National Research University Higher School of Economics

As a manuscript

Vorobyev Ivan Aleksandrovich

Research on the development of methods for combating fraud in financial organizations using machine learning

DISSERTATION SUMMARY

for the purpose of obtaining academic degree

Doctor of Philosophy in Technical Sciences

Academic supervisor:
Candidate of Technical Sciences
Los Aleksei Borisovich

Moscow — 2024

Glossary

Fraud monitoring is an automated system for analyzing transactions aimed at preventing fraudulent activity and protecting users' funds and personal data

The effectiveness of fraud detection method is a set of specific characteristics that allow evaluating the ability of the method to prevent fraudulent actions. Within the scope of research, this term can be applied to a specific algorithm or fraud monitoring system.

Machine Learning - methods based on the identification of empirical regularities in data. Mathematical statistics, numerical methods, mathematical analysis, optimization methods, probability theory, and graph theory are used to develop such methods.

Feature – is a characteristic or attribute that describes an object or data. Features are used to represent information about objects and serve as the basis for training machine learning models.

Data classification - is the process of assigning objects to different categories or classes based on specific features.

Data enrichment - is the process of adding information or attributes to existing data. This process may involve the use of external data sources, data analysis, and data transformation.

Banking operation - is a financial transaction that occurs between a bank and its clients. It includes various types of actions, such as money transfers, opening and closing accounts, issuing loans, debt repayment, and more.

Insurance claim - is a request or demand made by the policyholder to the insurance company in the event of an insured incident. Within the scope of an insurance claim, the policyholder contacts the insurance company to seek compensation for losses, coverage of expenses, or payment of insurance benefits according to the terms of the insurance contract.

1. Introduction

1.1. Problem Statement and Research Significance

The dissertation research focuses on the problem of enhancing the resilience of financial organizations against fraudulent attacks, which target both the assets of clients and the assets of the organizations themselves. Such attacks are referred to as financial fraud or simply fraud. In most cases, they are carried out by malicious individuals with the aim of illegally obtaining funds from clients or organizations. As stated in [1], financial fraud is a significant problem as it causes damage to both the economy of organizations and the economy of the state. Therefore, minimizing the consequences of fraudulent activities is one of the priority tasks for key participants in the financial sector: banks and insurance companies.

The development of data storage and processing technologies has enabled financial organizations to keep track of transactions, customer data, and other information in internal databases. It has become possible not only to accumulate this information but also to utilize big data and artificial intelligence (AI) technologies for automated decision-making in various processes, including the detection of fraudulent activities [2]. In addition, retrospective analysis of events based on data has become widely applied in the field of information security [3]. The majority of financial institutions have started implementing automated systems for transaction analysis, commonly known as fraud monitoring systems. Their main purpose is to identify unlawful actions against clients or the organizations themselves.

In the study [4] in section 6, financial fraud is classified into several types depending on the industry: banking, insurance, telecommunications, etc. Each industry has subtypes depending on the method of committing fraud. The proposed study examines the possibility of applying machine learning methods to enhance the effectiveness of combating fraud in two subtypes: banking with the use of cards in e-commerce and in auto insurance.

The main object of research is machine learning methods, the adaptation and integration of which into anti-fraud processes will allow financial organizations to be more resilient to risks associated with fraudulent activities.

The strategies employed by fraudsters in the examined subtypes may vary, but the challenges faced in applying machine learning methods are similar. These challenges include the class imbalance between fraudulent transactions and legitimate ones, as discussed, for example, in study [5], as well as the low interpretability of modeling results when using these methods [6].

Additional complexity in detecting fraud is introduced by the behavior of fraudsters themselves. Over time, in their attempts to circumvent the security measures of banks or insurance companies, the behavior of fraudsters changes, and the response of experts in financial organizations does not always keep up with these changes. On the other hand, security experts may have their own biases when

it comes to defining fraud indicators. For example, experts may assess the same fraud case differently. As a result, legitimate incidents may be mistaken for fraudulent ones because the expert makes an error due to unfamiliarity with a new scheme. For these reasons, in the data on fraud cases to which machine learning methods are applied, a phenomenon called "concept drift" [7] may occur, leading to the instability of machine learning models over time.

The presence of these issues and significant losses from threats associated with the activities of fraudsters highlights the relevance of the task of improving the effectiveness of algorithms created to prevent fraud.

1.2. Aims of the Thesis Research

The main **goal** of the proposed research is to develop a method that allows for more effective detection of fraud cases in financial organizations by utilizing machine learning and transactional data.

To achieve the goal, the following tasks were formulated:

1. Development of a method for preparing data on fraud cases that allows reducing the negative impact on the quality of machine learning algorithms caused by factors such as changes in fraud scheme scenarios and subjective expert assessment.
2. Development of new transaction attributes that have a positive impact on the effectiveness of fraud detection.
3. Development of an algorithm that allows to increase the precision of the fraud monitoring system through automatic generation of decision-making rules.
4. Conducting experiments on real data to evaluate the effectiveness of fraud detection achieved through the proposed methods.

1.3. Related Work

In the period of 1980-1990, scientific research focused on fraud detection was limited to the use of simple statistical and econometric methods [8–10]. Currently, artificial intelligence, particularly machine learning methods, are increasingly being applied to solve such tasks. Fraud detection methods have become of interest to both commercial companies and the scientific community. If there were 16,000 scientific papers published on this topic in 2015, the number has increased by 1.5 times in 2021.

Fraud detection algorithms can be divided into expert-based and statistical approaches. In the expert-based approach, fraud is detected based on rules created by experts, taking into account the analysis of typical fraudulent behavior in a manual manner. In the statistical approach, statistical methods, including machine learning models, are utilized to classify transactions as fraudulent or legitimate.

Statistical algorithms, according to [11], can be divided into classification tasks, clustering tasks, and graph analysis. The first type helps to separate transactions into fraudulent and legitimate even when fraudsters disguise their

activities as legitimate ones. The advantage of the second type of algorithms, although they may perform worse in recognizing disguised cases, lies in their ability to detect new events indicating fraudulent activities that have not been encountered in historical data. Graph analysis allows for considering the relationships between objects in a dataset. These three types of statistical algorithms focus on different aspects of fraud and complement each other.

The problem of fraud detection from the perspective of machine learning is a classification problem with two non-overlapping classes [12]. The main focus of the dissertation research is to improve the quality of classification by addressing the issues of class imbalance and the variability of fraudulent behavior. Additionally, the research aims to create new feature representations of transactions (for banking fraud monitoring) and claims (for insurance fraud monitoring) using existing data.

The effectiveness of classification depends on the quality of data and features [13,14]. In the study [13], where the classification results of various methods are compared on different datasets, all methods show decreased effectiveness on datasets with a large number of non-numeric features. By creating new feature representations in the research [14], the authors achieved greater improvement in effectiveness compared to transitioning from simple and interpretable statistical models to more complex ones.

Some researchers argue for the higher effectiveness of statistical algorithms. For example, in the study [15], the effectiveness of fraud detection procedures based on expert rules is compared to the effectiveness of a neural network developed by the authors. The results demonstrate that the neural network outperforms the expert rules, detecting fraud by an order of magnitude and with higher accuracy.

However, these methods can be used complementarily. For instance, in the study [16], combining a neural network focused on anomaly detection with an expert approach yields better results than using these two approaches individually.

In the insurance industry, fraud detection is a challenging task, both through the use of expert approaches and statistical methods, including machine learning. This is emphasized in many works, including the research study [17]. An additional challenge is the limited access to fraud data. As noted in [18], there is only one comprehensive dataset available for research on fraud detection using machine learning methods. This situation hinders progress in the field of fraud detection and leads to low classification performance.

To improve classification, various approaches are used. For example, in [19], an evaluation is applied that can vary depending on the duration of the claim and utilizes natural language processing. In [20] , the authors leverage the fact that fraudsters can distort questionnaire data, and when such an anomaly is detected, the insurance company can achieve an additional effect in reducing the level of fraud. Additionally, researchers strive to reduce the number of features used for

classification and enhance result interpretability [21]. In the study [22] , the authors improve fraud detection performance in auto insurance by applying genetic algorithms. The issue of class imbalance in fraud detection in auto insurance is explored in [23].

A comparative table of research results obtained in different years for the task of fraud detection in insurance is presented in [24].

This current scientific study proposes to continue research aimed at improving the quality of classification in the task of fraud detection in the banking and insurance sectors. In this regard, a range of methods is considered, which are applied in the decision-making process regarding transactions or claims to identify fraudulent activities.

1.4. Novelty of the Research

1. For the first time, a method is proposed to improve the effectiveness in fraud detection tasks by adjusting the target class using a neural network. This allows for data balancing for the use of machine learning methods and addresses concept drift issues.
2. A new approach is proposed for combining traditional expert knowledge with machine learning in order to enhance the effectiveness of fraud monitoring systems. The method involves using composite parts of rules created by experts to generate new, more effective rules through machine learning techniques.
3. Methods for creating new features for transactions and claims that improve the quality of fraud detection have been proposed.

2. Key Results

2.1. Key Ideas to be Defended:

1. A method has been developed that addresses the issue of class imbalance when using machine learning methods, while also mitigating concept drift in the data caused by changes in fraudulent schemes or incorrect data labeling. This approach improves the separating power of the classifier by enhancing the quality of the training data. A detailed description of the method and the obtained results are published in [25].
2. An approach is proposed that enhances the effectiveness of fraud monitoring systems by creating new attributes for transactions and claims. In the banking sector, transactions are enriched by integrating customer purchase history into the training data for evaluating transfers between clients. Claims are enriched with features derived from the graph of connections between participants in insurance events. Descriptions of the approaches for creating new features are published in [26].
3. A method has been developed, based on machine learning approaches, that allows financial organizations to reduce false positives in their fraud monitoring systems by implementing automatically generated decision-making algorithms for transaction evaluation. The method has been published in [27].
4. Methodologies for conducting experiments have been developed to assess the effectiveness of the proposed methods. A series of experiments has been conducted, the results of which demonstrate an improvement in the quality of fraud detection in the financial sector when using the developed approaches.

2.2. Personal Contribution to the Ideas to be Defended

During the dissertation research, the author developed an approach that enhances the effectiveness of applying machine learning methods by adjusting the expert labeling of data.

Additionally, a process for generating decision-making algorithms has been proposed, which involves the joint application of expert and statistical approaches and allows for improved accuracy in fraud classification without sacrificing interpretability. Within the research framework, an algorithm for constructing an insurance claims graph and a method for extracting new data from it to enhance the effectiveness of machine learning methods have also been proposed. The study demonstrated that enriching banking transaction data with customer purchase history improves the classification quality when using machine learning methods. A series of experiments was conducted, and the results showed that the proposed approaches have the potential to enhance the effectiveness of applying machine learning methods and can be a valuable tool in the field of fraud detection, where accurate data classification is required.

3. Publications and Approbation of Research

3.1. First-tier publications

1. Vorobyev I., Krivitskaya A. Reducing False Positives in Bank Anti-fraud Systems Based on Rule Induction in Distributed Tree-based Models // Computers and Security. 2022. Vol. 120, http://doi.org/10.1016/j.cose.2022.102786 (**Scopus, Q1**)
2. Vorobyev I. Fraud risk assessment in car insurance using claims graph features in machine learning // Expert Systems with Applications. 2024. Vol. 251, http://doi.org/10.1016/j.eswa.2024.124109 (**Scopus, Q1**)

3.2. Second-tier publications

1. Vorobyev I. A. ML methods for assessing the risk of fraud in auto insurance // Izvestiya of Saratov University. Mathematics. Mechanics. Informatics. 2024 (**Scopus**)
2. Festa Y. Y., Vorobyev I. A. A Hybrid Machine Learning Framework for E-commerce Fraud Detection // Model Assisted Statistics and Applications. 2022. Vol. 17. No. 1. P. 41-49, http://doi.org/10.3233/MAS-220006, (**Scopus**)

3.3. Reports at conferences and seminars

1. 2023, Interuniversity Scientific and Technical Conference of Students, Postgraduates, and Young Professionals named after E.V. Armensky (Moscow). Presentation: Research on the application of machine learning methods in detecting fraudulent activities regarding bank customers during transaction confirmation.
2. 2023, XII Congress of Young Scientists ITMO (Saint Petersburg). Presentation: Interpretability of machine learning models and the class imbalance problem in risk reduction tasks for credit-financial organizations.
3. 2023, XII International Scientific and Practical Conference "Mathematical and Computer Modeling in Economics, Insurance, and Risk Management" (Saratov). Presentation: ML methods for assessing the risk of fraud in auto insurance.
4. 2021, International Conference on Data Analytics and Computational Techniques, ICDACT-21 (Bhopal). Presentation: A Hybrid Machine Learning Framework for E-commerce Fraud Detection.
5. 2021, International Congress "Modern Problems of Computer and Information Sciences", VI International Scientific Conference on Convergent Cognitive-Information Technologies (Moscow). Presentation: The application of artificial intelligence for improving the efficiency of transactional fraudmonitoring.
6. 2020, XI International Forum "Fighting Fraud in the High-Tech Sector. Antifraud Russia - 2020" (Moscow). Presentation: Anti-fraud in Sber Acquiring.

4. Work content

The results of the dissertation research are presented in the following sections:

1. The use of machine learning methods in fraud detection tasks and approaches to assessing their effectiveness.
2. Architectures of fraud monitoring systems and potential areas for their improvement.
3. Preparation of data for training classifiers of fraudulent transactions and claims.
4. Generation of new rules to improve the quality of the fraud monitoring system.
5. Methodology for conducting experiments and research.

4.1. The use of machine learning methods in fraud detection tasks and approaches to assessing their effectiveness

Fraud is understood as the situation of theft of funds from a client or financial institution by professional fraudsters.

Fraud, according to [11], has the following specific features:

1) compared to the frequency of legitimate operations, fraud occurs rarely;
2) fraud is carefully thought out and planned;
3) fraudsters try to disguise their activity as legitimate;
4) the behavior of fraudsters changes over time;
5) fraudsters often operate in organized groups.

The assessment of transactions for fraud using machine learning methods is carried out using historical data. Each transaction has its own set of features, and if it has been processed by an expert or if the client has been asked to confirm its legitimacy, there is an answer to whether it contains fraud indicators. This allows us to reduce the task of fraud detection to a case of learning from precedents [12]. Specifically, we will consider a classification problem with two non-overlapping classes. The resulting decision function (referred to as the model or classifier) will then be used to evaluate a specific transaction for the presence of fraud based on its feature description (referred to as features).

Let us denote the set of evaluated transactions as X, and the set of answers to the question "is the transaction fraudulent?" as Y. Pairs of "transaction-answer" $(x_i,\ y_i)$ will be referred to as precedents. Let $\{x_1, \dots, x_l\} \subset X$ be a finite subset of transactions, and let the values of a certain function $y^*: X \to Y$ be known for this subset. Then $y_i = y^*(x_i)$. The function y^* will be referred to as the target function, and the collection of pairs $X^l = (x_i,\ y_i)_{i=1}^l$ will be called the training set.

The task of learning from precedents is to reconstruct the dependence y^* based on the sample X^l, i.e. to build a decision function $a: X \to Y$, that approximates the target function $y^*(x)$, not only on the transactions in X^l, but also on the entire set X.

The decision function a will also be referred to as an algorithm, and in some cases, as a classifier, when its role in evaluating transactions will be to classify them into fraudulent or legitimate categories. For practical application, the constructed algorithm a should provide efficient computer implementation, as it is expected that financial organizations will use it to analyze their transactional data stored on their servers.

The attributes of transactions x, obtained from the processes of financial organizations (such as transaction amount, customer age, insurance payout amount, etc.), from the perspective of learning from precedents, are features and formally represent the mapping $f: X \to D_f$, where D_f is the set of valid feature values.

There are several types of features, depending on the nature of the data.

- $D_f = \{0, 1\}$ – binary feature;
- $D_f = \mathbb{R}$ – quantitative feature;
- D_f – finite set, nominal or categorical feature.

In case all features in the data are the same, $D_{f_1} = \cdots = D_{f_n}$, and such data is called homogeneous, otherwise it is heterogeneous. In practice, transaction data stored in financial organizations is heterogeneous and contains all types of features. In this study, all categorical features will be transformed into binary using commonly known machine learning algorithms.

Let there be a set of features $f_1, \ldots, f_n$. The vector $(f_1(x), \ldots, f_n(x))$ is called the feature space of transaction $x \in X$. The collection of feature descriptions of all objects in the sample X^l, represented as a table of size $l \times n$, is called the matrix of objects-features:

$$F = \| f_j(x_i) \|_{l \times n} = \begin{pmatrix} f_1(x_1) & \ldots & f_n(x_1) \\ \ldots & \ldots & \ldots \\ f_1(x_l) & \ldots & f_n(x_l) \end{pmatrix} \quad (1)$$

An example of transaction descriptions for fraud detection task is presented in Table 1.

Table 1

Features of operations

Date and time of transaction	Card operation type	Type of service	Shop MCC	Transaction amount	Fraud
01.02.2024 13:03	Purchase via pos	Car service	5533	26720,00	0
01.02.2024 13:10	Purchase via pos	Car service	5533	1500,00	0
02.02.2024 14:12	Purchase via pos	Gas station	5541	2202,78	0
08.02.2024 10:00	Purchase via pos	Pet Shop	5995	7399,00	0
10.02.2024 23:00	Purchase via ecom	P2P	4900	4500,00	1

In this study, the set of allowable answers $Y = \{0, 1\}$, represents a classification task with two non-overlapping classes. In general, if $Y = \{1, \dots, M\}$, the set of transactions X can be divided into M non-overlapping classes $K_y = \{x \in X: y^*(x) = y\}$. The algorithm $a(x)$ provides an answer to the question "to which class does x belong?", and in the fraud detection task, the answer will indicate whether a transaction is fraudulent or not.

According to [12], a model of algorithms is defined as a parametric family of mappings $A = \{g(x, \theta) \mid \theta \in \Theta\}$, where $g: X \times \Theta \to Y$ is a fixed function, and θ, is the set of allowable parameter values, known as the parameter space.

The dissertation research aims to search for optimal model parameters for classifying transactions and embedding the obtained models at various stages of fraud detection in a financial organization. Currently, there are numerous different approaches and techniques for finding algorithms and optimal parameters (hyperparameters) that ultimately allow obtaining the necessary algorithm, $a(x)$, for decision-making in various tasks. The collection of these approaches is also known as machine learning methods (hereafter referred to as ML). In this study, the following well-known ML methods have been selected for integration into the fraud detection process.

The Decision Tree[1] (DT) is the most interpretable and simple tool used in machine learning. The modeling result can be represented as a tree-like structure, from which it is easy to extract a simple decision rule.

The Random Forest[2] (RF) algorithm has been chosen as the base algorithm for classification, as it has shown the best results in studies related to fraud detection [28]. The method involves using an ensemble of Decision Tree algorithms, each of which may not provide high classification quality individually, but by combining a large number of them, better results can be achieved. The choice of RF in this study is motivated by its low sensitivity to the size of the feature space and its high classification quality when trained on heterogeneous data with categorical and quantitative features.

To build an algorithm on data with a small number of features, a multi-layer perceptron[3] (MLP) has been selected. This method has also shown high results in research in the field of fraud prevention. When using MLP, the inclusion of hidden layers allows for the approximation of a nonlinear function for classification.

The search for the best model from the parameter space θ o is performed using the GridSearchCV[4] tool, which optimizes hyperparameters through cross-validation and grid search

To evaluate the results of the experiment, traditional metrics commonly used in fraud detection tasks were selected. In this study, classification into two non-

[1] https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html
[2] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
[3] https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html
[4] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

overlapping classes $Y = \{0, 1\}$ is considered. Let $y_i \in \mathbb{R}$, be the output of the trained model for the i-th transaction. To make a decision on whether a transaction is fraudulent or legitimate, we will use a threshold th, which converts the values of y_i into non-overlapping classes $y_i^p = [y_i > th]$[5].

From a statistical point of view, classification involves making a decision about the null hypothesis H_0 that a transaction belongs to class 1 and the alternative hypothesis H_1 that a transaction belongs to class 0. The decisions made can involve two types of errors: false positive (or Type I error), where a legitimate transaction is classified as fraudulent, and false negative (or Type II error), where a fraudulent transaction is classified as legitimate. Changing the threshold allows for adjusting the trade-off between these two types of errors, as increasing the probability of Type I error usually decreases the probability of Type II error, and vice versa.

The threshold th is chosen depending on the task at hand, and when it is fixed, it is possible to construct Table 2 (confusion matrix or error matrix):

Table 2

Confusion matrix

		Correct hypothesis	
		H_0	H_1
Result of decision making	H_0	TP, H_0 is correctly accepted	FP, H_0 is incorrectly accepted (Type II error)
	H_1	FN, H_0 is incorrectly rejected (Type I error)	TN, H_0 is correctly rejected

In traditional terms of machine learning, the implementation of hypotheses can be formulated in the following way:

- TP (True positive) – correctly identified fraudulent transaction,
- FP (False positive) – legitimate transaction identified as fraudulent,
- TN (True negative) – correctly identified legitimate transaction,
- FN (False negative) – fraudulent transaction identified as legitimate.

To evaluate the quality of classification, we will also use the following characteristics:

[5] Square brackets convert a logical value into a number according to the following rule: [false] = 0, [true] = 1

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Specificity = \frac{TN}{TN+FP} \tag{4}$$

Recall allows to assess the proportion of fraud detected by the classifier out of all fraudulent transactions. Precision is the probability that a transaction flagged by the classifier is truly fraudulent. Specificity is the proportion of legitimate transactions correctly identified by the classifier.

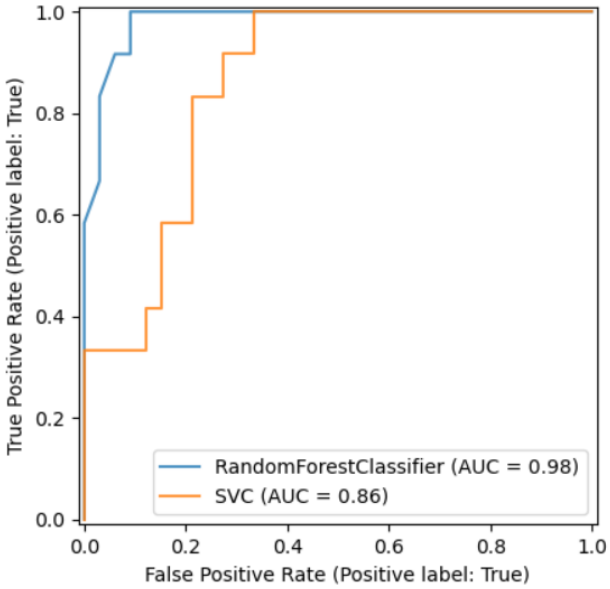
Also, a special characteristic called the ROC curve [29], will be used in the study, which shows what happens to the number of errors of both types as th changes. The false positive rate (FPR), computed for each threshold value, is plotted on the X-axis:

$$FPR = \frac{\sum_i [y_i^p = 1]}{\sum_i [y_i = 1]} \tag{5}$$

The proportion of true positive classifications (TPR) is plotted on the Y-axis, also computed for each threshold value:

$$TPR = \frac{\sum_i [y_i^p = 1]}{\sum_i [y_i = 0]} \tag{6}$$

An example of constructing an ROC curve[6] for two different machine learning methods is presented in Figure 1.



[6] https://scikit-learn.org/stable/auto_examples/miscellaneous/plot_roc_curve_visualization_api.html

Fig. 1. An example of comparing ROC curves for two different methods

The higher the ROC curve, the higher the classification quality. The ideal ROC curve passes through the upper left corner - the point (0, 1). The worst algorithm corresponds to the diagonal line connecting points (0, 0) and (1, 1). The area under the ROC curve (AUC) serves as a general characteristic of classification quality.

When working with highly imbalanced data, such as in the case of fraud detection, AUC (and ROC curves) can be overly optimistic. Therefore, it is suggested to use another evaluation metric for classifiers - the precision-recall (PR) curve. An example[7] of its construction is presented in Figure 2.

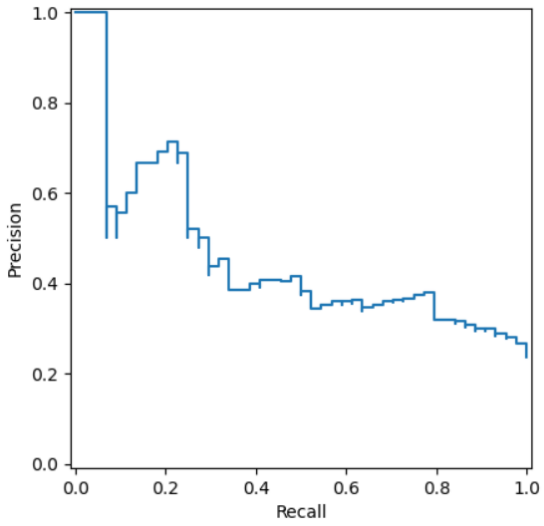


Fig. 2. An example of constructing a PR curve

The area under this curve (AUPRC) will also be used as a quantitative characteristic for model evaluation. As the name suggests, the precision-recall curve depicts precision (Y-axis) as a function of recall (X-axis) for each possible threshold. AUPRC also represents a value between 0 and 1.

The next section will provide a brief overview of fraud monitoring systems in financial organizations and their key components, which are expected to utilize machine learning methods.

4.2. Architectures of fraud monitoring systems and potential areas for their improvement

In the dissertation research, two processes are considered where financial organizations apply data-driven tools for fraud detection. These processes include customer banking transactions and insurance claim reviews.

Figure 3 schematically depicts the path of a banking payment through the fraud monitoring system.

[7] https://scikit-learn.org/stable/auto_examples/miscellaneous/plot_display_object_visualization.html

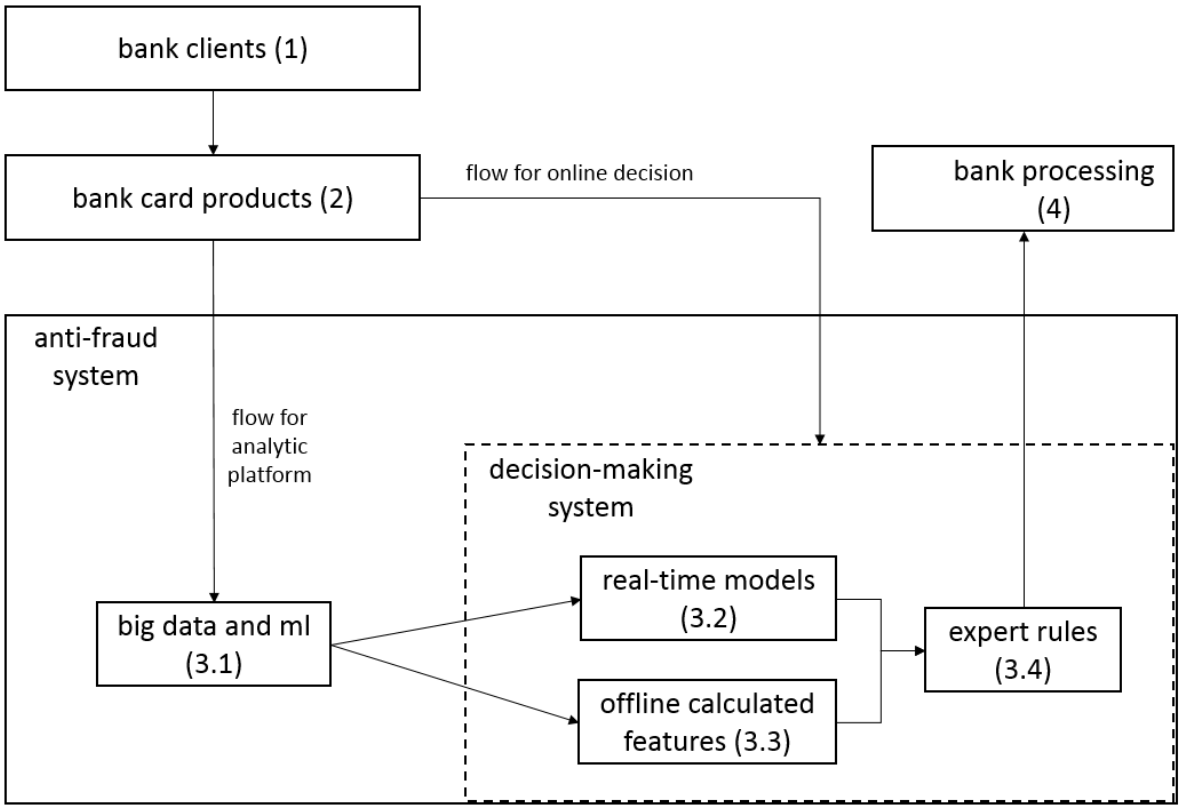


Fig. 3. An example of the stages of operation of a bank fraud monitoring system

(1) Bank customers who use banking services make purchases on websites, in retail stores, link cards for payment using smartphones, and so on

(2) Banking services for payment of goods and services, including online payments, money transfers, and cash withdrawals

(3) The bank's anti-fraud system (fraud monitoring)

(3.1) An analytical platform where the development of fraud detection algorithms takes place. In large banks with a high transactional flow, a stack of BIG DATA technologies is typically used for these purposes

(3.2) A technological block for executing machine learning models in real-time (model-based approach).

(3.3) Enrichment of transactions with additional features created on the analytical platform.

(3.4) A technological block for making the final decision on an operation based on expert rules (rule-based approach).

(4) Bank processing, in which the operation is executed directly after the fraud monitoring verdict.

The second case examines the process of an insurance company, where the risk of fraud by policyholders is reduced. The first barrier for fraudsters is the verification of the customer before entering into an insurance contract. In addition to actuarial calculations, insurers can refer to internal black or white lists, external sources of customer data, and apply their own models to assess the risk of fraud by the policyholder. Such procedures directly affect the insurer - they lengthen the

policy sales process, deteriorating the customer experience, while false rejections decrease the level of insurance premium collection. These facts compel insurance companies to simplify and automate checks at this stage. In this case, companies focus on the accuracy of fraud detection but do not pay sufficient attention to completeness, which allows professional fraudsters to successfully penetrate the insurance portfolio.

Next, the insurer analyzes the reported claims and, if signs of insurance fraud are detected, denies the payout. At this stage, both expert evaluation methods from the insurer's security department and techniques using machine learning methods are applied. The combination of expert work and systems that assess claims through data analysis and social networks yields positive results [30]. In the current process, the insurance company focuses on the completeness of fraud detection to stop a professional fraudster who has already manifested themselves and reduce their impact on the portfolio's loss ratio.

The schematically considered processes can be represented as follows (Figure 4).

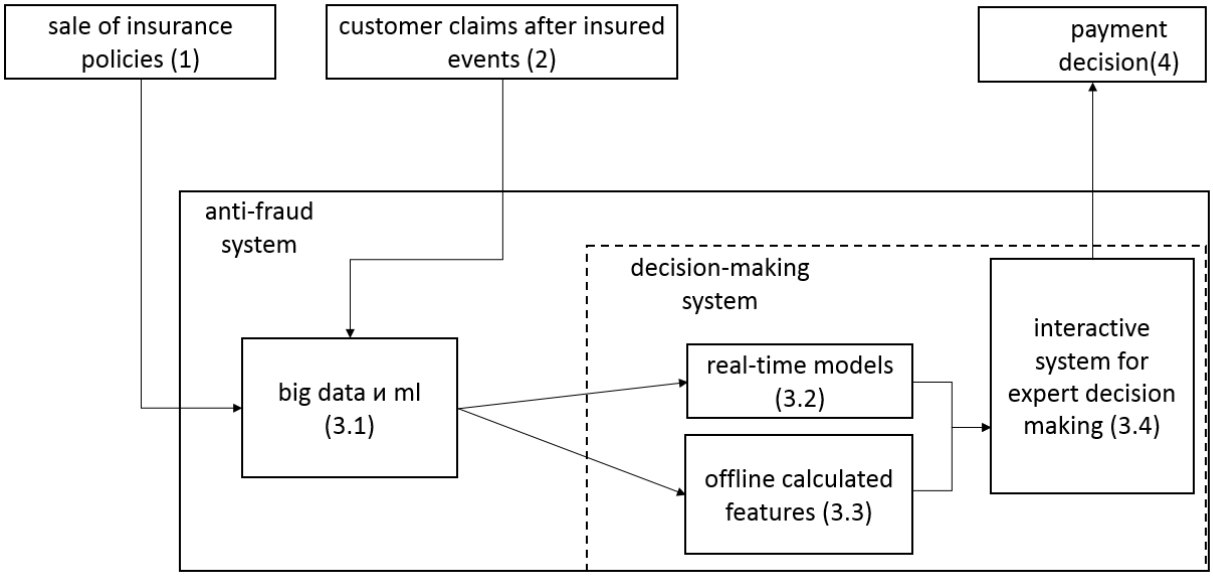


Fig. 4. An example of the stages of operation of an insurance company’s fraud monitoring system

In the present study, the following components of fraud monitoring systems have been selected, in which machine learning methods will be integrated to enhance the effectiveness of fraud detection:

1. data preparation for building transaction decision algorithms;
2. algorithm configuration in the decision-making system.

During the data preparation stage, class labeling correction is applied using a multilayer perceptron, as well as feature space expansion by using data that is different from the data contained in the evaluated transaction and data extracted from the graph.

The expert algorithm configuration stage will be automated using machine learning methods, which will enhance the accuracy of fraud detection.

4.3. Preparation of data for training classifiers of fraudulent transactions and claims

To improve the quality of fraud transaction classification, the following data preparation process is proposed:

1. the dataset is divided into four parts from different time periods;
2. the earlier part (D_{init}) is used to train the model M_L, which will be used to adjust the expert evaluation;
3. the next part (D_{train}) is used to train the model M_S on the re-labeled dataset, as well as to train the baseline model M_B, which will be compared to the experimental results;
4. the next part ($D_{control\_1}$), is used to find the cutoff points (TH_{fraud}, $TH_{legitimate}$) of the model M_L, based on which decisions will be made to adjust the labeling in D_{train};
5. finally, the $D_{control\_2}$ dataset is used for result validation - measuring the quality characteristics of the classification.

The data partitioning and classifiers for transaction evaluation are schematically presented in Figure 5.

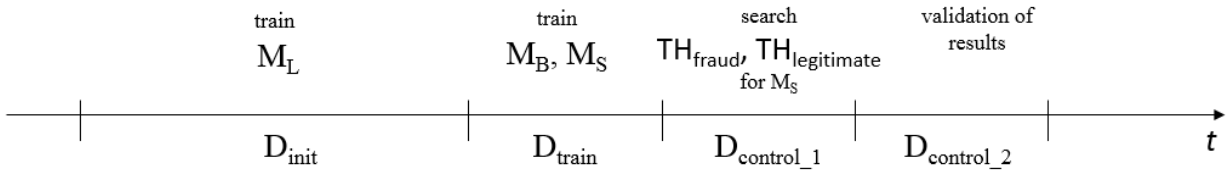


Fig. 5. Splitting data when adjusting markup

Table 3 presents the parameters used and references to the descriptions of classifiers applied in the proposed approach.

Table 3

Classifiers used in the training process

Classifier	Name	Link to description
M_L	Multilayer perceptron	https://scikit-learn.org/stable/modules/neural_networks_supervised.html (date of access: 12.02.2024)

M_B	RandomForest Classifier	https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (date of access: 12.02.2024)
M_S	RandomForest Classifier	https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (date of access: 12.02.2024)

The choice of RandomForest and multilayer perceptron as classifiers is justified by the research [28], which compares the main machine learning methods in fraud detection tasks in auto insurance.

The stages of label correction are schematically presented in Figure 6.

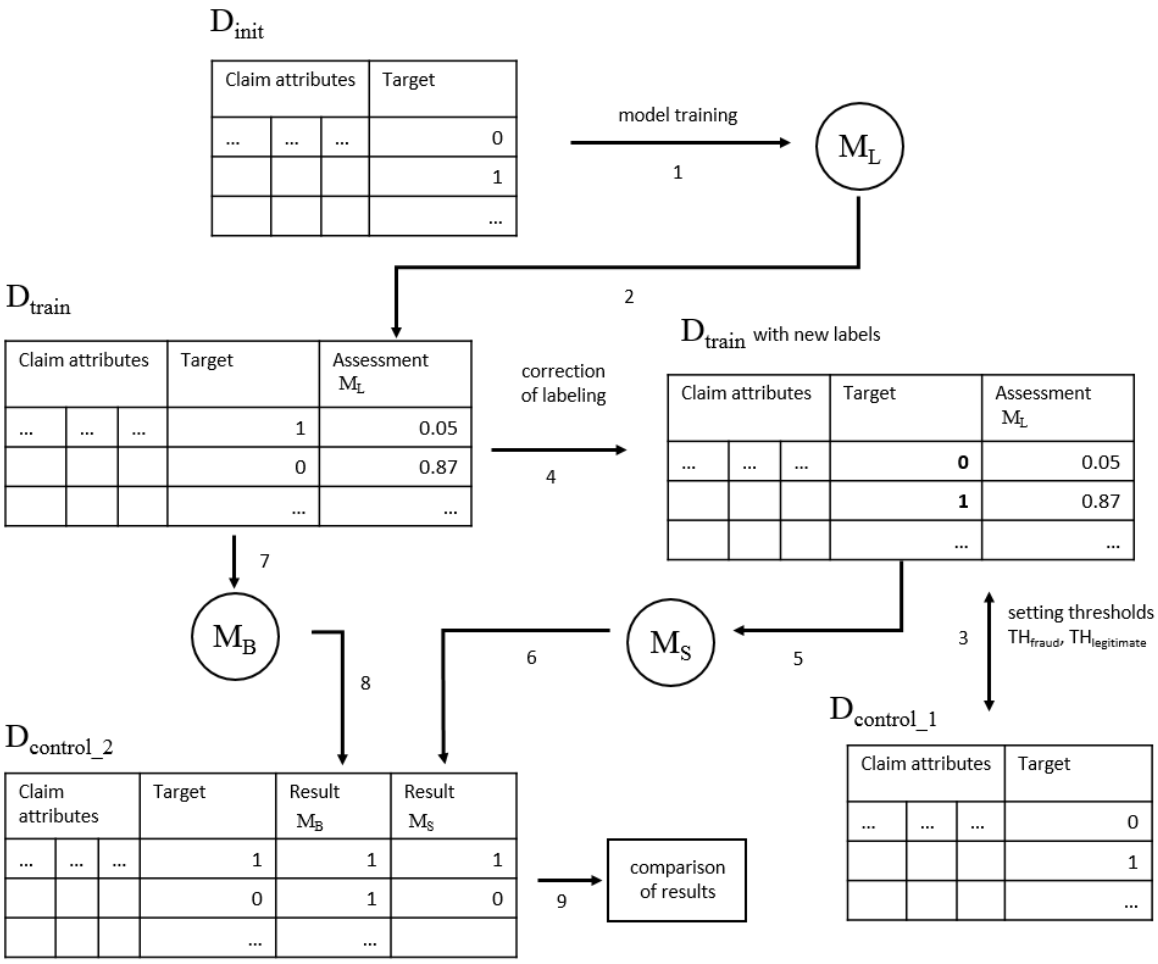


Fig. 6. Proposed sequence of stages for marking adjustments

To expand the feature space, it is proposed to include information from sources unrelated to the assessment of the current transaction. For example, in the banking industry, when evaluating transfers between clients, the history of customer purchase transactions can be integrated. In this case, the improvement in characteristics is achieved because fraudsters cannot provide a legitimate account history for the accounts they use in their fraudulent scheme.

To identify anomalous customer behavior in an insurance company, it is proposed to construct an undirected graph, where the vertices represent insurance claim requests and the edges represent accident-related entities (drivers, policyholders, etc.) as well as the incident itself. The graph is constructed over a certain period, such as a calendar year.

As new features for the claim, the properties of the vertices in the constructed graph can be considered. Table 4 presents the features examined within the scope of the dissertation research.

Table 4

Claim attributes built on the basis of the claims graph

Graph object	Attribute description
Vertex	• Vertex degree • Minimum degree of adjacent vertices • Number of adjacent vertices • Average degree of adjacent vertices
Connected Component	• Number of vertices in the connected component
Clique	• The size of the maximum clique that the bus vertex consists of • The number of cliques that the vertex consists of
Cycle	• The cycle length in which the vertex placed • The average degree of the vertices of the cycle in which the vertex is placed

4.4. Generation of new rules to improve the quality of the fraud monitoring system

Expert rule tuning in the decision-making system is proposed to be automated using machine learning methods. The approach consists of three stages:

- Data preparation and preprocessing;
- Application of machine learning methods;
- Extraction and evaluation of rules.

The data preparation stage includes:

1. loading historical data from the fraud monitoring system;
2. emulating the operation of the fraud monitoring system;
3. feature selection, filtering noisy data, and developing additional features.

In the next stage, the prepared data is subjected to Decision Tree or Random Forest methods.

At the final stage, rules are selected for integration into the fraud monitoring system. To do this, rules are extracted from trained algorithms built using Decision Tree or Random Forest methods. The rules are then compared based on classification quality characteristics. The best rules are implemented in the financial organization's fraud monitoring system to ensure they operate according to expert rules.

4.5. Methodology for conducting experiments and research

During the research, methodologies were developed to conduct four different experiments that allow for assessing the applicability of the proposed approaches.

4.5.1. Correction of target class labeling

For the experiment, two datasets on car insurance were selected. One of them is a well-known and widely used dataset called "carclaims.txt". It contains insurance claims registered in the USA from 1994 to 1996 [31].

Additionally, to demonstrate the applicability of the approach on different insurance data, the file "insurance_claims.csv"[8], was examined, which includes claims from January to February 2015. The features selected for the purposes of the dissertation research are listed in Table 5.

Table 5

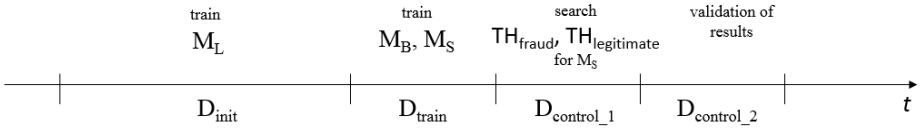
Description of claims characteristics for assessing the fraudulent component

Dataset	Feature name	Description
«carclaims.txt»	Age	Age of the policyholder
	DriverRating	Rating of the driver involved in the accident
	Gender	Gender of the policyholder
	BasePolicy	Policy type
	Fault	Guilty side
	NumberOfSuppliments	Number of additional options in the policy
	PastNumberOfClaims	Number of insured events under the current policy
	VehiclePrice	The cost of the car involved in the accident
	AgeOfPolicyHolder	Age of the policyholder
«insurance_claims.csv»	age	Age of the policyholder
	months_as_customer	Number of months as policyholder
	policy_annual_premium	Insurance premium
	insured_sex	Gender of the policyholder
	total_claim_amount	Claim amount
	incident_severity	Seriousness of the insured event

[8] https://www.kaggle.com/datasets/buntyshah/auto-insurance-claims-data

This small set of features was chosen to maintain the applicability of the proposed approach for evaluating the fraudulent component in various portfolios of insurance claims. This set of features can be obtained from the questionnaire data of policyholders and claims during the settlement of an insurance case.

Dataset "carclaims.txt" consists of 15,420 records, out of which 14,497 are legitimate claims and 923 (6%) have indicators of insurance fraud. The size of "insurance_claims.csv" is 1,000 records, with 247 (24.7%) being fraudulent. The claims can be arranged chronologically based on their submission to the insurance company, and the model's quality is proposed to be evaluated on more recent data. The data partitioning is schematically presented in Figure 7.



		D_{init}	D_{train}	$D_{control\_1}$	$D_{control\_2}$
«carclaims.txt»	number of claims:	6420	3000	3000	3000
	class imbalance:	6,6%	5,7%	5,8%	5,2%
«insurance_claims.csv»	number of claims:	255	241	253	251
	class imbalance:	26.6%	25.3%	28.4%	18.3%

Fig. 7. Splitting the data for an experiment

Table 6 presents the parameters used in the process of training the classifiers.

Table 6

Classifier hyperparameters

Classifier	Name	Hyperparameters
M_L	Multilayer perceptron	«carclaims.txt»: hidden_layer_sizes=(10), solver='lbfgs' «insurance_claims.csv»: hidden_layer_sizes=(2), solver='lbfgs' , activation = 'relu'
M_B	RandomForestClassifier	«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

M_S	RandomForestClassifier	«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

Experimental values of TH_{fraud}, $TH_{legitimate}$ were selected by measuring the classification quality on $D_{control\_1}$:

a) for the "carclaims.txt" dataset: $TH_{fraud} = 0,75$; $TH_{legitimate} = 0,1$;

b) for the «insurance_claims.csv» dataset: $TH_{fraud} = 0,8$; $TH_{legitimate} = 0,05$.

After that, the claims in D_{train} were re-labeled as follows:

- If the assessment result (probability of being classified as fraud) of a claim using M_L is greater than TH_{fraud}, it is relabeled as fraudulent;
- if the assessment result is less than $TH_{legitimate}$, the claim is relabeled as legitimate;
- the labeling of the claim remains unchanged in all other cases.

This class correction improved the balance in D_{train} to 36.3% in the "carclaims.txt" dataset and to 35.8% in the "insurance_claims.csv" dataset. Subsequently, M_B was trained on the D_{train} data before class correction, and M_S was trained on the D_{train} data after class correction.

The obtained models M_B and M_S were applied to the $D_{control\_2}$ dataset, which was not involved in the training of the classifiers or parameter tuning and is more recent in terms of the occurrence of claims by the insurer. Additionally, the labels in this dataset were not subjected to any correction. The ROC curves for these models are presented in Figure 8.

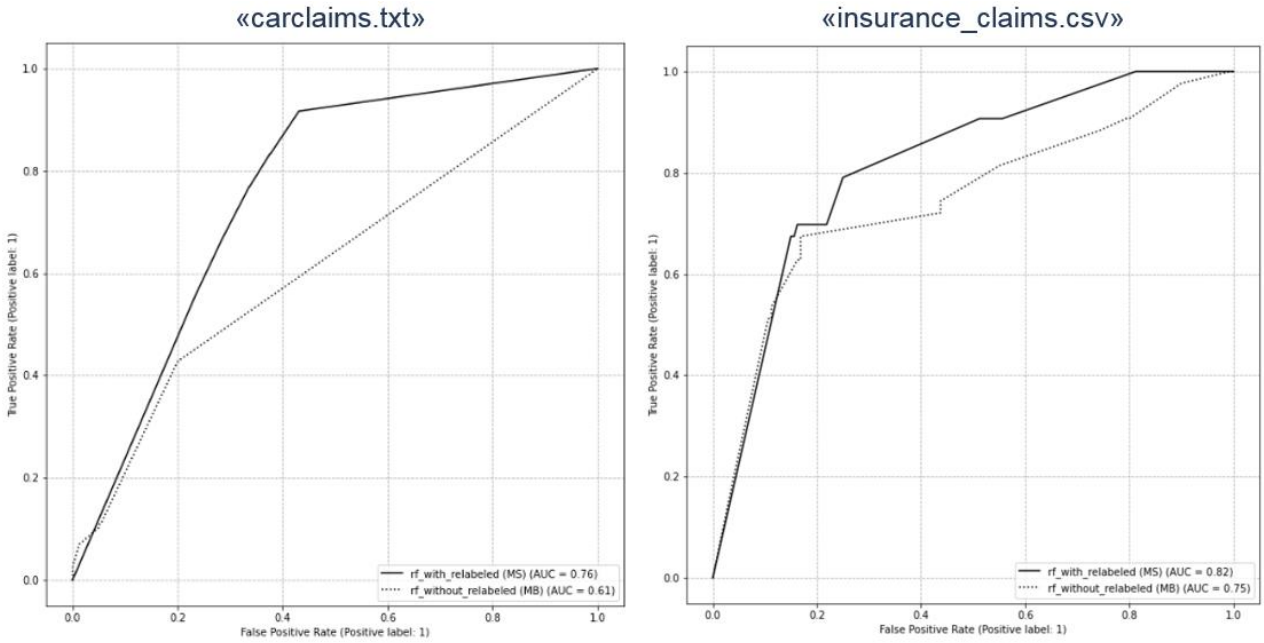


Fig. 8. Comparison of ROC curves for models M_B and M_S

The values of the areas under the curves (AUC) demonstrate significantly better classification quality for the proposed approach compared to the traditional training without label correction. Table 7 also presents the values of Recall at a fixed precision for comparing the models.

Table 7

Comparison of fraud detection quality metrics

Dataset	Characteristics	Proposed approach, M_S	Traditional approach, M_B
«carclaims.txt»	ROC AUC	**0,76**	0,61
	Precision	0,1	0,1
	Recall	**0,92**	0,43
«insurance_claims.csv»	ROC AUC	**0,82**	0,75
	Precision	0,55	0,55
	Recall	**0,7**	0,51

4.5.2. Using data of a different nature in the feature space

As a dataset for study, a set of transfer operations from a large bank over a weekly period was selected. These operations were flagged as suspicious by the fraud monitoring system, triggering one of the processing scenarios, such as issuing a warning to the client about possible fraud or outright rejection of the operation. Additionally, cases were added to the dataset where the fraud

monitoring system did not raise any alerts, but the client reported during the period under consideration that the operation was fraudulent. Missed and detected fraudulent operations are combined and will be treated as the target class when building the classifier. False positives are considered as the second class. Fraudulent operations will be labeled as 1, and false positives as 0. This way, the table (hits_fm) will be formed as follows:

Table 7

An example of a dataset (hits_fm) formed of a fraud monitoring system triggers.

Client initiating transfer	Client recipient of the transfer	Class label	Date of transaction
cl_1	cl_2	0	20.02.2021
cl_3	cl_4	1	20.02.2021
…	…	…	…
cl_m	cl_k	0	27.02.2021

Further, we will assume that if the class 1 is assigned in the hits_fm table, then the recipient's profile can be classified as fraudulent (drop). In the case of 0, the recipient is legitimate. The bank has the ability to build a table for each recipient based on the history of card transactions (Table 8). In order to limit the experiment, the history of transactions that occurred in a two-week period prior to the first transfer to the client from the hits_fm table is used. For example, in the "Transaction Amount in MCC_1 Group" column for the client cl_2, the value represents the sum of all payments made by the client in the two weeks preceding the operation in Table 7. In this case, during the period from 05.02.2021 to 19.02.2021, expenses in the MCC_1 category amounted to 40,000 rubles.

Table 8

Clients profile data set collected from customer card transactions history

Bank client	**Transaction amount in group MCC_1**	**Transaction amount in group MCC_2**	**…**	**Transaction amount in group MCC_N**
cl_2	40000	0	…	11112
cl_4	0	30000	…	0
…	…	…	…	…

The clients_profile table allows for an expansion of the feature space to apply machine learning methods in detecting fraud in bank transfers.

As a result of using the new data, the value of the Precision metric was significantly increased from 0.07 to 0.69.

4.5.3. Using data extracted from a graph in the feature space

In this experiment, two datasets, "carclaims.txt" and "insurance_claims.csv," are also considered.

The datasets under consideration do not have explicit attributes that would allow for the construction of a claims graph, such as a contract number. Therefore, the following assumptions were made to establish connections between two claims.

a. The claims are connected to each other by the participant of the insurance event if they have matching attributes: Make, Sex, MaritalStatus, Age, VehicleCategory, VehiclePrice, AgeOfVehicle, DriverRating, AgentType, NumberOfCars, BasePolicy for «carclaims.txt»; POSTAL_CODE for «insurance_claims.csv».

b. The claims are connected to each other by the insurance event if they have matching attributes: Year, Month, WeekOfMonth, DayOfWeek, AccidentArea, PoliceReportFiled, WitnessPresen for «carclaims.txt»; INCIDENT_CITY, INCIDENT_HOUR_OF_THE_DAY, INCIDENT_SEVERITY for «insurance_claims.csv».

The graph constructed in this way can be visualized as shown in Figure 9.

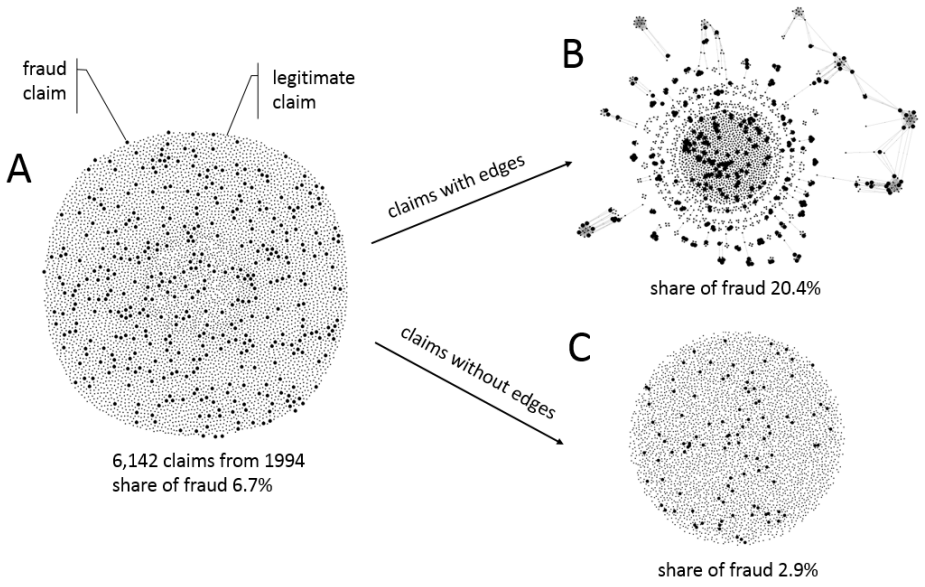


Рис. 9. An example of visualizing a graph built on connections between insurance events and policyholders

It can be noted that the very fact of the connection between claims increases the probability of fraud in the claim.

For each claim, a set of features is extracted from Table 4, and then the RandomForest machine learning method is applied to both the original dataset and the extended dataset.

Figure 10 demonstrates a comparison of performance characteristics when using new data extracted from the graph and when not using them. The experiment showed a significant increase in fraud detection efficiency for two independent datasets.

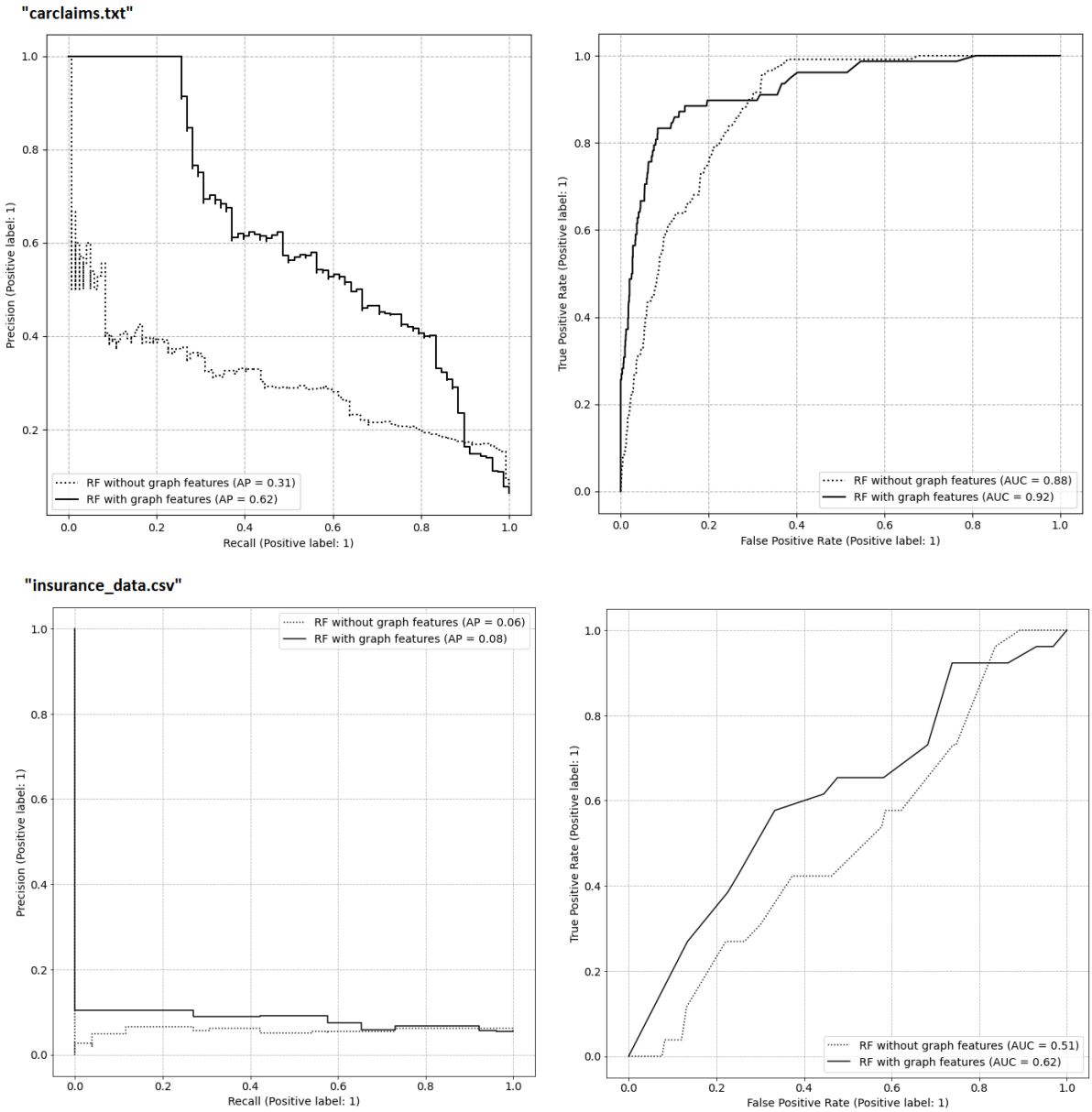


Fig. 10. Comparison of ROC and Precision-Recall curves for models trained with and without graph claims features.

4.5.4. Reconfiguring rules in the decision-making system

The decision rules developed using the approach proposed in Section 4.4 were successfully implemented in a real banking fraud monitoring system. The results showed that the average precision of the rules was 50%, and the average recall for fraud detection reached 0.6%.

5. Conclusion

It is important to note that the developed methods have the potential to improve the effectiveness of applying machine learning methods in combating fraud. These results can be valuable for various financial organizations facing classification challenges when using machine learning methods in their fraud monitoring systems.

Key results of the research study are as follows.

1. A method for improving the quality of data labeling for machine learning in fraud detection has been developed.
2. Methods for expanding the feature space have been developed to enhance the efficiency of fraud detection.
3. A method for tuning the decision-making system in fraud monitoring, incorporating elements of machine learning, has been proposed.
4. Experimental studies have been conducted to validate the effectiveness of the proposed approaches.

References

1. Al-Hashedi K.G., Magalingam P. Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019 // Computer Science Review. 2021. Vol. 40. P. 100402.
2. Bao Y., Hilary G., Ke B. Artificial Intelligence and Fraud Detection. 2022. P. 223–247.
3. Bezzateev S.V. et al. Risk assessment methodology for information systems, based on the user behavior and IT-security incidents analysis // Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2021. Vol. 21, № 4. P. 553–561.
4. Abdallah A., Maarof M.A., Zainal A. Fraud detection system: A survey // Journal of Network and Computer Applications. 2016. Vol. 68. P. 90–113.
5. Gupta P. et al. Unbalanced Credit Card Fraud Detection Data: A Machine Learning-Oriented Comparative Study of Balancing Techniques // Procedia Computer Science. 2023. Vol. 218. P. 2575–2584.
6. 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). IEEE, 2021. P. 1606–1611.
7. Jin C., Feng Y., Li F. Concept drift detection based on decision distribution in inconsistent information system // Knowledge-Based Systems. 2023. Vol. 279. P. 110934.

8. Anderson O.D. A Note on “Trial by Computer”—A Case Study of the Use of Simple Statistical Techniques in the Detection of a Fraud // Journal of the Operational Research Society. 1986. Vol. 37, № 4. P. 423–427.
9. Mercer L.C.J. Fraud detection via regression analysis // Computers & Security. 1990. Vol. 9, № 4. P. 331–338.
10. Wolf D., Greenberg D. The Dynamics of Welfare Fraud: An Econometric Duration Model in Discrete Time // Journal of Human Resources. 1986. Vol. 21, № 4. P. 437–455.
11. Baesens B., Vlasselaer V.V., Verbeke W. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques. Hoboken, NJ, USA: John Wiley & Sons, Inc, 2015.
12. Vorontsov K. V. Matematicheskie metody obucheniya po precedentam (teoriya obucheniya mashin) // Sajt «Mashinnoe obuchenie», kurs lekcij. 2011.
13. Kumari P., Mishra S.P. Analysis of Credit Card Fraud Detection Using Fusion Classifiers. 2019. P. 111–122.
14. Baesens B., Höppner S., Verdonck T. Data engineering for fraud detection // Decision Support Systems. 2021. Vol. 150. P. 113492.
15. Ghosh, Reilly. Credit card fraud detection with a neural-network // 1994 Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences. 1994. Vol. 3. P. 621–630.

16. Baader G., Krcmar H. Reducing false positives in fraud detection: Combining the red flag approach with process mining // International Journal of Accounting Information Systems. 2018. Vol. 31. P. 1–16.
17. Nian K. et al. Auto insurance fraud detection using unsupervised spectral ranking for anomaly // The Journal of Finance and Data Science. 2016. Vol. 2, № 1. P. 58–75.
18. 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, № 5. P. 568–575.
19. Yankol-Schalck M. The value of cross-data set analysis for automobile insurance fraud detection // Research in International Business and Finance. 2022. Vol. 63. P. 101769.
20. Vandervorst F., Verbeke W., Verdonck T. Data misrepresentation detection for insurance underwriting fraud prevention // Decision Support Systems. 2022. Vol. 159. P. 113798.
21. Aslam F. et al. Insurance fraud detection: Evidence from artificial intelligence and machine learning // Research in International Business and Finance. 2022. Vol. 62. P. 101744.
22. Yan C. et al. Improved adaptive genetic algorithm for the vehicle Insurance Fraud Identification Model based on a BP Neural Network // Theoretical Computer Science. 2020. Vol. 817. P. 12–23.

23. Salmi M., Atif D. Using a Data Mining Approach to Detect Automobile Insurance Fraud. 2022. P. 55–66.
24. Soufiane E. et al. Automobile Insurance Claims Auditing: A Comprehensive Survey on Handling Awry Datasets. 2022. P. 135–144.
25. Vorobyev I. ML methods for assessing the risk of fraud in auto insurance // Izvestiya of Saratov University. Mathematics. Mechanics. Informatics.
26. Festa Y.Y., Vorobyev I.A. A hybrid machine learning framework for e-commerce fraud detection // Model Assisted Statistics and Applications. 2022. Vol. 17, № 1. P. 41–49.
27. Vorobyev I., Krivitskaya A. Reducing false positives in bank anti-fraud systems based on rule induction in distributed tree-based models // Computers & Security. 2022. Vol. 120. P. 102786.
28. Itri B. et al. Performance comparative study of machine learning algorithms for automobile insurance fraud detection // 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS). IEEE, 2019. P. 1–4.
29. Fawcett T. An introduction to ROC analysis // Pattern Recognition Letters. 2006. Vol. 27, № 8. P. 861–874.
30. Šubelj L., Furlan Š., Bajec M. An expert system for detecting automobile insurance fraud using social network analysis // Expert Systems with Applications. 2011. Vol. 38, № 1. P. 1039–1052.
31. Phua C., Alahakoon D., Lee V. Minority report in fraud detection // ACM SIGKDD Explorations Newsletter. 2004. Vol. 6, № 1. P. 50–59.