированием.

Ключевые слова: оценка риска, выявление мошенничества, страховые претензии, машинное обучение, дрейф концепции, дисбаланс классов

Для цитирования: *Воробьев И. А.* Методы машинного обучения в задаче оценки риска мошенничества в автостраховании // Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. 2024. Т. 24, вып. 4. С. 619–628. https://doi.org/10.18500/1816-9791-2024-24-4-619-628, EDN: HLJKIS

Научный отдел

Article

ML methods for assessing the risk of fraud in auto insurance

I. A. Vorobyev

HSE Moscow Institute of Electronics and Mathematics, 34 Tallinskaya St., Moscow 123458, Russia

Ivan A. Vorobyev, vorobyev-ivan@yandex.ru, https://orcid.org/0000-0002-2886-6813

Abstract. The car insurance fraud level assessment is an urgent and complex task, which is largely due to the activities of fraudulent groups. For the confident management of insurance companies in the anti-fraud strategy, a tool to assess the current state of the claim's portfolio is needed. Modern machine learning methods make it possible to carry out such an assessment using data on policyholders and insurance cases. When applying these approaches, a number of problems arise that do not allow achieving the required quality of fraud detection. These include class imbalance and the so-called concept drift, which arises as a result of changes in the scenarios of fraudsters' schemes and the subjectivity of the expert assessment of a specific insurance case. This study proposes an approach to improve model metrics for detecting fraud in a claims portfolio. A numerical experiment conducted on two open data sets demonstrated a significant improvement in the detection rate of insurance fraud compared to classical modeling. Specifically, there was an increase in the completeness of fraud detection by 49 and 19 percentage points for the two datasets, respectively.

Keywords: risk assessment, fraud detection, insurance claims, machine learning, concept drift, class imbalance

For citation: Vorobyev I. A. ML methods for assessing the risk of fraud in auto insurance. *Izvestiya of Saratov University. Mathematics. Mechanics. Informatics*, 2024, vol. 24, iss. 4, pp. 619–628 (in Russian). https://doi.org/10.18500/1816-9791-2024-24-4-619-628, EDN: HLJKIS

Введение

Развитие технологий хранения и обработки данных способствовало распространению методов машинного обучения (ML) в различных областях, включая сферы, где требуется принятие решений. Благодаря накопленным массивам информации появилась возможность сделать переход от экспертных подходов к автоматизированным, основанным на данных. При этом основным инструментом в таких задачах стали модели классификации, извлекающие закономерности из наборов данных.

Для улучшения качества классификации используются различные подходы, которые зависят от области применения, объема данных, наличия разметки у объектов классификации, дисбаланса классов и других факторов. Сообществом ученых разработаны доступные инструменты (например, открытая библиотека машинного обучения scikit-learn (https://scikit-learn.org/)), облегчающие решение типовых проблем, с помощью которых исследователи достигают значительных результатов в задачах принятия решений.

Тем не менее определенные бизнес-процессы требуют индивидуального подхода к внедрению методов машинного обучения в них. В настоящем исследовании предлагается улучшить качество классификации путем корректировки разметки данных с использованием нейронной сети в тех наборах данных, где маркировка целевого класса проводится экспертами. Новизна такого подхода заключается в том, что улучшение классификации достигается не усложнением методов ML, а вследствие повышения качества входных данных для обучения, в которых высока вероятность наличия некорректной экспертной оценки или присутствия экземпляров, не рассмотренных экспертами.

Одной из областей, где в последние годы методы машинного обучения дали возможность решить ряд проблем, является обнаружение мошенничества в автостраховании [1]. Однако

повышение эффективности классификации в этой сфере остается актуальной проблемой. В связи с этим тестирование применимости предлагаемого подхода проведено на наборах данных о страховом мошенничестве.

По данным Всероссийского союза страховщиков (https://ins-union.ru/), в течение года страховые компании подают в суд более 8000 исковых заявлений, а сумма выплат мошенникам составляет около 0.5% от общего объема выплат по претензиям страхователей. Это существенно влияет на процесс формирования тарифов и в конечном счете отражается на законопослушных клиентах.

Традиционно процесс выявления мошеннических претензий состоит из анализа истории участников и сравнения полученных результатов с рассматриваемым страховым случаем [2]. В результате эксперт принимает решение по претензии о наличии признаков страхового мошенничества. Накопление таких вердиктов также является ценной информацией и в будущем может рассматриваться как разметка для применения методов машинного обучения с учителем.

С другой стороны, сотрудники страховых компаний могут иметь собственные предубеждения в части определения признаков мошенничества. Например, эксперты, имеющие интерес к риску мошенничества, могут оценить одну и ту же претензию по-разному. С течением времени в попытках обойти фрод-мониторинг меняется и поведение мошенников, при этом, согласно индустриальному опросу (https://www.friss.com/insight/insurance-fraud-report-2022/), реакция экспертов не всегда успевает за изменениями. Поэтому претензия, размеченная как легитимная, на деле может оказаться мошеннической, а вследствие незнания новой схемы эксперт допускает ошибку. По этим причинам в данных о мошенничестве может возникнуть так называемый concept drift [3], что приводит к неустойчивости моделей машинного обучения во времени.

Предлагаемый подход вносит вклад в исследования, посвященные выявлению мошенничества в автостраховании, и сосредоточен на повышении качества классификации моделей машинного обучения, применяемых к претензиям страхователей. Основной особенностью данной задачи является дисбаланс классов — поиск мошенничества схож с поиском иголки в стоге сена, что накладывает определенные ограничения на использование моделей классификации. В отличие от других работ из данной области (например, [4]), где проблема дисбаланса класса решается традиционными техниками (Undersampling, Oversampling, SMOTE и др.), в настоящем исследовании проводится корректировка целевого класса с помощью нейронной сети, что позволяет сбалансировать данные для использования методов машинного обучения совместно с устранением проблемы «concept drift». При этом предполагается снижение размерности пространства признаков исходного набора данных с сохранением метрик качества выявления мошенничества. Вследствие снижения достигается интерпретируемость принятого решения по претензии, что решает проблему нехватки бизнес-контекста в антифрод системах, рассмотренную, например, в финансовой сфере [5].

В данной статье:

1) представлен новый подход, который позволяет улучшить разделяющую способность классификатора путем повышения качества данных для обучения;
2) продемонстрировано, что использование нейронной сети для корректировки экспертной разметки позволяет решить проблемы дисбаланса классов и «concept drift», что приводит к улучшению качества классификации;
3) предлагаемый подход рассмотрен в задаче выявления мошенничества в автостраховании;
4) проведен численный эксперимент на двух открытых наборах данных с разметкой мошеннических претензий, который подтверждает эффективность предложенного подхода.

1. Область исследования

Рассматривается процесс страховой компании, в котором снижается риск мошенничества со стороны страхователей. Первой преградой для мошенников является проверка клиента

перед заключением договора страхования. Помимо актуарных расчетов страховщик может обратиться к внутренним черным или белым спискам, внешним источникам данных о клиенте (например, кредитной истории), применить модели оценки риска мошенничества со стороны страхователя. Такие процедуры напрямую влияют на страховщика — удлиняют процесс продажи полиса, ухудшая клиентский опыт, а ложные отказы снижают уровень сбора страховой премии. Эти факты заставляют страховые компании упрощать и автоматизировать проверки на данном этапе. В таком случае менеджмент компании ориентируется на точность выявления мошенничества, что позволяет профессиональным мошенникам проникать в страховой портфель.

Далее страховщик анализирует заявленные претензии и при выявлении признаков страхового мошенничества отказывает в выплате. На данном этапе применяются как экспертные методы оценки со стороны службы безопасности страховщика, так и техники с использованием методов машинного обучения. Положительный эффект дает совмещение работы эксперта и систем, позволяющих оценивать претензии с помощью анализа данных и социальных сетей. В текущем процессе страховая компания ориентируется на полноту выявления мошенничества, чтобы остановить уже проявившего себя профессионального мошенника и снизить его влияние на показатель убыточности портфеля.

В данном исследовании предлагается метод повышения качества оценки риска мошенничества в автостраховании с использованием машинного обучения при выплате претензии.

2. Методы и алгоритм оценки риска

Оценка претензии на риск мошенничества методами машинного обучения осуществляется с помощью исторических данных. Каждая претензия обладает своим признаковым описанием и, если она обрабатывалась экспертом, имеет ответ на вопрос о наличии в ней мошенничества со стороны страхователя. Тогда задачу выявления мошенничества можно свести к задаче обучения по прецедентам [6]. В частности, будем рассматривать задачу классификации с двумя непересекающимися классами. При этом найденная решающая функция (далее — модель или классификатор) будет использоваться для оценки конкретной претензии на риск мошенничества с помощью ее признакового описания (далее — признаки).

Для оценки результатов эксперимента выбраны традиционно используемые в задачах выявления мошенничества метрики:

$$Recall = \frac{TP}{(TP + FN)}, \quad Precision = \frac{TP}{(TP + FP)},$$

где TP (True positive) — мошенническая претензия идентифицирована корректно, FP (False positive) — легитимная претензия идентифицирована как мошенническая, TN (True negative) — легитимная претензия идентифицирована корректно, FN (False negative) — мошенническая претензия идентифицирована как легитимная.

Recall (полнота) позволит оценить долю мошенничества, выявленную классификатором по отношению ко всем мошенническим претензиям; Precision (точность) — вероятность того, что подозреваемая классификатором претензия действительно мошенническая.

Также для сравнения моделей и прироста качества выявления мошенничества будет использоваться ROC кривая [7]. Увеличение площади под кривой ROC (AUC) будет выступать в качестве меры улучшения построенного классификатора.

Открытые наборы данных автострахования, пригодные для исследования с точки зрения машинного обучения и с размеченными случаями мошенничества, встречаются крайне редко [8]. Один из них — широко известный и используемый в различных исследованиях «carclaims.txt». В нем содержатся страховые случаи, зарегистрированные в США за период с 1994 по 1996 гг. [9].

Также для демонстрации применимости подхода на различных страховых данных рассмотрен файл «insurance_claims.csv» (https://www.kaggle.com/datasets/buntyshah/auto-insurance-claims-data/), в котором содержатся претензии за период с января по февраль 2015 г.

Для целей данного исследования выбраны признаки, приведенные в табл. 1.

Таблица 1 / Table 1

Описание признаков претензий для оценки мошеннической составляющей
Description of claims characteristics for assessing the fraudulent component

Набор данных	Название признака	Описание
«carclaims.txt»	Age	Возраст страхователя
	DriverRating	Рейтинг водителя
	Gender	Пол страхователя
	BasePolicy	Тип полиса
	Fault	Виновная сторона
	NumberOfSuppliments	Количество доп. опций
	PastNumberOfClaims	Количество страховых случаев
	VehiclePrice	Стоимость автомобиля
	AgeOfPolicyHolder	Возраст страхователя
«insurance_claims.csv»	age	Возраст страхователя
	policy_annual_premium	Страховая премия
	insured_sex	Пол страхователя
	total_claim_amount	Сумма претензии
	incident_severity	Серьезность страхового случая

Такой малочисленный состав признаков выбран с целью сохранения применимости предложенного подхода для оценки мошеннической составляющей в различных портфелях претензий страховых компаний. Данный набор признаков можно получить из анкетных данных страхователей и претензий при урегулировании страхового случая. Категориальные признаки в экспериментах были закодированы в числовые по принципу one-hot (https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html/).

Набор данных «carclaims.txt» состоит из 15 420 записей, из которых 14 497 являются легитимными претензиями, а 923 (6.0%) — с признаками страхового мошенничества. Размер «insurance_claims.csv» составляет 1000 записей, из которых 247 — мошеннические (24.7%). Претензии можно выстроить в хронологическом порядке по тому, как они поступали в страховую компанию, поэтому качество модели предлагается проверять на более поздних данных, так называемой out-of-time выборке.

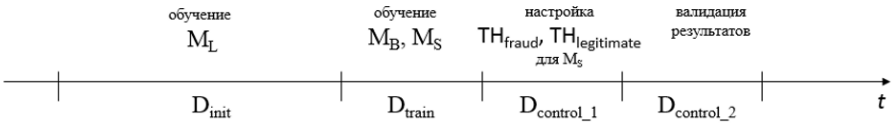
Рассмотрим следующий процесс моделирования для снижения дисбаланса классов и решения проблемы «concept drift»:

1) набор данных разбивается на четыре части из разных временных периодов;
2) более ранняя часть (D_{init}) предназначена для обучения модели M_L, с помощью которой будет корректироваться экспертная оценка;
3) следующая по времени (D_{train}) используется для обучения модели M_S предлагаемым подходом на переразмеченном наборе данных, а также для обучения базовой модели M_B, с которой будет сравниваться результат эксперимента;
4) следующая часть ($D_{control\_1}$) предназначена для поиска точек отсечения (TH_{fraud}, $TH_{legitimate}$) модели M_L, в зависимости от значений которых будет приниматься решение о корректировки разметки в D_{train};
5) наконец на выборке $D_{control\_2}$ будет проводиться валидация результатов.

Разбиение данных и классификаторы для оценки претензий схематично представлены на рис. 1.

В табл. 2 приведены используемые параметры и ссылки на описание моделей, примененных в процессе обучения.

Выбор Random Forest и нейронной сети в качестве классификатора обусловлен исследованием [10], в котором показано сравнение основных методов машинного обучения в задачах выявления мошенничества в автостраховании. Для получения лучших параметров моделей использовался широко известный подход кросс-валидации, в частности GridSearchCV



«carclaims.txt»	количество претензий:	6420	3000	3000	3000
	дисбаланс классов:	6,6%	5,7%	5,8%	5,2%
«insurance_claims.csv»	количество претензий:	255	241	253	251
	дисбаланс классов:	26.6%	25.3%	28.4%	18.3%

Рис. 1. Применение предлагаемого подхода на наборах данных претензий автострахования
Fig. 1. Application of the proposed approach on automobile insurance claims datasets

Таблица 2 / Table 2

Используемые в процессе обучения классификаторы
Classifiers used in the training process

Название	Классификатор	Параметры	Ссылка на описание
Multilayer perceptron	M_L	«carclaims.txt»: $hidden\_layer\_sizes = (10)$, $solver =' lbfgs'$ «insurance_claims.csv»: $hidden\_layer\_sizes = (2)$, $solver =' lbfgs'$, $activation =' relu'$	https://scikit-learn.org/stable/modules/neural_networks_supervised.html (дата обращения: 22.09.2023)
Random Forest Classifier	M_B	«carclaims.txt»: $class\_weight = \{0:1, 1:1\}$, $criterion =' entropy'$, $n\_estimators = 5$ «insurance_claims.csv»: $class_weight = \{0:1, 1:1\}$, $criterion =' entropy'$, $n\_estimators = 2$, $max\_depth = 3$	https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (дата обращения: 22.09.2023)
	M_S	«carclaims.txt»: $class\_weight = \{0:1, 1:3\}$, $criterion =' entropy'$, $n\_estimators = 5$ «insurance_claims.csv»: $class_weight = \{0:1, 1:1\}$, $criterion =' entropy'$, $n\_estimators = 2$, $max\_depth = 3$	

(https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html/) при оптимизации метрики roc_auc.

Экспериментально с помощью измерения качества классификации на $D_{control\_1}$ подобраны значения TH_{fraud}, $TH_{legitimate}$:

а) для набора данных «carclaims.txt»: $TH_{fraud} = 0.75$; $TH_{legitimate} = 0.1$;

б) для «insurance_claims.csv»: $TH_{fraud} = 0.8$; $TH_{legitimate} = 0.05$.

После этого претензии в D_{train} были переразмечены следующим образом:

1) если результат оценки (вероятность отнесение к мошенничеству) претензии с помощью M_L больше TH_{fraud}, то она переразмечается как мошенническая;
2) если результат оценки менее $TH_{legitimate}$, то претензия переразмечается как легитимная;
3) в остальных случаях разметка претензии не подвергается изменению.

Данная корректировка классов улучшила баланс в D_{train} до 36.3% в наборе данных «carclaims.txt» и до 35.8% в «insurance_claims.csv». Далее проведено обучение M_B на данных D_{train} до корректировки классов и M_S на данных D_{train} после корректировки.

Этапы проведенного эксперимента схематично показаны на рис. 2.

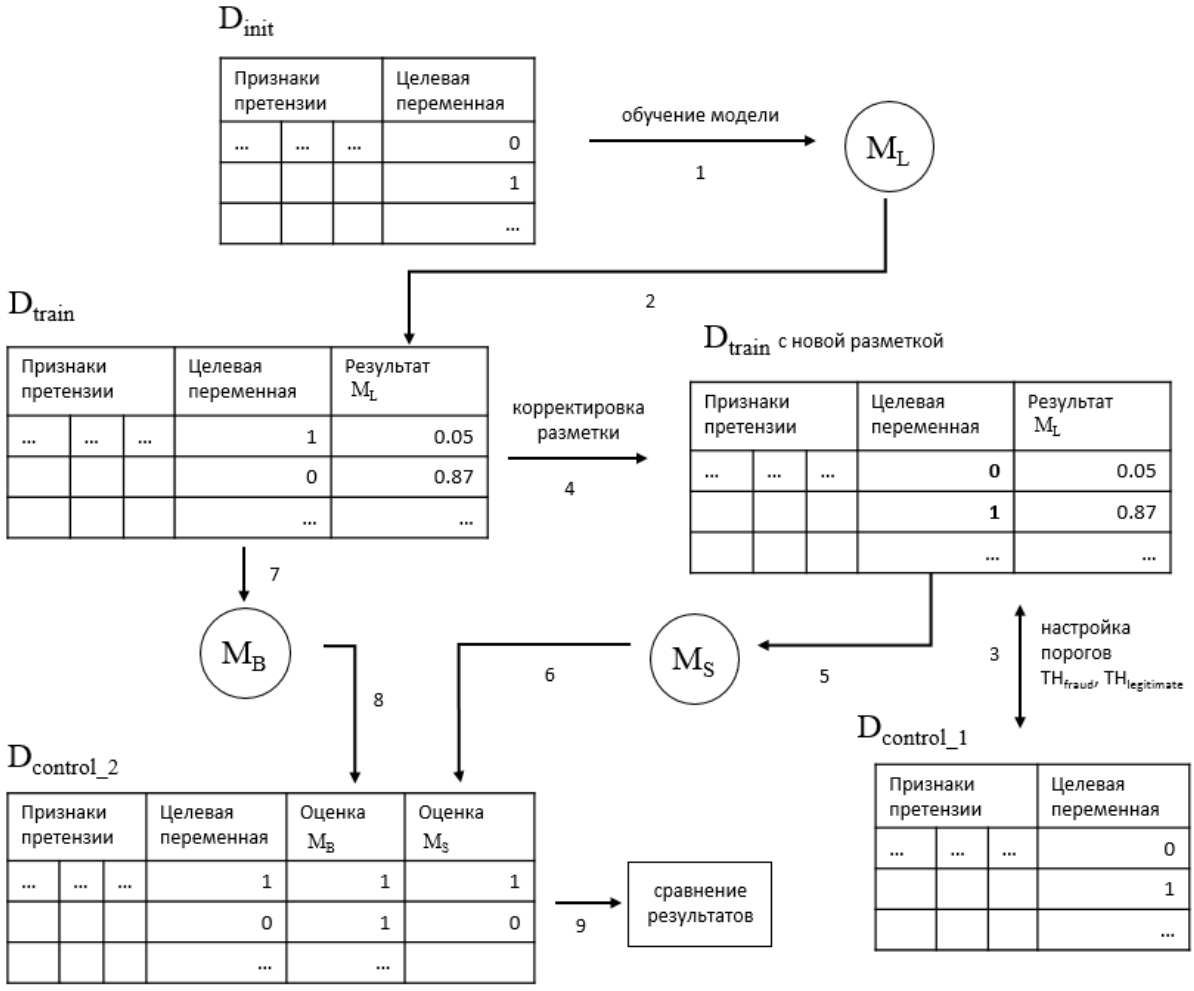


Рис. 2. Этапы проведения эксперимента
Fig. 2. Stages of conducting an experiment

3. Результаты эксперимента

Полученные модели M_B и M_S применены в выборке $D_{control\_2}$, которая никаким образом не участвовала в обучении классификаторов или настройке параметров и является более поздней по времени с точки зрения появления претензии у страховщика. Также в данной выборке разметка не подвергалась корректировке. На рис. 3 представлены ROC кривые для этих моделей.

Значения площадей под данными кривыми (AUC) показывает значительно лучшее качество классификации для предложенного подхода в сравнении с традиционным обучением без

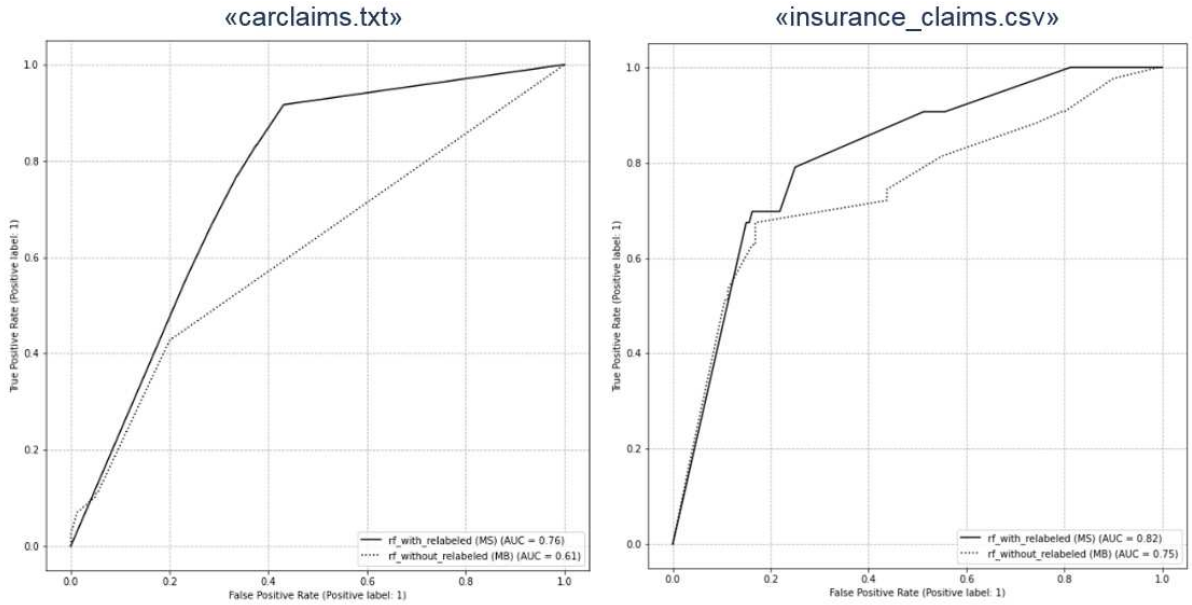


Рис. 3. Сравнение ROC кривых для моделей M_B и M_S
Fig. 3. Comparison of ROC curves for M_B and M_S models

корректировки разметки. В табл. 3 также приведены значения Recall при фиксированной точности для сравнения моделей.

Таблица 3 / Table 3

Сравнение метрик качества выявления мошенничества
Comparison of fraud detection quality metrics

Набор данных	Метрика	Предложенный подход, M_S	Традиционный подход, M_B
«carclaims.txt»	ROC AUC	**0.76**	0.61
	Precision	0.10	0.10
	Recall	**0.92**	0.43
«insurance_claims.csv»	ROC AUC	**0.82**	0.75
	Precision	0.55	0.55
	Recall	**0.70**	0.51

Заключение

В настоящей работе исследована возможность корректировки экспертной разметки данных с помощью нейронной сети для дальнейшего применения методов машинного обучения. Прирост эффективности выявления мошенничества в данном случае подтвержден приростом основных метрик качества классификации. Для иллюстрации эффективности использованы два открытых набора данных о страховом мошенничестве.

Предложенный подход позволяет решить проблемы дисбаланса классов и «concept drift» за счет добавления в обучение мошеннических кейсов, ошибочно помеченных экспертами как легитимные.

На следующем этапе исследования предполагается расширить область применения данного подхода на оценку риска мошенничества в банковских операциях, где также существует проблема некачественной разметки данных.

Список литературы

1. *Bao Y., Hilary G., Ke B.* Artificial intelligence and fraud detection // Innovative technology at the interface of finance and operations / ed. by V. Babich, J. R. Birge, G. Hilary. Cham :

Springer, 2022. P. 223–247. (Springer Series in Supply Chain Management, vol. 11). https://doi.org/10.1007/978-3-030-75729-8_8
2. *Subelj L., Furlan S., Bajec M.* An expert system for detecting automobile insurance fraud using social network analysis // Expert Systems with Applications. 2011. Vol. 38, iss. 1. P. 1039–1052. https://doi.org/10.1016/j.eswa.2010.07.143
3. *Jin C., Feng Y., Li F.* Concept drift detection based on decision distribution in inconsistent information system // Knowledge-Based Systems. 2023. Vol. 279. Art. 110934. https://doi.org/10.1016/j.knosys.2023.110934
4. *Gupta P., Varshney A., Khan M., Ahmed R., Shuaib M., Alam S.* Unbalanced credit card fraud detection data: A machine learning-oriented comparative study of balancing techniques // Procedia Computer Science. 2023. Vol. 218. P. 2575–2584. https://doi.org/10.1016/j.procs.2023.01.231
5. *Pant P., Srivastava P.* Cost-sensitive model evaluation approach for financial fraud detection system // 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC). Coimbatore, India, 2021. P. 1606–1611. https://doi.org/10.1109/ICESC51422.2021.9532741
6. *Воронцов К. В.* Математические методы обучения по прецедентам (теория обучения машин) // «Машинное обучение», курс лекций. 2011. 141 с. URL: http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf (дата обращения: 22.09.2023).
7. *Fawcett T.* An introduction to ROC analysis // Pattern Recognition Letters, 2006. Vol. 27, iss. 8. P. 861–874. https://doi.org/10.1016/j.patrec.2005.10.010
8. *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. Vol. 32, iss. 5. P. 568–575. https://doi.org/10.1016/j.jksuci.2017.09.010
9. *Phua C., Alahakoon D.* Minority report in fraud detection: Classification of skewed data // ACM SIGKDD Explorations Newsletter. 2004. Vol. 6, iss. 1. P. 50–59. https://doi.org/10.1145/1007730.1007738
10. *Itri B., Mohamed Y., Mohamed Q., Omar B.* Performance comparative study of machine learning algorithms for automobile insurance fraud detection // 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS). Marrakech, Moroko, 2019. P. 1–4. https://doi.org/10.1109/ICDS47004.2019.8942277

References

1. Bao Y., Hilary G., Ke B. Artificial intelligence and fraud detection. In: Babich V., Birge J. R., Hilary G. (eds.) *Innovative technology at the interface of finance and operations*. Springer Series in Supply Chain Management, vol. 11. Cham, Springer, 2022, pp. 223–247. https://doi.org/10.1007/978-3-030-75729-8_8
2. Subelj L., Furlan S., Bajec M. An expert system for detecting automobile insurance fraud using social network analysis. *Expert Systems with Applications*, 2011, vol. 38, iss. 1, pp. 1039–1052. https://doi.org/10.1016/j.eswa.2010.07.143
3. Jin C., Feng Y., Li F. Concept drift detection based on decision distribution in inconsistent information system. *Knowledge-Based Systems*, 2023, vol. 279, art. 110934. https://doi.org/10.1016/j.knosys.2023.110934
4. Gupta P., Varshney A., Khan M., Ahmed R., Shuaib M., Alam S. Unbalanced credit card fraud detection data: A machine learning-oriented comparative study of balancing techniques. *Procedia Computer Science*, 2023, vol. 218, pp. 2575–2584. https://doi.org/10.1016/j.procs.2023.01.231
5. Pant P., Srivastava P. Cost-sensitive model evaluation approach for financial fraud detection system. *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*. Coimbatore, India, 2021, pp. 1606–1611. https://doi.org/10.1109/ICESC51422.2021.9532741
6. Voroncov K. V. Mathematical methods of learning from precedents (the theory of machine learning). *"Mashinnoe obuchenie", kurs lektsiy* ["Machine Learning", course of lectures], 2011. 141 p. Available at: http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf (accessed September 22, 2023) (in Russian).
7. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*, 2006, vol. 27, iss. 8, pp. 861–874. https://doi.org/10.1016/j.patrec.2005.10.010
8. 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, vol. 32, iss. 5, pp. 568–575. https://doi.org/10.1016/j.jksuci.2017.09.010

9. Phua C., Alahakoon D. Minority report in fraud detection. *ACM SIGKDD Explorations Newsletter*, 2004, vol. 6, iss. 1, pp. 50–59. https://doi.org/10.1145/1007730.1007738
10. Itri B., Mohamed Y., Mohamed Q., Omar B. Performance comparative study of machine learning algorithms for automobile insurance fraud detection. *2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS)*. Marrakech, Moroko, 2019, pp. 1–4. https://doi.org/10.1109/ICDS47004.2019.8942277

Поступила в редакцию / Received 19.12.2023
Принята к публикации / Accepted 26.12.2023
Опубликована / Published 29.11.2024