ISSN Online: 2821-1936

Transactions on Data Analysis in Social Science

Journal Homepage: https://transoscience.ir

An Intelligent Method for Credit Card Fraud Detection Using Data Mining Techniques

F. Fallah Ziarani[1,*], A. Jalalian[2]

[1] Department of Information Technology Engineering, Faculty of Computer Science, Raja University, Qazvin, Iran
[2] Department of Information Technology Engineering, Faculty of Computer Science, Raja University, Qazvin, Iran

ARTICLE INFO

Article History:
Received 21 July 2020
Received in revised form 4 October 2020
Accepted 20 November 2020
Available online 2 December 2020

Keywords:
Electronic Banking, Data Mining, Credit Card Fraud, Imbalanced Data

ABSTRACT

In recent years, billions of dollars in losses have been caused by fraudulent credit card transactions, representing a serious and growing problem. To mitigate the damage from such transactions, data mining techniques are widely employed for credit card fraud detection, utilizing approaches such as classification and clustering with machine learning algorithms. This study focuses on supervised learning, in which machine learning algorithms are trained on labeled datasets to construct predictive models. A major challenge in this domain is the highly imbalanced class distribution within the datasets, as the number of fraudulent transactions is significantly lower than that of legitimate transactions. This paper examines strategies for addressing imbalanced data in machine learning algorithms and proposes an optimized method for detecting and identifying fraud on both original and balanced datasets. In this study, the performance of classification techniques is compared using well-known methods, including C5.0 decision trees, Support Vector Machines (SVM) with Sigmoid, Linear, and RBF kernels, and neural networks.

1. INTRODUCTION

In recent years, e-commerce has emerged as one of the main channels for global trade. Electronic payment systems constitute a critical component of e-commerce. Among these, credit cards have become one of the primary payment methods. While e-commerce provides significant opportunities across various sectors, particularly global trade, it also introduces risks for banks, financial institutions, and credit organizations [1, 2]. The nature of fraud has evolved over the past decades with technological advancements. Fraud has existed as long as human society itself and can take various forms [3]. Credit card fraud is currently one of the most significant threats to commercial institutions. Typically, fraud involves the unauthorized and illegal use of the facilities of a legitimate account [4] and generally refers to obtaining goods, services, or money through illegal means [5]. Over time, as fraud detection methods have evolved, fraudsters have continuously adapted their techniques to evade detection [6].

* Corresponding Author: f.fallah0035@gmail.com
Department of Information Technology Engineering, Faculty of Computer Science, Raja University, Qazvin, Iran

http://dx.doi.org/10.47176/TDASS.2020.180

© 2020 by the authors. Licensee T.D.A.S, Tehran, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).

Credit card fraud can be broadly categorized into three types [7]:

Physical card fraud: In this type, the cardholder physically presents the card to a merchant for payment. The card may be lost or stolen, and then used by another person, potentially resulting in substantial financial loss if the cardholder becomes aware of it too late [8].

Virtual card fraud: In this form, the fraudster does not have the physical card but requires the card's information. Fraud occurs through the unauthorized use of the card's details, sometimes facilitated by counterfeit cards [5, 8].

Application fraud: This occurs when a credit card is fraudulently obtained using false personal information. This type of fraud is less common because it can often be detected during the application process by verifying the applicant's information, unlike other forms of fraud that are unpredictable [5].

Two main approaches are employed in combating credit card fraud: prevention and detection. Fraud prevention differs from detection, as it involves actions taken at the outset to stop or neutralize fraud, while detection comes into play when prevention fails, utilizing appropriate tools to identify fraudulent activities [1]. For human analysts, recognizing patterns of fraud in transaction datasets is often impractical due to the large number of online, high-dimensional data points. Therefore, automated systems for fraud detection are essential [9].

The design of efficient fraud detection algorithms is key to reducing losses. Many algorithms rely on advanced machine learning techniques to assist researchers in uncovering fraud [5]. However, designing effective fraud detection algorithms is challenging due to factors such as non-stationary data distributions, highly imbalanced class distributions, and the limited availability of labeled transactions verified by fraud inspectors. Since fraudulent transactions are far fewer than legitimate ones, datasets are typically imbalanced [10]. Many machine learning algorithms perform poorly when applied to imbalanced datasets [11], as most learning systems are not designed to handle large disparities between class sizes [12]. The continuous evolution of fraud methods requires adaptive learning algorithms capable of tracking these illegal activities [5].

In prior literature, traditional classification methods for imbalanced datasets have relied on sampling techniques to balance the dataset [11]. Methods such as resampling have been proposed to improve performance. However, imbalance is not the only factor affecting classification difficulty; another critical factor is class overlap, which limits the information available from transaction records regarding the underlying process [13]. Another major challenge in credit card fraud detection is the limited availability of data due to confidentiality concerns, restricting the opportunity to share real datasets and evaluate existing techniques [9].

This paper aims to provide an empirical comparison of different algorithms and modeling techniques using a real-world dataset, focusing on open questions such as: Which machine learning algorithm should be employed? Should the data be analyzed in its original imbalanced form, or should it be balanced first? If balanced, what is the most effective approach? Which performance metrics are most suitable for evaluating the proposed method?

In this study, we examine and compare the performance of three classification techniques and assess the impact of data balancing on performance in the presence of class imbalance. These frameworks are capable of handling evolving, imbalanced data streams. All results are based on experiments conducted using real credit card transaction data from European cardholders in September 2013 [10].

2. RELATED WORK

Many researchers have developed credit card fraud detection and tracking techniques based on data mining approaches. Gash and Reilly proposed a three-layer feedforward neural network (FFNN) for credit card fraud detection, which requires a long training time [14]. Malini and colleagues discussed credit card fraud detection in banking and compared several techniques, including machine learning, genetic programming, fuzzy logic, sequence analysis, and others, for detecting fraudulent transactions. In addition, K-Nearest Neighbors (KNN) and advanced detection methods were implemented to optimize solutions for fraud detection problems. These approaches have proven effective in minimizing false alarm rates while increasing fraud detection accuracy.[15]

In another study, Maria et al. developed an anti-fraud framework by combining two unsupervised algorithms. They used classification to generate data packets, which were subsequently grouped using clustering concepts. This

model was applied for manual implementation on existing data across multiple bank accounts. Principal Component Analysis (PCA), as an unsupervised classification scheme, was employed for transaction classification.[16]

Siddhartha and colleagues evaluated two advanced data mining methods Support Vector Machines (SVM) and Random Forests alongside logistic regression to enhance the identification, control, and legal prosecution of credit card fraud. This study was based on real-world international credit card transaction data.[17]

The focus of much of the literature has been on supervised methods, including neural networks [18], rule-based approaches such as RIPPER [19, 20], and tree-based algorithms like C4.5 [21] and CART [22]. These studies aimed to provide empirical comparisons of various algorithms and modeling techniques on real-world datasets.

Ai-Hua and colleagues examined the effectiveness of classification models for credit card fraud detection. Two different classification methods decision trees, neural networks, and logistic regression were tested for their applicability in fraud detection. Their study provided a useful framework for selecting the optimal model for assessing credit card fraud risk based on various performance criteria.[23]

In several studies, 26 data mining techniques have been applied for financial fraud detection. Extensive research has been conducted on credit card fraud detection, with most systems relying on supervised algorithms, such as neural networks [14, 18, 23–27] and Support Vector Machines.[32–28 ,17]

3. METHODOLOGY

As discussed in the previous sections, the primary challenge in applying machine learning techniques for fraud detection is the imbalance between normal and fraudulent transaction classes. Since the number of fraudulent transactions is significantly smaller than that of normal transactions, the dataset exhibits a highly imbalanced distribution. The proposed methodology is illustrated in Figure 1.

In this study, we aim to evaluate the performance of machine learning methods on both imbalanced and balanced datasets. In the first step, machine learning algorithms are applied to the original imbalanced dataset to assess the impact of class imbalance on algorithm performance. Classification models are constructed and evaluated using well-established methods, including C5.0 decision trees, Support Vector Machines (SVM), and neural networks.

Studies have shown that a balanced dataset generally provides better overall classification performance compared to imbalanced datasets [33]. Accordingly, based on prior research, in the second step, before implementing the learning algorithms, resampling techniques are employed to balance the sample space of an imbalanced dataset, thereby reducing the impact of class distribution on the learning process [5]. Resampling methods are versatile, as they are independent of the chosen classifier [34].

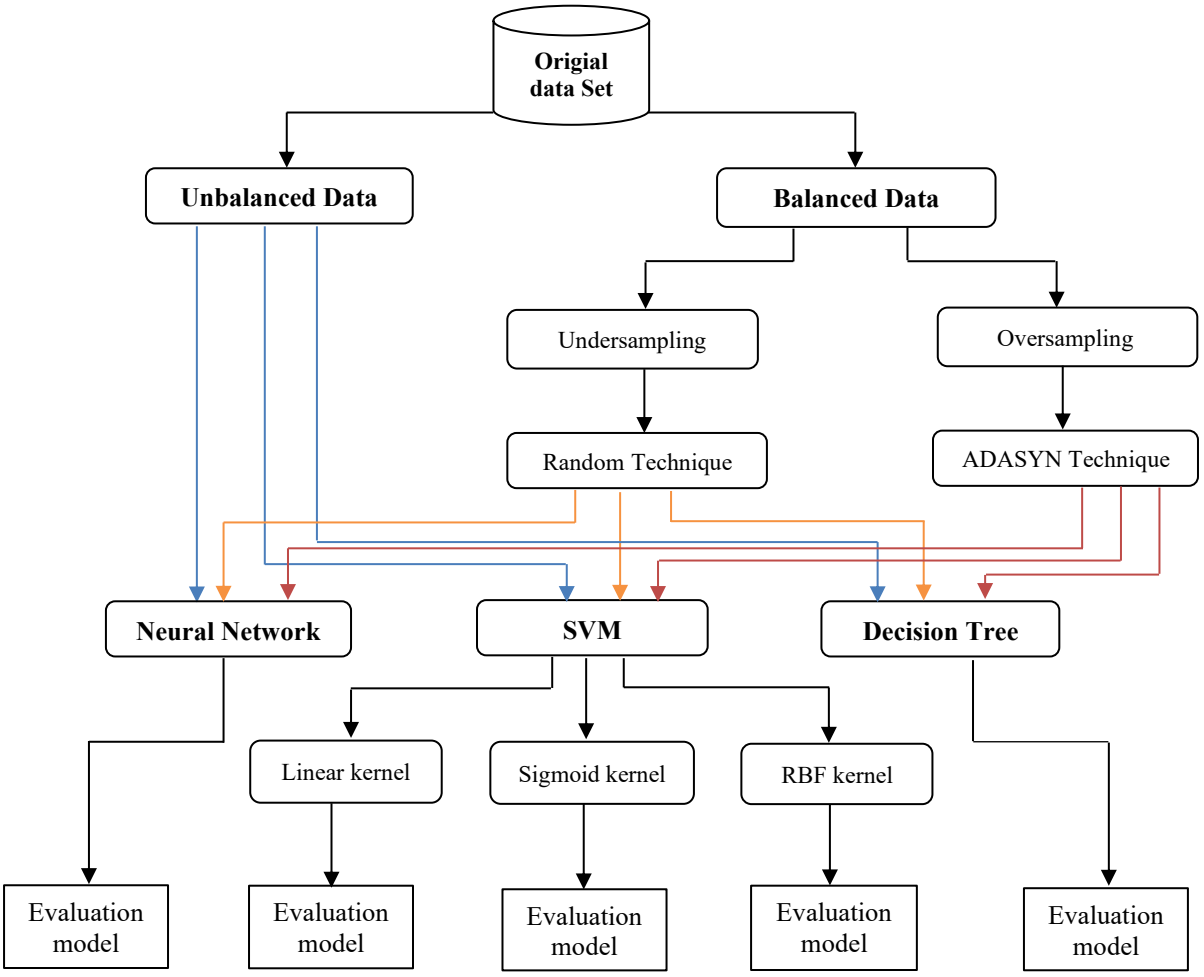
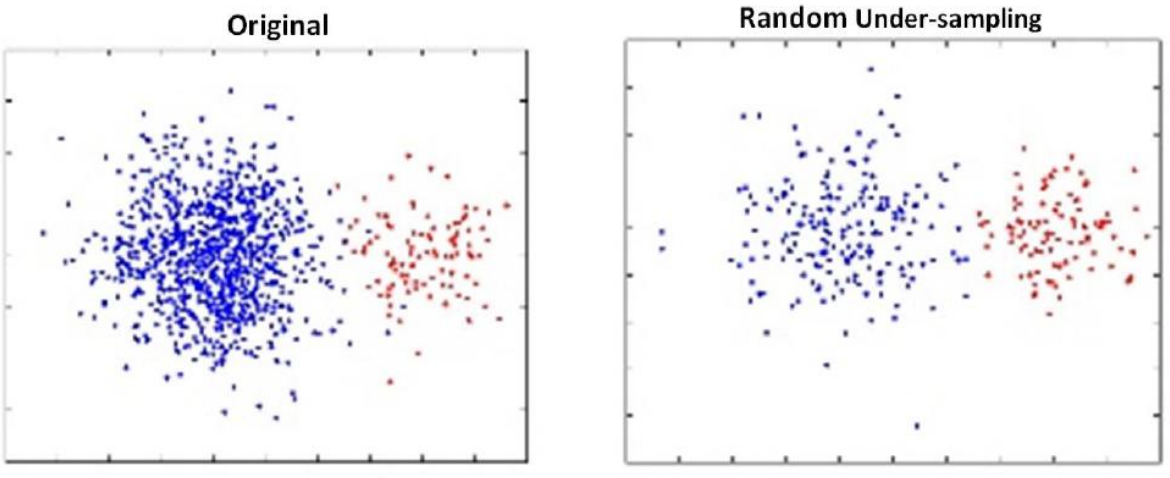


Fig. 1. Proposed Methodology Diagram

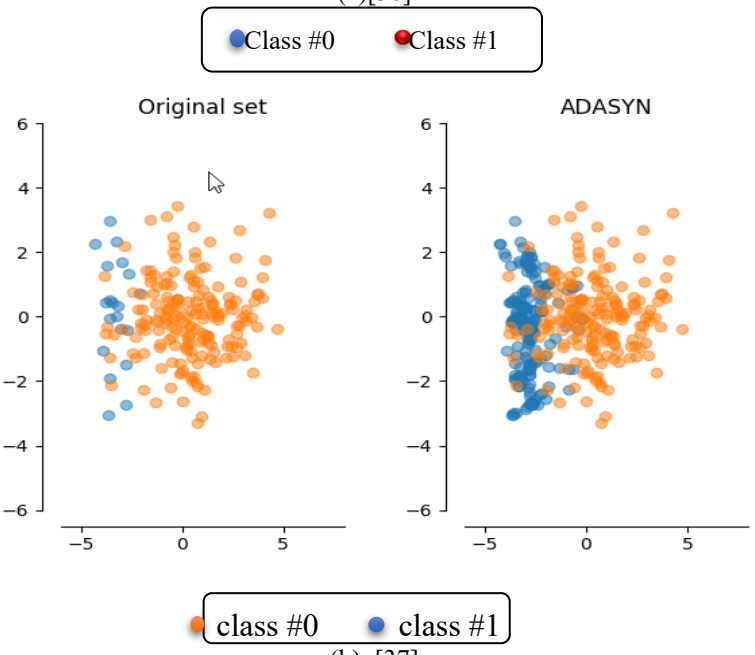
3.1. Data Balancing Methods

Under-Sampling: This method mitigates the adverse effects of imbalanced class distribution by discarding inherent samples from the majority class. The simplest approach is Random Under-Sampling (RUS), illustrated in Figure 2(a), which involves randomly removing examples from the majority class [35]. In this study, this method has been employed.

Over-Sampling: Over-sampling addresses class imbalance by increasing the number of samples in the minority class, repeating them until both classes have equal frequency [10, 36]. Adaptive Synthetic Sampling (ADASYN) is a widely used over-sampling method for generating minority class samples and is considered an improved version of SMOTE [37], as shown in Figure 2(b). In this study, ADASYN is applied. ADASYN not only reduces the learning bias introduced by the original imbalanced data distribution but also adjusts the decision boundary to focus on learning from difficult-to-classify samples [37].



(a)[38]



(b) [37]

Fig. 2. (a) RUS Sampling Technique; (b) ADASYN Sampling Technique

3.2. Data Classification Techniques

Decision Trees (DT): Decision trees are among the most widely used machine learning algorithms [39]. They are considered a white-box technique, meaning they are easily interpretable and computationally inexpensive [40]. Depending on the modeling objective, decision trees can be used for either classification or regression [41]. In this study, we focus on binary classification decision trees. A decision tree is a hierarchical structure that attempts to separate records into distinct binary subgroups. Each decision tree method uses its own splitting algorithm and metric. Well-known decision tree algorithms include ID3, C5.0, and CART [42]. ID3 uses information gain, C5.0 uses gain ratio, and CART employs the Gini index to measure impurity [2, 28]. In this research, the C5.0 classification model is applied.

Support Vector Machines (SVM): SVMs are statistical learning techniques [43] that have proven highly effective in various classification tasks. Unlike decision trees, SVMs aim to find a hyperplane that separates two classes while minimizing classification errors. In its simplest form, SVM uses a linear hyperplane to construct a maximum-margin classifier [44], as proposed by Vapnik and his team at AT&T Bell Laboratories [45].

Due to the highly imbalanced nature of the data (fraudulent vs. non-fraudulent cases), extracting meaningful features for detecting fraudulent transactions is challenging. The simplicity of linear classification and the ability to operate in a high-dimensional feature space make SVMs particularly suitable for fraud detection tasks [17]. The strength of SVMs arises from two key properties: kernel representation and margin optimization [17]. In many cases, finding a suitable linear hyperplane in the input space is limited [46]. A kernel function [47] allows the inner product of two data points to be represented in a high-dimensional feature space. Several kernel functions exist for SVMs; in this study, we use three common kernels: Linear, RBF, and Sigmoid. Selecting an appropriate kernel depends on the nature of the classification task and the input dataset.

Artificial Neural Networks (ANN): Neural networks are defined as a collection of interconnected nodes designed to mimic the functionality of the human brain [48]. ANNs, or artificial neural networks, are mathematical or computational models inspired by the structure and functional aspects of biological neural networks. ANNs are typically adaptive systems that adjust their structure based on external or internal information flowing through the network during the learning phase. Modern neural networks are generally used to model complex relationships between inputs and outputs or to detect patterns in data [49].

A Multilayer Perceptron (MLP) is an ANN model that maps input datasets to a set of corresponding outputs. MLPs are trained using the backpropagation algorithm, a supervised learning method divided into two phases: forward propagation and weight updating. These phases are repeated iteratively until the network performance is satisfactory. MLPs consist of multiple layers of computational units, usually connected in a feedforward manner. In many applications, these units employ a sigmoid activation function. When the learning data represents numerical analysis, the output nodes typically use a sigmoid function.

After balancing the dataset using the resampling techniques described previously, the three classification algorithms (C5.0, SVM, and ANN) are applied to classify the data. Finally, the resulting models are compared using various evaluation metrics to assess how machine learning algorithms perform when exposed to balanced versus imbalanced datasets.

4. EXPERIMENTAL RESULTS

In this section, we present the evaluation metrics used for analyzing results, introduce the dataset, and provide the obtained outcomes. The selected methods for constructing classification models include C5.0 decision trees, SVM with three kernels (Sigmoid, RBF, and Linear), and artificial neural networks (ANNs). All methods are applied to three datasets. In the first phase, the models are trained on the original imbalanced data. In the next phase, the data are balanced using preprocessing techniques such as Random Under-Sampling (RUS) and ADASYN over-sampling. Finally, the results are compared to assess the impact of data balancing.

For experiments involving SVM, the k-fold cross-validation method [50] is employed with k=10k=10k=10, meaning that one subset is used for testing and the remaining nine for training, repeated ten times. This validation approach helps determine how well the model generalizes to unseen data.

Data preprocessing was performed using Python, and the proposed methodology was implemented in RapidMiner Studio 9.

4.1. Evaluation Metrics

Several classification performance metrics commonly cited in the literature are employed. Overall accuracy is an insufficient performance indicator in this study because of the significant class imbalance. Predicting all cases as the majority class can yield a misleadingly high accuracy. Therefore, additional evaluation metrics are used, including:

Sensitivity

Specificity

Accuracy

Recall

F-measure

Area Under the Curve (AUC)

For binary classification problems, it is standard to construct a confusion matrix [9]. A confusion matrix is one of the most widely used methods for evaluating classification models, with actual labels along one axis and predicted labels along the other. Each cell in the table counts the number of predictions falling into a given category. Correct predictions appear along the main diagonal, allowing for quick comparison of performance across different classes.

Each metric is defined in terms of the confusion matrix, as illustrated in Figure 3. The abbreviations used in the matrix are: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), where positive refers to fraudulent cases and negative to legitimate cases [10].

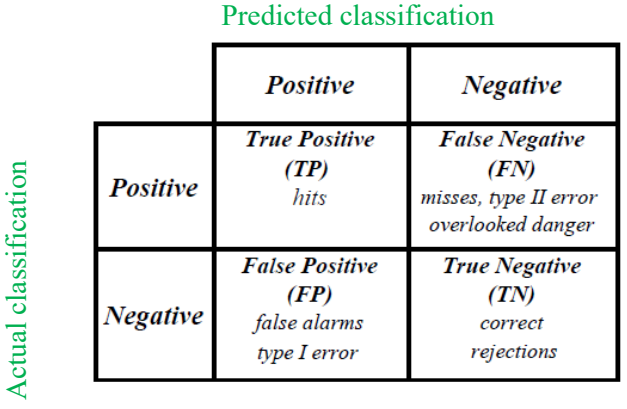


Fig. 3. Confusion Matrix [5]

Accuracy [29] measures the proportion of correctly classified instances to the total number of classified instances:

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Sensitivity (True Positive Rate, TPR) indicates the model's ability to correctly identify fraudulent cases:

$$\text{Sensitivity or TPR}=\frac{TP}{P}=\frac{TP}{TP+FN} \tag{2}$$

Specificity (True Negative Rate, TNR) measures the accuracy in identifying legitimate cases:

$$\text{Specificity or TNR}=\frac{TN}{N}=\frac{TN}{TN+FP} \tag{3}$$

Recall is defined as:

$$\text{RECALL}=\frac{TP}{TP+FN} \tag{4}$$

Precision is calculated as:

$$\text{Precision}=\frac{TP}{TP+FP} \tag{5}$$

F-Measure is the harmonic mean of recall and precision:

$$\text{F\_Measure} = \frac{2PR}{P+R} \quad (6)$$

When evaluating or visualizing the performance of multiclass classification problems, the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) are commonly used. The ROC curve plots the True Positive Rate (TPR) on the Y-axis against the False Positive Rate (FPR) on the X-axis. The AUC is a critical metric for assessing the performance of a classification model [51].

The AUC considers all possible thresholds, which can be used to evaluate probabilistic predictions. Different thresholds yield different TP and FP rates. Lowering the threshold allows the model to predict more positive cases but may increase the FP rate simultaneously [52].

An excellent model will have an AUC close to 1, indicating strong discriminative ability between the two classes. A poor model will have an AUC near 0, indicating poor distinction capability. An AUC of 0.5 suggests that the model has no discriminative power and performs no better than random guessing [51].

4.2. Dataset

The dataset contains information on credit card transactions, including examples of fraudulent activities. It was first utilized in the experiments reported in [10]. The dataset comprises credit card transaction records from European cardholders collected in September 2013. It is publicly available at: http://www.ulb.ac.be/di/map/adalpozz/data/creditcard.Rdata.

The dataset represents transactions over a two-day period and includes 30 numerical input features and a binary output variable. Only three features Time, Amount, and Class are directly interpretable; the remaining features (V1–V28) have been transformed using Principal Component Analysis (PCA) to preserve confidentiality. The “Time” feature indicates the elapsed time between each transaction and the first transaction in the dataset. The “Amount” feature denotes the transaction amount, which can be used, for example, in cost-sensitive learning. The “Class” variable represents the response: 1 indicates a fraudulent transaction, and 0 indicates a legitimate transaction. Cardholder identifiers are not available, so each transaction can be considered independent.

This is one of the rare publicly available datasets for credit card fraud detection. The dataset is highly imbalanced: fraudulent transactions constitute only 0.172% of the total, with 492 fraudulent transactions among 284,807 records [5].

5. RESULTS

5.1. Experiments on the Imbalanced Dataset

The performance of the SVM algorithm with three kernel functions on the imbalanced dataset is summarized in Table 1. Although all three kernels achieved an overall accuracy of 99%, Figure 4 shows that they all appear to have excellent accuracy. However, this result is misleading and unacceptable because both sensitivity and recall are extremely low, while specificity is 99%. This indicates the presence of severe class imbalance and the dominance of the negative class over the positive class. Consequently, the model classifies all positive (fraudulent) instances as negative (legitimate).

Table 1. Cross-validation results of three SVM kernels on the imbalanced dataset

SVM Kernel	Accuracy	Sensitivity	Specificity	Precision	Recall	F-measure	AUC
Linear	99.87%	33.95%	99.99%	80.96%	33.95%	47.64%	0.965
RBF	99.87%	27.83%	99.99%	83.81%	27.83%	41.64%	0.975
Sigmoid	99.84%	13.41%	99.99%	63.01%	13.41%	21.91%	0.555

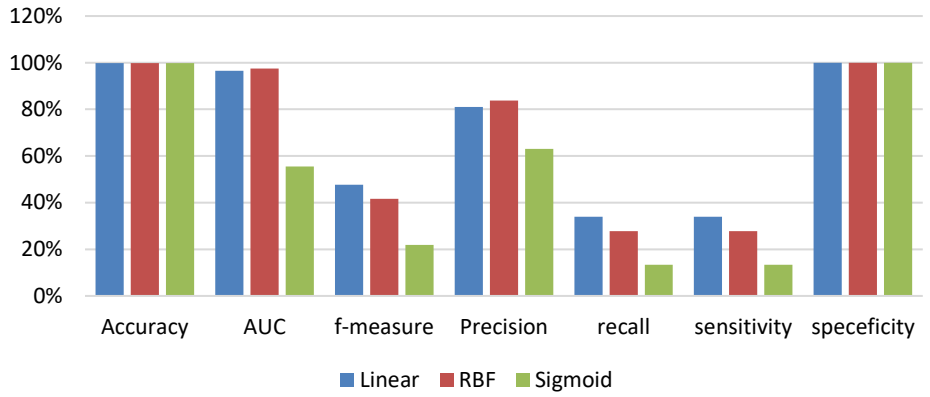


Fig. 4. Comparison of the performance of three SVM kernels on the imbalanced dataset

The cross-validation results of the SVM algorithm with its best-performing kernel from the previous stage, compared with Neural Network (NN) and Decision Tree (DT) algorithms on the imbalanced dataset, are shown in Table 2. Both NN and DT techniques exhibit high values in sensitivity and specificity, which indicate the detection rates for fraudulent and legitimate transactions. However, given the severe imbalance between the two classes, these methods fail to properly reflect the impact of data imbalance on the model. Overall, SVM demonstrates more reliable performance on imbalanced data.

Table 2. Cross-validation results of SVM, Neural Network, and Decision Tree on the imbalanced dataset

Algorithm	Accuracy	Sensitivity	Specificity	Precision	Recall	F-measure	AUC
SVM	99.87%	33.95%	99.99%	80.96%	33.95%	47.64%	0.965
NN	99.95%	80.08%	99.98%	88.30%	80.08%	83.76%	0.952
DT	99.93%	72.75%	99.98%	85.98%	72.75%	78.63%	0.858

As illustrated in Figure 5, the SVM algorithm better captures the imbalance present in the dataset compared to the other two techniques, as reflected by the sensitivity and specificity metrics.

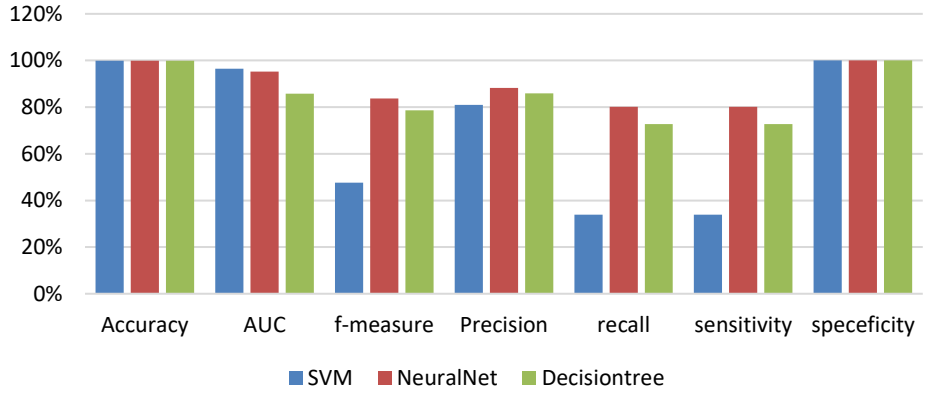


Fig. 5. Comparison of the performance of SVM, NN, and DT on the imbalanced dataset

Applying the RUS (Random Under Sampling) technique to this dataset leads to the removal of intrinsic samples from the majority class and the loss of useful and valuable information [33]. This explains why some metrics in Table 3 reach 100%, while the accuracy parameter which reflects overall classification performance retains reasonable and reliable values across all three SVM kernels.

From the sensitivity and specificity values across the three kernels, it is evident that the dataset has achieved a more balanced level, yielding more precise results. As shown in Figure 6, the Linear kernel demonstrates superior performance in most metrics, such as Accuracy, Precision, F-measure, and AUC. Among the remaining kernels, RBF ranks second due to higher values in key metrics like Accuracy, AUC, and Precision, while the Sigmoid kernel ranks last.

Table 3. Cross-validation results of three SVM kernels on the dataset balanced using RUS

SVM	Accuracy	Sensitivity	Specificity	Precision	Recall	F-measure	AUC
Linear	97.12%	85.91%	100.00%	100.00%	85.91%	92.37%	0.983
RBF	96.64%	83.54%	100.00%	100.00%	83.54%	90.98%	0.983
Sigmoid	96.98%	86.58%	99.65%	98.52%	86.58%	92.12%	0.972

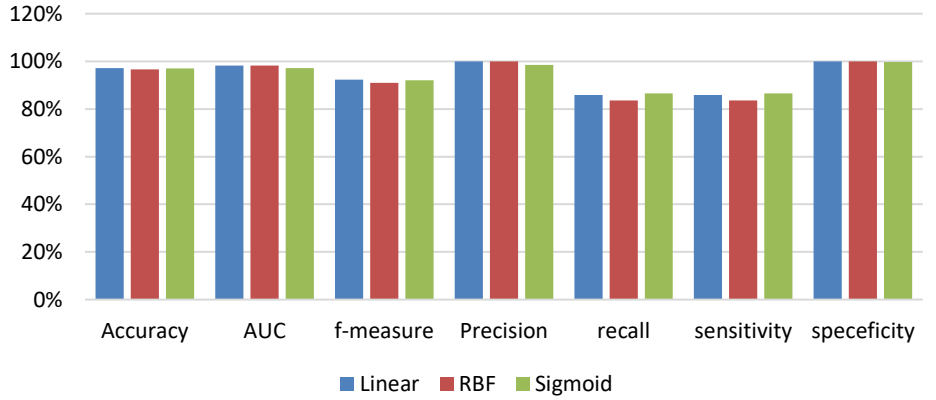


Fig. 6. Comparison of the performance of three SVM kernels on the dataset balanced using RUS

In Table 4 and Figure 7, a comparison of the three classifiers SVM, DT, and NN shows that SVM performs better in terms of Precision, Specificity, and AUC, while maintaining a balanced performance across both classes. The Decision Tree (DT) demonstrates slightly higher overall accuracy, whereas the Neural Network (NN) achieves higher Sensitivity, reflecting its better detection rate for fraudulent transactions. The differences in the F-measure among the three techniques are minimal.

Given that the AUC metric is crucial for evaluating classification models, SVM can be considered the superior classifier in this study, as it achieves the highest AUC value.

Table 4. Cross-validation results of SVM, DT, and NN on the RUS-balanced dataset

Classifier	Accuracy	Sensitivity	Specificity	Precision	Recall	F-measure	AUC
SVM	97.12%	85.91%	100.00%	100.00%	85.91%	92.37%	0.983
NN	97.15%	90.26%	98.92%	88.34%	90.26%	92.83%	0.976
DT	97.22%	87.41%	99.74%	98.89%	87.41%	92.76%	0.889

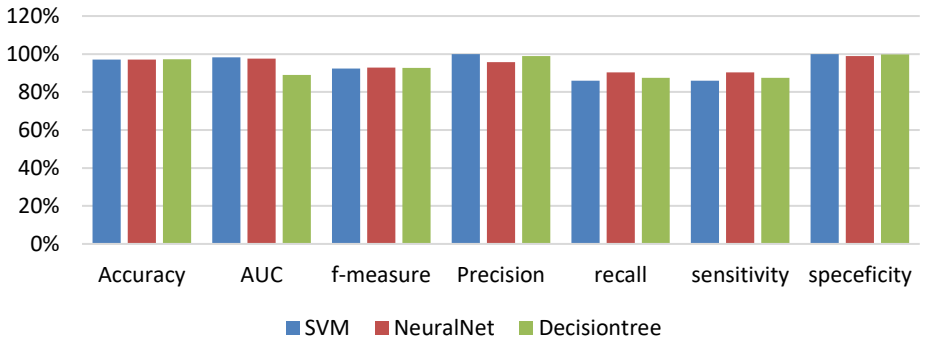


Figure 7. Comparison of the performance of SVM, DT, and NN on the RUS-balanced dataset

As shown in Figure 8, the ROC curve illustrates the True Positive Rate (TPR) against the False Positive Rate (FPR) for different validation thresholds. At this stage, after balancing the dataset, the AUC values for all three techniques are close to 1, which is considered ideal. This indicates the strong capability of the classifiers to distinguish between the two classes with minimal overlap. Among the three, the ROC curve of the SVM method demonstrates a slightly superior performance.

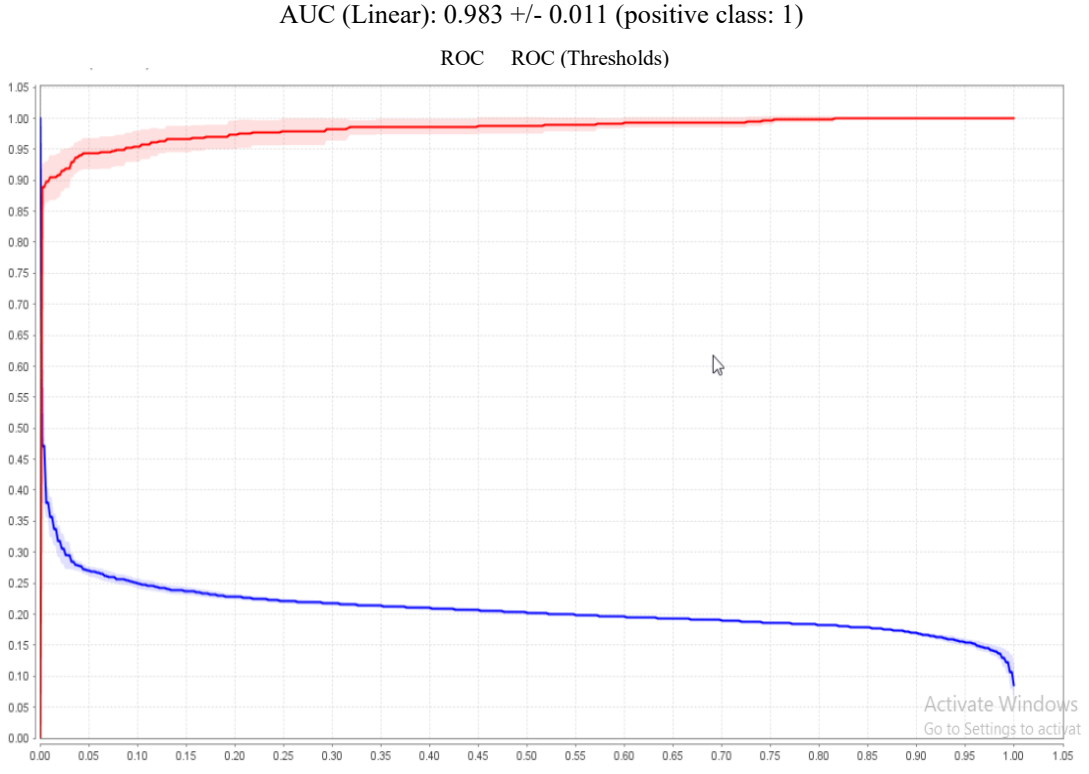


Fig. 8. Comparison of AUC and ROC of SVM, NN, and DT tests on the RUS-balanced dataset

As shown in Table 5 and Figure 9, it is clearly observed that the Linear kernel achieves the highest values and the best overall performance, while the Sigmoid kernel shows the worst performance. However, the RBF kernel also demonstrates strong performance, with only a slight difference compared to the Linear kernel, making their results nearly equivalent. No significant superiority can be assigned between these two kernels; their performance is almost identical.

Table 5. Cross-validation results of three SVM kernels on the ADASYN-balanced dataset

SVM	Acc	Sensitivity	Specificity	Precision	Recall	F	AUC
Linear	97.76%	88.98%	99.95%	99.79%	88.98%	94.08%	0.996
RBF	97.17%	86.02%	99.96%	99.79%	86.02%	92.40%	0.996
Sigmoid	74.80%	37.13%	84.22%	37.05%	37.13%	37.09%	0.637

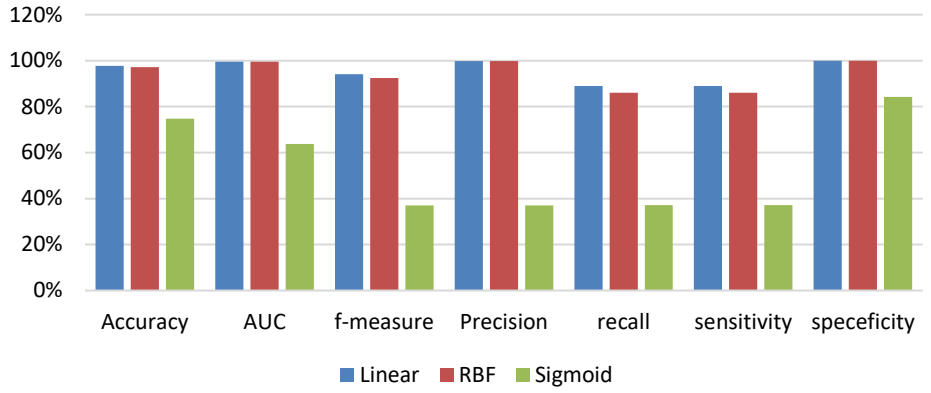


Fig. 9. Comparison of SVM, DT, and NN performance on the ADASYN-balanced dataset

Based on Table 6 and the corresponding chart (Figure 10), the NN algorithm demonstrates the highest values for metrics such as Accuracy, Sensitivity, F-measure, and AUC. This indicates that NN provides a better model on the balanced dataset because metrics like Accuracy, which evaluates the correctness of the generated model and classification precision, and Sensitivity, which measures the detection of true positives (i.e., frauds), are high. Additionally, AUC, which is a valuable metric for evaluating classifier performance, is close to 1.

SVM also performs well in metrics like Sensitivity, Precision, and AUC, but DT shows better results than SVM for Accuracy, Sensitivity, Recall, and F-measure, placing it second after NN in overall performance.

Table 6. Cross-validation results of SVM, DT, and NN on the ADASYN-balanced dataset

Technique	Acc	Sensitivity	Specificity	Precision	Recall	F	AUC
SVM	**97.76%**	**88.98%**	**99.95%**	**99.79%**	**88.98%**	**94.08%**	**0.996**
NN	**99.73%**	**99.05%**	**99.91%**	**99.62%**	**99.05%**	**99.33%**	**0.999**
DT	**98.09%**	**91.30%**	**99.79%**	**99.08%**	**91.30%**	**95.03%**	**0.956**

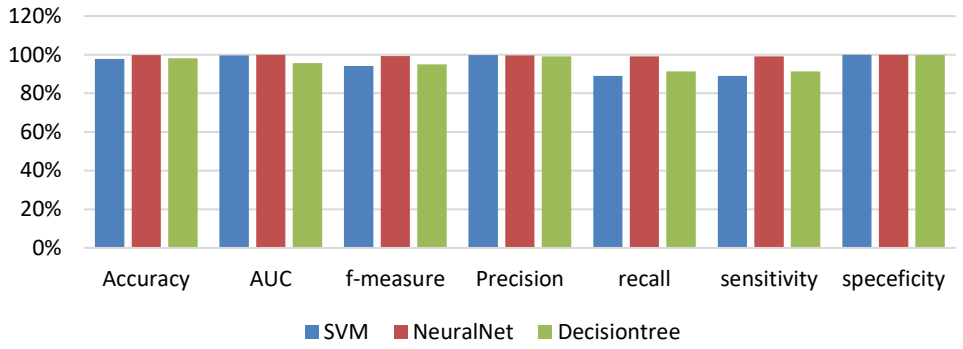


Figure 10. Comparison of SVM, DT, and NN performance on the ADASYN-balanced dataset

At this stage, all three techniques demonstrate very good results on the balanced dataset. However, the NN technique, with its excellent AUC value and ideal ROC curve, is identified as the best model. See Figure 11.

AUC(NN): 0.999 +/- 0.000 (positive class: 1)

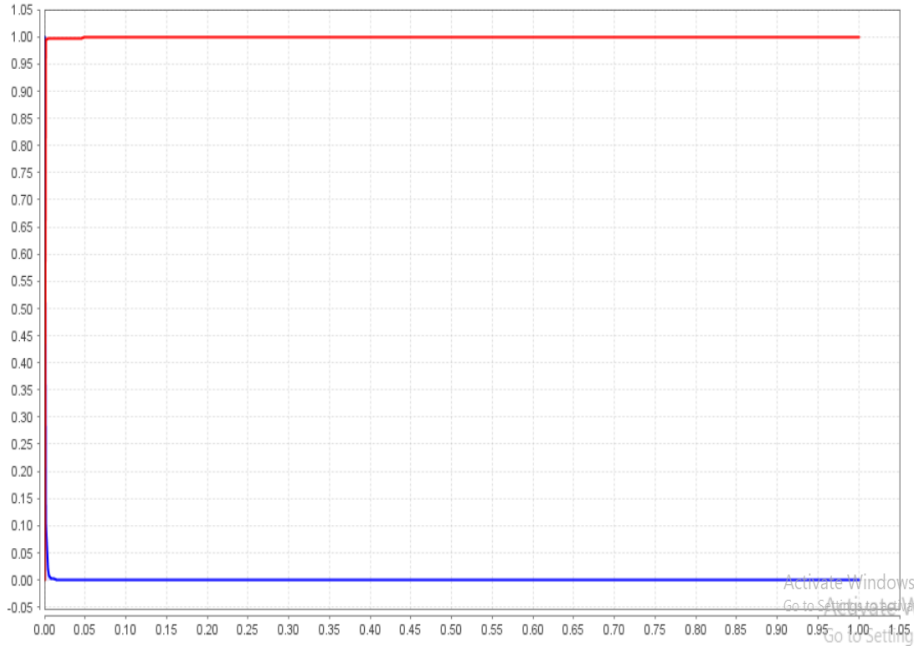


Figure 11. Comparison of AUC and ROC for SVM, NN, and DT on the ADASYN-balanced dataset

6. CONCLUSION

In this study, the performance of three advanced data mining techniques C5.0 decision tree, artificial neural network (ANN), and support vector machines (SVM) with three kernel functions was evaluated for credit card fraud detection. A real dataset containing credit card transactions with fraudulent samples, conducted by European cardholders in September 2013, was analyzed.

SVM and neural networks have recently attracted attention due to their superior performance across various applications. Because fraudulent transactions are far fewer than legitimate ones, sampling techniques are necessary to obtain a training dataset with a sufficient proportion of fraudulent to non-fraudulent cases. Both undersampling and oversampling methods were applied, and the performance of the three techniques was compared. This study

considers both traditional comparative performance metrics and specific criteria relevant to the practical deployment of such models.

Overall, all three techniques demonstrated sufficient capability for modeling fraud detection on the dataset. Their performance differed on imbalanced and balanced datasets. On the imbalanced dataset, all three techniques achieved high accuracy because the majority class (legitimate transactions) dominated, causing fraudulent transactions to be classified as legitimate. However, these predictions are biased due to the lack of minority-class information. Machine learning algorithms generally assume datasets are balanced across classes and, therefore, tend to classify any test instance into the majority class to improve apparent accuracy this represents a significant challenge that is often overlooked.

Comparing the evaluation metrics of the three techniques on the imbalanced dataset, SVM showed strong ability to detect class imbalance. Among its three kernel functions, Linear and RBF performed similarly well, while the Sigmoid kernel performed poorly.

Regarding sampling methods, ADASYN produced more accurate models and greater improvements in algorithm performance than RUS. While all three classifiers performed similarly when undersampling the majority class, valuable information was lost, limiting the reliability of this approach. After oversampling the minority class and applying all three algorithms, NN outperformed the other two techniques.

Among the parameters used to evaluate model performance, Accuracy and AUC were the most appropriate overall, whereas Sensitivity and Specificity are particularly informative for highly imbalanced datasets like this one.

Transparency Statement

The data supporting this study are available upon reasonable request to the corresponding author, subject to ethical and confidentiality considerations.

Acknowledgments

We would like to express our gratitude to all individuals who contributed to this project.

Declaration of Interest

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

REFERENCES

[1] Nasiri, N., Minayi, B., & Farjami, Y. (2010). Application of data mining methods in electronic banking for detecting suspicious financial transactions [in Persian]. Qom University – Faculty of Engineering.

[2] Chye Koh, H., & Kee Low, C. (2004). Going concern prediction using data mining techniques. Managerial Auditing Journal, 19(3), 462–476. https://doi.org/10.1108/02686900410524436

[3] Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. Statistical Science, 17(3), 235–249. https://doi.org/10.1214/ss/1042727940

[4] Nisbet, R., Elder, J., & Miner, G. (2009). Handbook of statistical analysis and data mining applications.

Academic Press.

[5] Dal Pozzolo, A. (2015). Adaptive machine learning for credit card fraud detection (Doctoral dissertation).

[6] Bolton, R. J., & Hand, D. J. (2001). Unsupervised profiling methods for fraud detection. Credit Scoring and Credit Control VII, 235–255.

[7] Zaslavsky, V., & Strizhak, A. (2006). Credit card fraud detection using self-organizing maps. Information & Security: An International Journal, 18(3), 48–63. https://doi.org/10.11610/isij.1803

[8] Vala, H. M., & Nejad, D. F. (2015). Detecting fraud in banking transactions using data mining: A case study on Mehr Eqtesad Bank transactions [in Persian].

[9] Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. Expert Systems with Applications, 41(10), 4915–4928. https://doi.org/10.1016/j.eswa.2014.02.026

[10] Dal Pozzolo, A., et al. (2015). Calibrating probability with undersampling for unbalanced classification. 2015 IEEE Symposium Series on Computational Intelligence, 33–40. IEEE. https://doi.org/10.1109/SSCI.2015.33

[11] Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. Intelligent Data Analysis, 6(5), 429–449. https://doi.org/10.3233/IDA-2002-6504

[12] Batista, G. E., Carvalho, A. C., & Monard, M. C. (2000). Applying one-sided selection to unbalanced datasets. In Mexican International Conference on Artificial Intelligence (pp. 315–325). Springer. https://doi.org/10.1007/10720076_29

[13] Holte, R. C., Acker, L., & Porter, B. W. (1989). Concept learning and the problem of small disjuncts. In IJCAI (pp. 813–818).

[14] Ghosh, S., & Reilly, D. L. (1994). Credit card fraud detection with a neural network. In Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences (pp. 621–630). IEEE. https://doi.org/10.1109/HICSS.1994.323314

[15] Malini, N., & Pushpa, M. (2017). Analysis on credit card fraud identification techniques based on KNN and outlier detection. 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB) (pp. 255–259). IEEE. https://doi.org/10.1109/AEEICB.2017.7972424

[16] Lepoivre, M. R., et al. (2016). Credit card fraud detection with unsupervised algorithms. Journal of Advances in Information Technology, 7(1), 34–38. https://doi.org/10.12720/jait.7.1.34-38

[17] Bhattacharyya, S., et al. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), 602–613. https://doi.org/10.1016/j.dss.2010.08.008

[18] Dorronsoro, J. R., et al. (1997). Neural fraud detection in credit card operations. IEEE Transactions on Neural Networks, 8(4), 827–834. https://doi.org/10.1109/72.595879

[19] Clark, P., & Niblett, T. (1989). The CN2 induction algorithm. Machine Learning, 3(4), 261–283. https://doi.org/10.1023/A:1022641700528

[20] Cohen, W. W. (1995). Fast effective rule induction. In Machine Learning Proceedings 1995 (pp. 115–123). Elsevier.
https://doi.org/10.1016/B978-1-55860-377-6.50023-2

[21] Quinlan, J. R. (2014). C4.5: Programs for machine learning. Elsevier.

[22] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Wadsworth International Group.

[23] Shen, A., Tong, R., & Deng, Y. (2007). Application of classification models on credit card fraud detection. In 2007 International Conference on Service Systems and Service Management (pp. 1–4). IEEE. https://doi.org/10.1109/ICSSSM.2007.4280163

[24] Aleskerov, E., Freisleben, B., & Rao, B. (1997). Cardwatch: A neural network based database mining system for credit card fraud detection. In Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr) (pp. 220–226). IEEE.

[25] Brause, R., Langsdorf, T., & Hepp, M. (1999). Neural data mining for credit card fraud detection. In Proceedings 11th International Conference on Tools with Artificial Intelligence (pp. 103–106). IEEE.

[26] Maes, S., et al. (2002). Credit card fraud detection using Bayesian and neural networks. Proceedings of the 1st International Naiso Congress on Neuro Fuzzy Technologies.

[27] Syeda, M., Zhang, Y.-Q., & Pan, Y. (2002). Parallel granular neural networks for fast credit card fraud detection. IEEE World Congress on Computational Intelligence (FUZZ-IEEE'02). https://doi.org/10.1109/FUZZ.2002.1007255

[28] Şahin, Y. G., & Duman, E. (2011). Detecting credit card fraud by decision trees and support vector machines. 2011 International Symposium on Innovations in Intelligent Systems and Applications, 594–598. https://doi.org/10.1109/INISTA.2011.5946108

[29] Ravale, U., Marathe, N., & Padiya, P. (2015). Feature selection based hybrid anomaly intrusion detection system using K-means and RBF kernel function. Procedia Computer Science, 45, 428–435. https://doi.org/10.1016/j.procs.2015.03.174

[30] Krupka, T. (2016). SVM classifiers and heuristics for feature selection.

[31] Romero, R., Iglesias, E., & Borrajo, L. (2015). A linear-RBF multikernel SVM to classify big text corpora. BioMed Research International, 2015, 878291. https://doi.org/10.1155/2015/878291

[32] Singh, G., et al. (2012). A machine learning approach for detection of fraud based on SVM. International Journal of Scientific Engineering and Technology, 1(3), 194–198.

[33] Rout, N., Mishra, D., & Mallick, M. K. (2018). Handling imbalanced data: A survey. In International Proceedings on Advances in Soft Computing, Intelligent Systems and Applications (pp. 431–443). Springer. https://doi.org/10.1007/978-981-10-5272-9_39

[34] López, V., et al. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information Sciences, 250, 113–141. https://doi.org/10.1016/j.ins.2013.07.007

[35] Tahir, M. A., et al. (2009). A multiple expert approach to the class imbalance problem using inverse random undersampling. In International Workshop on Multiple Classifier Systems (pp. 82–91). Springer. https://doi.org/10.1007/978-3-642-02326-2_9

[36] Chawla, N. V., et al. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953

[37] He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence) (pp. 1322–1328). IEEE. https://doi.org/10.1109/IJCNN.2008.4633969

[38] Ganesh Kumar, R. D., Mohan, K. R., Jagan Mohan, R., & Chakraborty, G. (2016). Predicting rare events using specialized sampling techniques in SAS (p. 7).

[39] Rokach, L., & Maimon, O. Z. (2008). Data mining with decision trees: Theory and applications (Vol. 69). World Scientific. https://doi.org/10.1142/9789812771728

[40] Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

[41] Bahnsen, A. C., Aouada, D., & Ottersten, B. (2015). Example-dependent cost-sensitive decision trees. Expert Systems with Applications, 42(19), 6609–6619. https://doi.org/10.1016/j.eswa.2015.04.042

[42] Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. Expert Systems with Applications, 32(4), 995–1003. https://doi.org/10.1016/j.eswa.2006.02.016

[43] Vapnik, V. N. (1998). Statistical learning theory. Wiley.

[44] Kecman, V. (2001). Learning and soft computing: Support vector machines, neural networks, and fuzzy logic models. MIT Press.

[45] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1023/A:1022627411411

[46] Phienthrakul, T., & Kijsirikul, B. (2005). Evolutionary strategies for multi-scale radial basis function kernels in support vector machines. In Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation (pp. 987–993). ACM. https://doi.org/10.1145/1068009.1068160

[47] Schölkopf, B., Smola, A. J., & Bach, F. (2002). Learning with kernels: Support vector machines, regularization, optimization, and beyond. MIT Press. https://doi.org/10.7551/mitpress/4175.001.0001

[48] Chaudhary, K., Yadav, J., & Mallick, B. (2012). A review of credit card fraud detection techniques. International Journal of Computer Applications, 45(1), 39–44.

[49] Chen, W. H., Hsu, S. H., & Shen, H. P. (2005). Application of SVM and ANN for intrusion detection. Computers & Operations Research, 32(10), 2617–2634. https://doi.org/10.1016/j.cor.2004.03.019

[50] Anohhin, I., Võhandu, L., & Emeritus, P. (2017). Data mining and machine learning for fraud detection (Master's thesis). Tallinn University of Technology, Faculty of Information Technology.

[51] Beitzel, S. (2006). On understanding and classifying web queries (PhD thesis). Illinois Institute of Technology.

[52] Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Systems with Applications, 39(3), 3446–3453. https://doi.org/10.1016/j.eswa.2011.09.033