Торайғыров университетінің
ҒЫЛЫМИ ЖУРНАЛ

НАУЧНЫЙ ЖУРНАЛ
Торайгыров университета

ТОРАЙҒЫРОВ УНИВЕРСИТЕТІНІҢ ХАБАРШЫСЫ

Физика, математика және компьютерлік ғылымдар сериясы
1997 жылдан бастап шығады

TORAIGHYROV UNIVERSITY

ВЕСТНИК ТОРАЙГЫРОВ УНИВЕРСИТЕТА

Серия: Физика, математика и компьютерные науки
Издается с 1997 года

ISSN 2959-068X

№ 4 (2023)
Павлодар

НАУЧНЫЙ ЖУРНАЛ
ТОРАЙГЫРОВ УНИВЕРСИТЕТА

Серия: Физика, математика и компьютерные науки
выходит 4 раза в год

СВИДЕТЕЛЬСТВО
о постановке на переучет периодического печатного издания,
информационного агентства и сетевого издания
№ KZ91VPY00046988
выдано
Министерством информации и общественного развития
Республики Казахстан

Тематическая направленность
публикация материалов в области физики, математики,
механики и информатики

Подписной индекс – 76208

https://doi.org/10.48081/EKIW3862

SRSTI 50.03.05

https://doi.org/10.48081/NMPU3955

R. K. Murat[1], F. Tursunmetova[2], N. K. Nadirov[1]
[1]Astana IT University, Republic of Kazakhstan, Astana;
[2]Nazarbayev University, Republic of Kazakhstan, Astana;
*e-mail: m.raikhan@astanait.edu.kz

MULTI-CLASSIFIERS SYSTEM FOR CREDIT CARD FRAUD DETECTION

The business of issuing credit cards is extremely important to the functioning of the economy since it facilitates the use of a straightforward method of payment in a variety of contexts, such as online banking, commercial transactions, and financial dealings. Nonetheless, using credit cards is also associated with fraudulent activity and non-payment, both of which constitute a considerable danger to customers and the business as a whole. Detecting and preventing fraudulent activity involving credit cards is complex and time-consuming because of the ever-changing nature of fraudulent and expected behavior and the unequal class label and overlapping of class instances inside the data sets.

The class imbalance in credit card data sets poses a significant obstacle in detecting fraud. Biased models may result from a substantially lower number of fraudulent cases compared to non-fraudulent cases.

The study underlines the influence of oversampling and under-sampling techniques on single-classifier methods and stresses the significance of selecting an appropriate classifier algorithm. The results indicate that multi-classifier methods, particularly COPOD + RFC and IForest + RFC, can substantially improve credit card fraud detection compared to single-classifier methods. These results demonstrate the potential advantages of integrating multiple unsupervised and supervised learning algorithms to improve credit card fraud detection while decreasing false positives.

Overall, the study emphasizes the significance of employing a combination of machine learning techniques to resolve the difficulties of credit card fraud detection. The proposed method can assist credit card companies in accurately and efficiently identifying fraudulent activities, thereby reducing the risk of financial loss and enhancing customer confidence.

Keywords: fraud, multi-classifier methods, detecting, single-classifier methods, transactions, accuracy, fraudulent activity, the combination of machine learning, credit card.

Introduction

In recent years, digital banking has evolved as an essential element of the modern financial system. It has been utilized by numerous industries, including banks, e-commerce, and personal finance, because of its reliability and regularity [1]. On the other hand, a spike in fraudulent transactions has coincided with a rise in digital banking usage. As a defense mechanism against this fraudulent activity, banks have created several various fraud detection systems [1]. The ongoing development of new methods for committing digital fraud has necessitated the implementation of advanced fraud detection systems by financial institutions [1].

Machine learning algorithms have emerged as an effective approach to fraud detection [2]. These models have advantages over traditional rule-based systems as they can analyze vast amounts of data and identify subtle patterns indicative of fraudulent activity. Unlike rule-based systems, machine learning models can adapt and update their algorithms to respond to newly identified fraudulent schemes [2].

The technique based on machine learning is more adept at keeping up with the constantly evolving strategies employed by fraudsters. While a combination of rule-based and machine-learning approaches may still be necessary, the focus should be on further developing sophisticated machine-learning methods to overcome the limitations of traditional rule-based systems. Continuous research is crucial in identifying novel fraud detection techniques and addressing the challenges of machine learning algorithms [2–3].

Worldwide, fraud using credit cards is currently increasing, and it can have significant repercussions for individuals, organizations, and financial institutions in terms of financial loss and reputation. As reported by the Nilson Report, worldwide fraud involving credit card losses hit $27.85 billion in 2019, and these losses are anticipated to continue to rise over the next few years. [4]

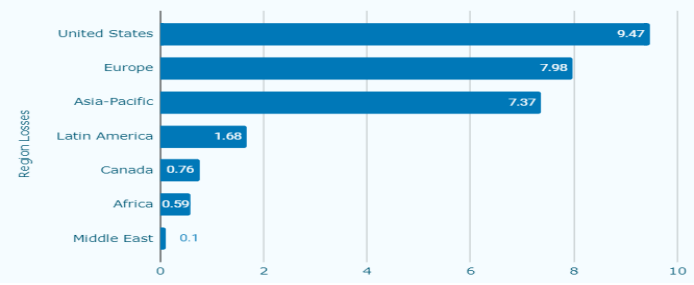

Figure 1 – Credit Card Fraud Losses by Region (in billions USD)

The data presented in Figure 1 provides an overview of each region's credit card fraud losses, expressed in billions of U.S. dollars. The table lists many regions, including the Middle East, Canada, the United States, Asia- Pacific, South America, Europe, and Africa. According to the data shown in the graph, the United States experienced the most credit card fraud losses in 2020, totaling USD 9.47 billion. Europe ranked second with losses totaling 7,98 billion USD, followed by Asia-Pacific with losses totaling 7,37 billion USD. The losses in Latin America were 1.68 billion U.S. dollars. In Canada, they were 0.76 billion U.S. dollars. In Africa, they were 0.59 billion U.S. dollars; in the Middle East, they were 0.1 billion U.S. dollars. This table's data is essential for comprehending the geographic distribution of credit card fraud losses and identifying which locations may require additional measures to prevent and detect fraud.

In Kazakhstan, online fraudulent transactions are a prevalent issue, with criminals using social engineering tactics to obtain personal information. The National Bank of Kazakhstan reported 7,151 incidents of bank card fraud in the first half of 2020, resulting in a total loss of KZT 3.3 billion [5]. To combat this, financial institutions and the government have implemented measures such as fraud detection systems, biometric verification, and consumer education.

However, fraudsters constantly modify their techniques, posing an ongoing challenge to financial institutions and law enforcement agencies. Continuous attention and innovation are necessary to ensure secure payment systems and protect consumers and businesses from the negative impacts of credit card fraud [6].

To address this problem, this article aims to develop a multi-classifier system for detecting credit card fraud. The system will utilize machine learning techniques to identify and prevent fraudulent transactions. It will be trained on large datasets to detect subtle patterns of fraudulent behavior and adapt to new fraud schemes by updating its models accordingly.

Problem statement

Data mining techniques like machine learning, neural networks, artificial intelligence, and frequency mining can be used to prevent financial fraud, including credit card fraud [7–9]. However, these techniques face obstacles such as restricted access to data for testing, limited information in datasets, and security concerns that limit the exchange of ideas [8].

Researchers have developed solutions to address these obstacles, such as the hybrid fuzzy logic model that combines machine learning and fuzzy logic to enhance anti-fraud precision [9]. In this study, the focus is on credit card fraud detection, and the main objectives are to address the issues of imbalanced class distribution and overlapping class samples.

To achieve these objectives, a Multi-Classifiers System will be created, utilizing machine learning classifiers and techniques like ensemble learning, feature selection, and class balancing to improve accuracy and handle overlapping class samples and imbalanced class distribution [10–11].

The goal is to build a robust and effective fraud detection system capable of enhancing the accuracy and dependability of credit card fraud detection systems, helping businesses and financial institutions safeguard themselves and their customers from the negative repercussions of fraud.

Materials and Methods

Machine learning algorithms, such as neural networks, Decision Trees, Support Vector Machines (SVM), and Naive Bayes, have shown their effectiveness in detecting fraudulent transactions by identifying patterns and anomalies in large datasets. These algorithms outperform humans in automated fraud detection, saving time and reducing errors. However, the success of these models relies on high-quality training and validation data. Decision Trees excel in financial fraud detection, but they can struggle with imbalanced class distribution or overlapping samples, which researchers address through techniques like class underrepresentation and overrepresentation. Linear and logistic regression models are commonly used, but their performance may suffer from unbalanced classes and overlapping samples, which can be improved by combining them with other machine-learning approaches. The Local Outlier Factor algorithm detects fraud by identifying deviations from the average, although it faces limitations with sparse fraudulent data. Naive Bayes is reliable but less accurate with imbalanced or overlapping data, yet it remains popular due to its simplicity and speed in fraud detection.

C4.5, a decision tree method used for fraud detection, has limitations due to imbalanced class distribution and overlapping class samples, but strategies like under-sampling, oversampling, and cost-sensitive learning can mitigate these challenges while maintaining its effectiveness in spotting fraudulent transactions.

SVM is a powerful ML model for fraud detection, particularly for high-dimensional and complex datasets. However, imbalanced class distribution and overlapping class samples are challenges that can affect its accuracy. Various techniques like oversampling, under-sampling, and feature selection can address these issues. Despite these limitations, SVMs have shown high accuracy in detecting fraudulent transactions, with some studies reporting 99.8 % accuracy. Ensemble techniques can also enhance fraud detection efficacy. Additional research is needed to build more effective algorithms to manage overlapping class samples.

It is essential to evaluate the effectiveness of fraud detection models to identify in- stances of fraudulent behavior precisely. Based on the findings of an exhaustive Literature Review, it has been determined that the performance of various models for fraud detection varies significantly. Unbalanced and overlapping data sets pose obstacles that a few models can effectively address.

In detecting fraud, the effectiveness of singular classification models like decision trees, the likelihood of occurrence function (LOF), Naive Bayes, C4.5, and support vector machines (SVMs) vary. Critical evaluative criteria for these models include the data set, precision, constraints, and real-time efficacy. While each model has advantages and disadvantages, combining them into a multi-classifier system may effectively overcome their shortcomings and improve fraud detection accuracy. In conclusion, integrating multiple models may improve fraud detection performance, allowing for more accurate identification of fraudulent behavior.

Dataset Description

The chapter on data set description and preprocessing is vital to every data analysis endeavor. The focus of this chapter is on detailing the obtained raw data and the actions taken to clean, transform, and prepare it for future analysis. The quality and precision of the study depend heavily on the data collection quality and the preprocessing processes' efficacy. Thus, it is essential to describe the data set thoroughly, including its size, characteristics, and any missing or incorrect data. In addition, all preprocessing methods, including missing value management, outlier detection, and normalization, must be exhaustively documented and justified. By providing a clear and straightforward description of the data set and preprocessing processes, the future analysis will be more accurate and dependable, resulting in more solid results.

These datasets are valuable resources for training and assessing machine learning models for fraud detection and credit risk assessment and have been widely utilized in research studies. The Credit Card Fraud (CCF) datasets is extensively utilized in finance and credit card fraud detection.

The Credit Card Fraud (CCF) dataset, released in 2016 by European cardholders, has been a valuable resource for fraud detection research. It provides transactional data that has been extensively used to develop and test various

machine-learning models and fraud detection techniques. The availability of this dataset has facilitated the study and evaluation of novel strategies and methodologies for detecting fraudulent credit card transactions. The CCF dataset has played a significant role in advancing fraud detection capabilities [12].

The dataset contains 30 attributes, 28 of which are numerical and 2 of which are categorical. PCA was used to change the numerical characteristics to keep the data private.

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V24	V25	V26	V27	V28	Amount	Class
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.066928	0.128539	-0.189115	0.133558	-0.021053	149.62	0
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288	-0.339846	0.167170	0.125895	-0.008983	0.014724	2.69	0
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	0.247998	0.771679	0.909412	-0.689281	-0.327642	-0.139097	-0.055353	-0.059752	378.66	0

Figure 2 – Dataset visualization

The two defining characteristics are «Time» and «Amount» The «Time» attribute indicates the number of seconds between each transaction and the first transaction in the collection. The «Amount» field displays the transaction total in euros. Table 1 shows a detailed description of the features in the CCF dataset. «Class» is the target variable identifying whether a transaction is fraudulent. Several 1 indicates a fraudulent transaction, while a value of 0 suggests a transaction that is not fraudulent.

Table 1 – Credit Card Fraud (CCF) dataset features

Feature	Description	Type
Time	Number of seconds elapsed	Categorical
V1-V28	PCA transformed features	Numerical
Amount	Transaction amount	Categorical
Class	Fraudulent or non-fraudulent	Categorical

Preprocessing The data supplier has already preprocessed the dataset, and the dataset contains no missing values.

Limitation The CCF dataset contains 492 fraudulent transactions out of a total of 284,807 transactions, resulting in a class imbalance issue. To address this imbalance, techniques such as oversampling and under-sampling can be employed to balance the dataset before training machine learning models.

The CCF dataset includes anonymized features V1 to V28, derived from PCA transformation for data privacy. Although the specific attributes and their meanings are not disclosed for confidentiality reasons, researchers can utilize these features as inputs for machine learning models aimed at detecting fraudulent transactions.

The class imbalance issue in the CCF dataset, with a small number of fraudulent transactions compared to valid transactions, poses challenges for effectively detecting fraud using machine learning models.

Nevertheless, the CCF dataset has become a widely used benchmark for fraud detection tasks in the academic community. Its availability has significantly contributed to the development and testing of novel fraud detection techniques, ultimately enhancing the security of credit card transactions.

Methodology

Detecting fraud in credit card transactions takes multiple steps of data preparation and analysis using various machine-learning algorithms. Data preparation includes cleaning and altering the data to make it usable for machine learning techniques. SMOTE (Synthetic Minority Over-sampling Technique) is a typical oversampling approach used in the preprocessing phase to handle the problem of imbalanced data. Several metrics, including accuracy, precision, recall, F1 - score, and AUC-ROC, are used to assess the performance of each model (Area Under the Receiver Operating Characteristic curve).

Finally, a multi-classification model will implement a multi-classifier system based on a sequential combination approach. This will allow us to integrate the outputs from multiple classifiers to produce a more robust and accurate prediction.

The methodology intends to give a comprehensive approach to detecting credit card fraud, considering the data's uneven nature and the necessity for high precision in recognizing fraudulent transactions.

Single classifiers models with datasets

We must first partition our data. The train test split function from the Scikit-Learn module is used in the code fragment to split the data. The initial dataset is divided into two sets, X and y, representing the input features and the output label, respectively. The data is separated into two sets: training and testing, with a test size of 30 % and a random state of 42. This means that 70 % of the data will be used to train the model and 30 % to evaluate its performance. The training set will be used to fit the model, while the testing set will be used to evaluate the model's performance.

```
Splitting for the first dataset:
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3)

Splitting for the second dataset:
X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.2,random_state=42)
```

Figure 3 – Code of splitting the dataset

For the first dataset, we will use several single-class classification models, including Logistic Regression, Naive Bayes Classifier, Random Forest Classifier, and Support Vector Machine (SVM).

Logistic Regression is a statistical model used to study the relationship between a dependent variable and one or more independent variables by estimating probability using a logistic function. Any real-valued number is mapped to a probability between 0 and 1 by the logistic function (sigmoid function). Based on the input features, the model computes the probability ratio for the occurrence of an event, such as fraud detection. The outcome of the logistic Regression is a binary classification model with a threshold for classifying a transaction as fraudulent or not.

SVMs are commonly used in machine learning to handle classification and regression problems. The program projects the samples of a given data set with n attributes into an n-dimensional space where each sample is represented as a point, and the value of each attribute corresponds to a specific position. As demonstrated in Figure 4, the SVM algorithm next calculates the best hyper-plane that splits the data into classes.

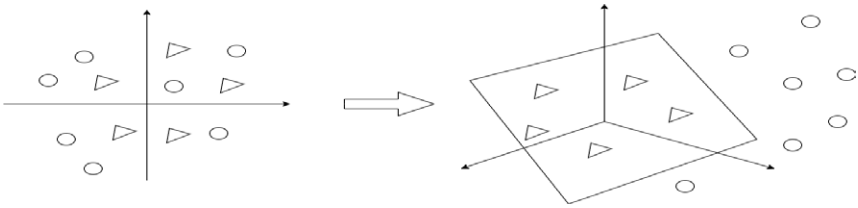

Figure 4 – Support Vector Machines (SVM) algorithm

The technique works by creating several decision trees using a random subset of the features and a random selection of the data. For each tree, a bootstrap sample of the training data is used to produce a new dataset of the same size as the original, but some of the samples are repeated and others are left out. This is referred to as bagging (bootstrap aggregating).

To predict the class of a new sample, the algorithm sends the sample through all the trees and combines the results by taking the majority vote. The final output is the class with the highest number of votes.

A random forest decision tree is generated using a collection of features and data. This randomization helps to reduce overfitting and increase the generalization of the model. Using techniques such as cross-validation, the algorithm's hyperparameters, such as the number of trees and the maximum depth of each tree, can be changed to improve the model's performance on the validation set.

The Multi-Classifiers System, often known as MCS, is an ensemble technique that enhances the accuracy of predictions by merging the findings of multiple classifiers into a single model.

Before going on to more complex models, simpler classifiers that produce less precise results are often applied first. In the sequential combination strategy, two or more separate classifiers process the input data by collaborating sequentially. Following this, the output of the initial classifier serves as the input for subsequent classifiers. The boosting method is an example of a learning algorithm that uses sequential combining. Figure 5.

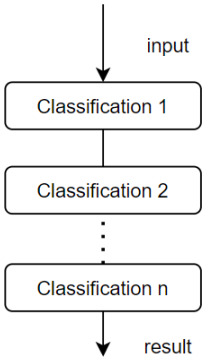

Figure 5 – Sequential combination

In this research project, a sequential multi-classifier combination strategy will be implemented to improve the accuracy of fraud detection models. The concept of merging supervised and unsupervised models to improve the identification of fraudulent transactions will also be investigated. Support Vector Machine (SVM) and Random Forest Classifier are the base classifiers employed in this work (RFC).

Combining the COPOD algorithm, an unsupervised outlier detection methodology, with the RFC model constitutes the first combination method. COPOD identifies data outliers and eliminates them from the training set, enhancing the RFC model's performance.

The second combination approach will entail combining the COPOD and SVM algorithms. Like the first method, COPOD is used to preprocess the data by finding and removing outliers, improving the SVM model's performance.

Using the Isolation Forest (IForest) algorithm, another unsupervised outlier detection tool, in conjunction with SVM, constitutes the third combination method. The IForest algorithm functions by isolating data outliers through recurrent dataset segmentation. By deleting these outliers, the SVM model's performance can be enhanced.

All unsupervised models (COPOD and IForest) will be combined with RFC in the fourth combination approach. Before feeding the data into the RFC model, this

strategy seeks to discover and eliminate data outliers by combining the strengths of two unsupervised outlier detection algorithms.

In the final combination procedure, all unsupervised models (COPOD and IForest) will be combined using SVM. This method tries to enhance the performance of SVM by reducing data outliers.

By applying sequential multi-classifier combination approaches with supervised and unsupervised models, we hope to boost the accuracy of fraud detection models and the capacity to identify fraudulent transactions.

Results and Discussion

The chapter on findings will compare the performance of single and multi-classification models for credit card fraud detection.

Single classifier's performance with the dataset

The table demonstrates the performance metrics of four distinct single classifiers on the Credit Card Fraud (CCF) dataset. Each model is evaluated based on its accuracy, precision, recall, and F1 score.

Table 2 – Single classifiers performance

Dataset	Model	Accuracy	Precision	Recall	F1 score
CCF	Logistic Regression	99.92	0.87	0.63	0.73
CCF	Naive Bayes	97.8	0.65	0.84	0.70
CCF	Random Forest	99.93	0.77	0.86	0.81
CCF	Support Vector Machine	99.94	0.83	0.83	0.83

The logistic regression model had the highest accuracy for the CCF dataset, at 99.92 %, but its recall was relatively low, at 0.63, showing that it struggled to identify fraudulent transactions correctly. The Naive Bayes model's accuracy was lower, at 97.8 %, but its recall performance was superior, with a score of 0.84. The Random Forest model achieved the most excellent F1 score of 0.81, indicating that its precision and recall were well-balanced. The Support Vector Machine model achieved a balanced performance with an accuracy of 99.94 % and an F1 score of 0.83.

Overall, these results indicate that the selection of a single classifier can substantially impact a model's performance and that other models may be better suited for particular datasets and aims.

Multi-classification model performance

The table depicts the evaluation of the performance of various single-classifier models trained on Credit Card Fraud (CCF)

Table 3 – Multi classifiers performance

Dataset	Model	Accuracy	Precision	Recall	F1 score
CCF	COPOD + RFC	99.95	0.92	0.78	0.85
CCF	COPOD + SVC	99.96	0.93	0.89	0.91
CCF	IForest + RFC	99.94	0.92	0.76	0.83
CCF	IForest + SVC	99.93	0.81	0.82	0.82
CCF	All Unsupervised + RFC	99.94	0.88	0.78	0.83
CCF	All Unsupervised + SVC	99.92	0.86	0.73	0.73

The performance metrics evaluated are accuracy, precision, recall, and the F1 score. COPOD + RFC, COPOD + SVC, IForest + RFC, IForest + SVC, All Un- supervised + RFC, and All Unsupervised + SVC are the models evaluated for the CCF dataset.

Conclusions

Based on the tables presented, it is evident that several machine-learning models have been employed to detect fraudulent behavior with varying degrees of success. Logistic regression, Naive Bayes, random forest, and Support Vector Machines (SVM) have all shown high accuracy in identifying fraud in credit card transactions. However, imbalanced datasets can negatively impact the precision of some models, such as Naive Bayes. In order to overcome these challenges, researchers have proposed combining multiple models and using unsupervised learning techniques. The combination of COPOD and random forest, and COPOD and SVM, have shown excellent performance in detecting fraud, with high accuracy, precision, recall, and F1 scores.

Overall, the use of machine learning models and unsupervised learning techniques has the potential to improve the detection of fraudulent behavior significantly. However, it is crucial to consider the limitations and challenges associated with these models, such as imbalanced datasets and overlapping class samples. Further research is needed to develop more robust and effective models that can address these challenges and provide greater accuracy and reliability in detecting fraudulent behavior

REFERENCES

1 **Litan, A.** The evolving nature of digital banking fraud. Forbes. 2020, [Electronic resource]. – https://www.forbes.com/sites/forbestechcouncil/2020/01/06/the-evolving-%20nature-of-digital-banking-fraud/?sh=692c479a78e9.

2 **Ankur, R.** Comparative Analysis of Various Classification Algorithms in the Case of Fraud Detection // International Journal of Engineering Research& Technology. – 2017. – 6(9). http://dx.doi.org/10.17577/IJERTV6IS090047.

3 J E T A. Supervised Machine Learning Algorithms: Classification and Comparison // International Journal of Computer Trends and Technology (IJCTT). – 2017. – 48(3), 128–138 p. http://dx.doi.org/10.14445/22312803/IJCTT-V48P126.

4 Nilson Report. The Nilson Report. – [Electronic resource]. – https://www.nilsonreport.com/upload/content_promo/The_Nilson_Report_12-15-2020.pdf. 2020.

5 National Bank of Kazakhstan. Financial Stability Review. 2020. – [Electronic resource]. – https://nationalbank.kz/files/fsr/fsr_2020_2.pdf.

6 Euromonitor International. Credit cards in Kazakhstan. – 2020. – [Electronic resource]. – https://www.euromonitor.com/credit-cards-in-kazakhstan/report.

7 **Bahnsen, A. C., Villegas, S., Aouada, D., & Ottersten, B. E.** Fraud Detection by Stacking Cost-Sensitive Decision Trees. Data Science for Cyber-Security, 2017. – 251–266 p. [Electronic resource]. – https://doi.org/10.1142/9781786345646012.

8 **Navanshu, K., Saad Yunus, S.** Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models // International Journal of Pure and Applied Mathematics. – 2018. – 118(20). – P 825–838. [Electronic resource]. – https://www.acadpubl.eu/hub/2018-118-21/articles/21b/90.pdf.

9 **Sadgali, I., Sael, N., Benabbou, F.** Performance of machine learning techniques in the detection of financial frauds // Procedia Computer Science, 2019. – 148, 45–54. https://doi.org/10.1016/j.procs.2019.01.007

10 **Kalid, S. N., Ng, K. H., Tong, G.-K., Khor, K. C.** A Multiple Classifiers System for Anomaly Detection in Credit Card Data With Unbalanced and Overlapped Classes. // IEEE Access. – 2020. – 8, 1–1 p. http://dx.doi.org/10.1109/ACCESS.2020.2972009.

11 **Jenny, D., Elena, K.** Identification of non-typical in international transactions on bank cards of individuals using machine learning methods. Procedia Computer Science. – 2021. – 190. – 178–183 p. https://doi.org/10.1016/j.procs.2021.06.023.

12 **Dal Pozzolo, A., Caelen, O., Johnson, R. A., Bontempi, G.** Cali- brating Probability with Undersampling for Unbalanced Classification. In 2015 IEEE

Symposium Series on Computational Intelligence, Cape Town, South Africa : IEEE. – 2015. – 159–166 p. https://doi.org/10.1109/SSCI.2015.33

Accepted for publication 15.12.23.

Р. К. Мұрат[1], Ф. Турсунметова[2], Н. Қ. Нәдіров[1]
[1]Astana IT University, Қазақстан Республикасы, Астана қ.;
[2]Nazarbayev University, Қазақстан Республикасы, Астана қ.
Басып шығаруға 15.12.23 қабылданды.

КРЕДИТТІК КАРТАЛАРДЫ АНЫҚТАУҒА АРНАЛҒАН КӨП-КЛАССИФИКАТОРЛАР ЖҮЙЕСІ

Несие карталарын шығару бизнесі экономиканың жұмыс істеуі үшін өте маңызды, өйткені ол онлайн-банкинг, коммерциялық транзакциялар және қаржылық мәмілелер сияқты әртүрлі контексттерде тікелей төлем әдісін пайдалануға мүмкіндік береді. Дегенмен, несие карталарын пайдалану алаяқтық әрекетпен және төлем жасамаумен де байланысты, олардың екеуі де клиенттерге және тұтастай алғанда бизнеске айтарлықтай қауіп төндіреді. Несие карталарына қатысты алаяқтық әрекетті анықтау және алдын алу, алаяқтық және күтілетін мінез-құлықтың үнемі өзгеріп тұратын сипатына және тең емес сынып белгісіне және деректер жиынындағы сынып даналарының қабаттасуына байланысты күрделі және уақытты қажет етеді.

Несие карталарының деректер жинағындағы сыныптық теңгерімсіздік алаяқтықты анықтауда айтарлықтай кедергі жасайды. Біржақты модельдер алаяқтық емес істермен салыстырғанда алаяқтық істер санының айтарлықтай аз болуынан туындауы мүмкін.

Зерттеу бір классификаторлық әдістерге артық іріктеу және төмен таңдау әдістерінің әсерін көрсетеді және сәйкес жіктеуіш алгоритмін таңдаудың маңыздылығын атап көрсетеді. Нәтижелер көп жіктеуіш әдістері, атап айтқанда COPOD + RFC және IForest + RFC, бір жіктеуіш әдістермен салыстырғанда несие картасының алаяқтығын анықтауды айтарлықтай жақсарта алатынын көрсетеді. Бұл нәтижелер жалған позитивтерді азайта отырып, несие картасының алаяқтығын анықтауды жақсарту үшін бірнеше бақыланбайтын және бақыланатын оқыту алгоритмдерін біріктірудің ықтимал артықшылықтарын көрсетеді.

Жалпы алғанда, зерттеу несие картасының алаяқтығын анықтау қиындықтарын шешу үшін машиналық оқыту әдістерінің комбинациясын қолданудың маңыздылығын атап көрсетеді. Ұсынылған әдіс несие картасы компанияларына алаяқтық әрекеттерді дәл және тиімді анықтауға көмектесе алады, осылайша қаржылық жоғалту қаупін азайтады және тұтынушылардың сенімін арттырады.

Кілтті сөздер: алаяқтық, көп жіктеуіш әдістері, анықтау, бір жіктеуіш әдістері, транзакциялар, дәлдік, алаяқтық әрекет, машиналық оқыту комбинациясы, несие картасы.

Р. К. Мұрат[1], Ф. Турсунметова[2], Н. Қ. Нәдіров[1]
[1]Astana IT University, Республика Казахстан, г. Астана;
[2]Nazarbayev University, Республика Казахстан, г. Астана
Принято к изданию 15.12.23.

СИСТЕМА МУЛЬТИКЛАССИФИКАТОРОВ ДЛЯ ВЫЯВЛЕНИЯ МОШЕННИЧЕСТВА С КРЕДИТНЫМИ КАРТАМИ

Бизнес по выпуску кредитных карт чрезвычайно важен для функционирования экономики, поскольку он облегчает использование простого метода оплаты в различных контекстах, таких как онлайн банкинг, коммерческие операции и финансовые операции. Тем не менее использование кредитных карт также связано с мошенническими действиями и неплатежами, которые представляют значительную опасность для клиентов и бизнеса в целом. Обнаружение и предотвращение мошеннических действий с использованием кредитных карт является сложным и требует много времени из-за постоянно меняющегося характера мошеннического и ожидаемого поведения, а также неравных меток классов и перекрытия экземпляров классов в наборах данных.

Дисбаланс классов в наборах данных кредитных карт представляет собой серьезное препятствие для обнаружения мошенничества. Предвзятые модели могут быть результатом значительно меньшего количества случаев мошенничества по сравнению с делами, не связанными с мошенничеством.

В исследовании подчеркивается влияние методов пере дискретизации и недостаточной выборки на методы с одним классификатором и подчеркивается важность выбора соответствующего алгоритма классификатора. Результаты

показывают, что методы с несколькими классификаторами, особенно COPOD + RFC и IForest + RFC, могут существенно улучшить обнаружение мошенничества с кредитными картами по сравнению с методами с одним классификатором. Эти результаты демонстрируют потенциальные преимущества интеграции нескольких алгоритмов обучения без учителя и с учителем для улучшения обнаружения мошенничества с кредитными картами при одновременном снижении количества ложных срабатываний.

В целом, исследование подчеркивает важность использования комбинации методов машинного обучения для решения проблем обнаружения мошенничества с кредитными картами. Предлагаемый метод может помочь компаниям, выпускающим кредитные карты, точно и эффективно выявлять мошеннические действия, тем самым снижая риск финансовых потерь и повышая доверие клиентов.

Ключевые слова: мошенничество, много классификационные методы, обнаружение, одно классификационные методы, транзакции, точность, мошенническая деятельность, сочетание машинного обучения, кредитная карта.

Теруге 15.12.2023 ж. жіберілді. Басуға 29.12.2023 ж. қол қойылды.
Электрондық баспа
7,50 Mb RAM
Шартты баспа табағы 10,01. Таралымы 300 дана. Бағасы келісім бойынша.
Компьютерде беттеген: Е. Е. Калихан
Корректор: А. Р. Омарова
Тапсырыс № 4178

Сдано в набор 15.12.2023 г. Подписано в печать 29.12.2023 г.
Электронное издание
7,50 Mb RAM
Усл.печ.л. 10,01. Тираж 300 экз. Цена договорная.
Компьютерная верстка Е. Е. Калихан
Корректор: А. Р. Омарова
Заказ № 4178

«Toraighyrov University» баспасынан басылып шығарылған
«Торайғыров университеті» КЕ АҚ
140008, Павлодар қ., Ломов к., 64, 137 каб.

«Toraighyrov University» баспасы
«Торайғыров университеті» КЕ АҚ
140008, Павлодар қ., Ломов к., 64, 137 каб.
+7(718)267-36-69

e-mail: kereku@tou.edu.kz
www.vestnik.tou.edu.kz
https://vestnik-pm.tou.edu.kz/