Indonesian Journal of Computer Science
ISSN 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam No. 1, Padang, Indonesia
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Credit Card Fraud Detection Based on Machine Learning Classification Algorithm

Bareq Mardan[1], Adnan Mohsin Abdulazeez[2]
bareq.noaman@dpu.edu.krd [1] , adnan.mohsin@dpu.edu.krd [2]
[1]Technical College of Duhok, Duhok Polytechnic University, Kurdistan Region, Iraq
[2]Technical College of Engineering, Duhok Polytechnic University, Kurdistan Region, Iraq

Article Information

Submitted : 9 May 2024
Reviewed : 20 May 2024
Accepted : 15 Jun 2024

Keywords

Risk Analysis, Prediction, Credit Card, Machine Learning Algorithm, Credit Card Risk

Abstract

Credit risk analysis is a critical process in the financial industry, as it helps lenders assess the likelihood of borrowers defaulting on their loans. With the advent of machine learning algorithms, there has been a growing interest in leveraging these techniques for more accurate and efficient credit risk prediction. Traditional credit risk models often rely on manual processes and limited data sources, resulting in potential biases and inaccuracies. Additionally, the rapid growth of credit card usage and the increasing complexity of financial transactions have made it challenging to accurately assess credit risk using conventional methods. This review paper aims to provide a comprehensive overview of machine learning algorithms used for credit risk prediction in the context of credit card lending. It explores classification techniques and their applications in credit risk analysis. The paper also discusses the challenges and limitations associated with these algorithms, including data quality, overfitting, and interpretability.

A. Introduction

Credit card fraud poses major risks and costs for financial institutions globally, with losses estimated at over $30 billion annually [1]. Traditional rule-based fraud detection systems rely on cumbersome manual rule engineering which struggles to keep pace with the evolving tactics of sophisticated fraudsters [2]. Moreover, such systems often suffer from unacceptably high false positive rates, negatively impacting the customer experience [3][4]. Machine learning has emerged as a promising approach for developing more accurate predictive models that can adapt to changing fraud patterns without extensive manual work . Recent studies have shown machine learning algorithms such as random forests, neural networks, and ensemble methods achieve high fraud detection performance when applied to credit card transaction data [5][6]. However, open questions remain regarding several key factors important for real-world implementation, including model performance across different environments, explain ability of predictions, and suitability for operational use in high-risk financial applications that demand transparency and accountability [7][8].

This paper presents a rigorous comparative evaluation of popular machine learning algorithms for the task of credit card fraud risk analysis and prediction. Models are trained and tested on a large real-world transaction dataset and objectively assessed based on predictive power as well as issues like class imbalance handling and interpretability [9][10]. The most effective and transparent models are identified according to their ability to balance predictive performance with characteristics necessary for use in financial risk analysis systems that demand trusted decision-making [11][12].The main aim of this work A performance benchmark of classification algorithms for credit card fraud detection, Identification of suitable machine learning approaches through consideration of multiple factors beyond predictive accuracy alone, Guidance for stakeholders on responsibly applying advanced analytics for credit risk assessment. The results provide insights for progressing fraud detection capabilities in a manner aligned with expectations of the financial sect. Figure 1. It focuses on the ML integration approach to support credit card detection

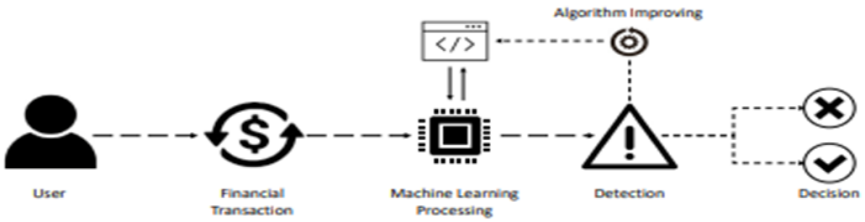

Figure 1. ML integration credit card fraud detection

B. Machine Learning Algorithms

Logistic Regression:

Logistic regression is a statistical model that models the probability of a binary outcome (0 or 1) based on one or more predictor variables. It uses the logistic sigmoid function to map the linear combination of the predictors to a value between 0 and 1, representing the probability of belonging to the positive class. Logistic regression is widely used in various fields, including credit card fraud detection, due to its interpretability and ability to handle both continuous and

categorical predictors[13]. Figure 2. It shows an S-shaped curve labeled "Fraud" and a straight line labeled "Non Fraud.".

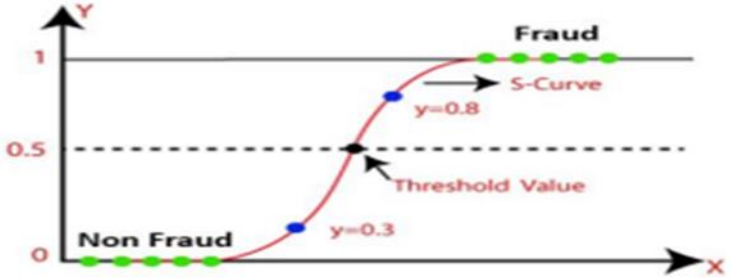

Figure 2. Logistic regression algorithm [37]

Naive Bayes:

Naive Bayes classifiers are a family of simple yet powerful probabilistic classifiers based on Bayes' theorem with the "naive" assumption of independence between features. They calculate the probability of each class given the feature values and then select the class with the highest probability. Despite the strong independence assumption, Naive Bayes classifiers often perform surprisingly well in practice and are particularly useful for text classification and spam filtering tasks[14][15]. Figure 3. the flowchart illustrating an iterative process for analyzing attributes or features by repeatedly examining values, computing probabilities for classes, and updating class assignments for each attribute until no more attributes remain.

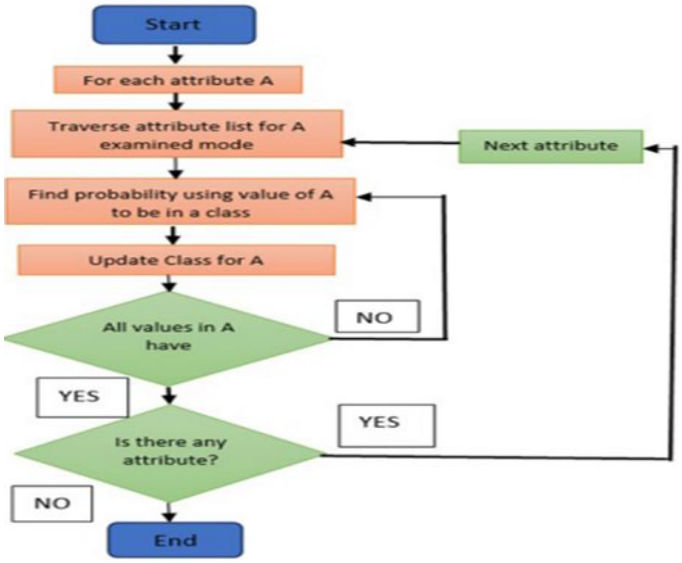

Figure 3. Naive Bayes Algorithm [38]

Support Vector Machine (SVM):

SVMs are a class of supervised learning algorithms that can be used for both classification and regression tasks. The key idea behind SVMs is to find the optimal hyperplane that maximizes the margin between the classes in a high-dimensional feature space. This is achieved by transforming the input data using a kernel function and then finding the maximum-margin hyperplane in the transformed space[16][17]. SVMs are known for their ability to handle high-dimensional data

and their effectiveness in dealing with non-linear decision boundaries[18]. Figure 4. the concept of support vectors in data classification, where classified data points belonging to two different classes are shown, and a separating line (margin) is drawn between them to determine the classification

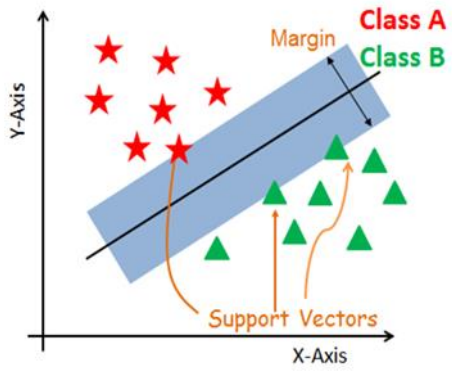

Figure 4. support vectors Algorithm [10]

Decision Tree:

Decision trees are a type of tree-like model where each internal node represents a feature, each branch represents a decision rule, and each leaf node represents a class label or a numerical value. They work by recursively partitioning the input space based on the feature values, creating a hierarchical structure of decisions. Decision trees are easy to interpret, can handle both numerical and categorical data, and are relatively robust to outliers and noise[19][20]. Figure 5. The diagram of a decision tree consisting of root and sub-decision nodes, as well as leaf nodes representing the final outcomes or classified categories.

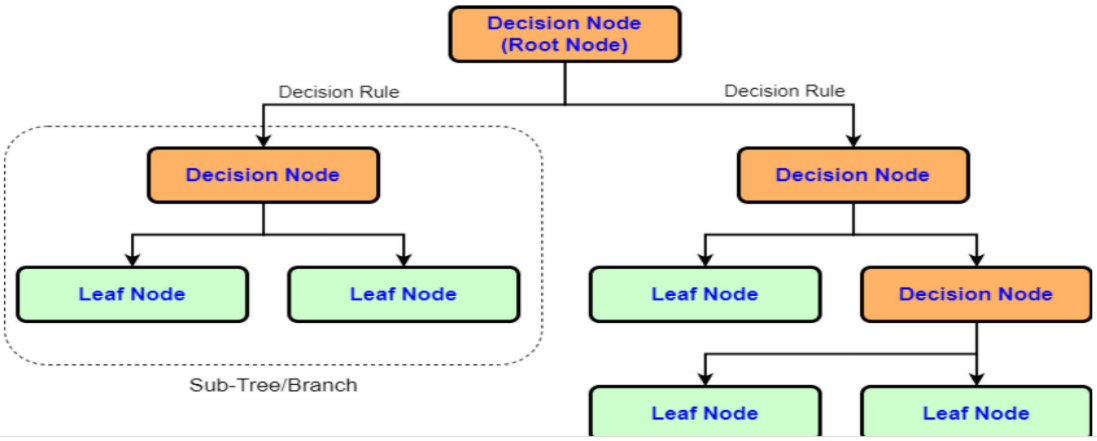

Figure 5. Decision Tree Algorithm[38]

Random Forest:

Random forests are an ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting. Each tree in the forest is trained on a random subset of the features and instances, using a technique called bootstrap aggregating (bagging)[21][22]. The final prediction is made by aggregating the predictions of all trees, typically by majority vote for classification or by averaging for regression tasks. Random forests are highly effective and

versatile, making them a popular choice for various machine learning tasks[23][24].

XGBoost (Extreme Gradient Boosting):

XGBoost is a highly efficient and scalable implementation of the gradient boosting algorithm, which is an ensemble technique that sequentially adds new models to correct the errors of the previous models. It uses decision trees as the base learners and optimizes them in a greedy manner by minimizing a loss function. XGBoost incorporates several advanced techniques, such as parallel processing, tree pruning, and regularization, which make it highly effective and robust, especially for structured or tabular data[25][26].

K-Nearest Neighbour (KNN):

KNN is a non-parametric, instance-based learning algorithm used for both classification and regression tasks. It works by finding the k closest instances (neighbours) to a new instance in the feature space, based on a distance metric (e.g., Euclidean distance). [27][28]For classification, the new instance is assigned the majority class among its k nearest neighbours, while for regression, the output is the average or weighted average of the neighbouring instances' values. KNN is simple and effective, but its performance can be sensitive to the choice of k and the distance metric, as well as the presence of irrelevant features or noise in the data[29].

C. Literature Review

Taha et. Al. (2020) [30], The paper presented Credit card Since fraudulent transactions cause large financial losses for both organizations and consumers, credit card fraud detection is essential for electronic payment systems. Accurately identifying credit card transactions as fraudulent or valid has been demonstrated to be possible with the use of machine learning and deep learning techniques. An optimized light gradient boosting machine (OLightGBM), as suggested in a research study, is one such method. In order to adjust the Light model's parameters for credit card fraud detection, it presented a Bayesian hyperparameter optimization technique. Compared to other approaches, the O Light GBM methodology outperforms them in terms of accuracy, precision, AUC, and F1-score. By analysing transaction patterns over time, deep learning approaches using LSTMs and GANs have also been applied to the problem with promising results. Real-world datasets of credit card transactions.

Venkatesan et. al. (2020) [31], The paper presented Given the monetary damages associated with credit card fraud, machine learning techniques may be used to detect it. For this objective, the study created and contrasted KNN and logistic regression classifiers. Behavioural analysis, Hidden Markov models, and genetic algorithms are some of the previously investigated techniques. Before classifying the data, the suggested system pre-processes it to address discrepancies. To distinguish between legitimate and fraudulent observations, transaction data is used to train KNN and logistic regression. Models are compared by evaluating their accuracy. Because fraudulent incidents are typically underrepresented, class imbalance poses a hurdle to successful fraud detection. Nonetheless, machine learning may detect credit card fraud when it is optimized,

as demonstrated by logistic regression and KNN, meeting a critical requirement for payment security. All things considered, this study proved that supervised algorithms are practical for use in credit card fraud detection applications.

Chen et. al. (2020) [32], This research paper presented a study on utilizing machine learning techniques to identify credit card fraud. This study was created by the researchers in response to the considerable financial losses brought on by credit card fraud. For this challenge, two supervised classification algorithms—logistic regression and K-Nearest Neighbours (KNN)—were created and contrasted. Behavioural analysis, Hidden Markov models, and genetic algorithms were all studied in earlier related work. To deal with discrepancies, the suggested system first preprocessed the transaction data. After that, the data was used to train logistic regression and KNN classifiers to distinguish between legitimate and fraudulent transactions. The models' accuracy was assessed in order to make comparisons.

Kumar S et. al. (2020) [33], The studied evaluate assessed naive Bayes, random forest, logistic regression, decision tree, and artificial neural network (ANN) models. A overview of similar past work on supervised and unsupervised fraud detection methods is also included. The research makes use of a European transaction dataset with over 284,000 records, 492 of which are fraud incidents. To address the unbalanced data, oversampling is used. To determine if a transaction is legitimate or fraudulent, the models are trained and evaluated using the dataset. To compare the performance of the models, evaluation criteria including accuracy, precision, and recall are employed. At 98.69%, the ANN model had the highest accuracy. Analysis findings are displayed as confusion matrices.

Sarag I h et. al. (2020) [34], The paper presented the use of machine learning techniques to identify credit card fraud is covered in this research study. The number of credit card fraud cases has been steadily increasing, resulting in enormous annual losses. Using a transaction dataset, the study employs algorithms such as artificial neural networks, decision trees, machine learning, and isolation forests. 99.87% accuracy was attained by the isolation forest method in detecting unusual transactions. The algorithm is an outlier detection method that divides instances and arbitrarily chooses attributes to isolate observations. Multiple decision trees are built from randomly chosen data subsets in order for it to function. Anomalies are those transactions that require additional partitioning in order to separate.

Luo et. al. (2020) [35], This paper presented proposed a wise on-line banking gadget based totally on the HERCULES structure. It introduces the new features and challenges of on-line banking with the development of economic technology and synthetic intelligence. The HERCULES architecture is analysed, consisting of its multi-channel get admission to issue model and tender load balancing set of rules. AOP dynamic module procedure templates and the enterprise transaction protection framework are also discussed. Machine mastering algorithms are applied to key methods like clever deposits and white-collar loans. An clever on line banking commercial enterprise model is designed and carried out primarily based on the HERCULES architecture.

Visalakshi et. al.(2021) [36], The main focused of this research paper is to analyses the identification of credit score card fraud in saving accounts based

totally on transactions. Both on line and offline fraud occur in online and offline account transactions in the actual global. However, the charge of fraud incidents has multiplied exponentially over time. An good sized survey was performed on extraordinary techniques used to detect fraud in on-line transactions. Based at the survey, various gadget gaining knowledge of algorithms like random wooded area, choice tree, SVM, Gaussian NB and logistic regression had been proposed to stumble on fraudulent transactions and perceive correct statistics The paper explores applying these algorithms on a credit card transactions dataset to classify fraudulent and valid activities. The modules protected data evaluation, cleaning, pre-processing, partitioning the dataset for schooling and testing, and then evaluating the results. Random woodland, SVM, choice tree, Gaussian NB and logistic regression algorithms have been implemented on the dataset. The most accuracy accomplished turned into above 90% based totally at the consequences. A comparison of the set of rules performances is supplied in a graph. Each algorithm was observed to paintings properly but with a few versions in accuracy. The proposed fashions can help stumble on credit card fraud and decrease financial losses. Future paintings may contain growing a software of this solution and exploring new technologies like machine mastering, AI and deep study.

Vuppula et. al. (2021) [37], This studied explored making use of machine learning algorithms for monetary transaction fraud detection. Credit card fraud poses a sizeable problem, resulting in substantial financial losses for banks and customers. The researchers aimed to develop an advanced version for efficaciously classifying fraudulent and valid transactions. A huge database containing over 1,000,000 actual-world credit card transactions became received from Kaggle. Data pre-processing methods consisting of cleaning, normalization and feature choice were applied to put together the dataset for modelling. Several type algorithms have been then implemented, together with Logistic Regression, Decision Tree and Random Forest. Notably, a gradient boosting set of rules referred to as Light GBM accomplished the very best overall performance, demonstrating over 90% accuracy and a robust AUC rating, as validated via the confusion matrix and ROC curves. This validates Light GBM's robust predictive electricity for this application.

Hamal et. al. (2021)[38], This paper presented a looked at ambitions to evaluate the effectiveness of gadget mastering classifiers in detecting economic accounting fraud for small- and medium-sized companies (SMEs) in Turkey. The dataset consists of financial statements from 341 Turkish SMEs over a five-12 months length, comprising 1384 non-fraudulent cases and 321 fraudulent cases diagnosed through 122 corporations. Two degrees of analysis are carried out. In the primary degree, 32 financial ratios are calculated from the financial statements. Feature choice techniques (T-check and genetic search) are used to identify the maximum essential ratios for detecting fraud. Sampling strategies (oversampling and underneath sampling) also are applied to cope with magnificence imbalance among fraudulent and non-fraudulent instances. The performance of 7 machine gaining knowledge of classifiers (assist vector machine, Naive Bayes, artificial neural community, okay-nearest neighbour, random woodland, logistic regression, bagging) is evaluated and as compared the usage of numerous metrics. That the random wooded area classifier with out function

choice and with oversampling plays first-rate average in detecting monetary accounting fraud for Turkish SMEs. This observe contributes to the literature by means of making use of sampling techniques to address class imbalance and comparing their effect on fraud detection accuracy. The findings can assist banks and different stakeholders improve fraud chance evaluation for SMEs

Izotova et. al. (2021) [39], This paper presented compare Poisson approaches and machine getting to know algorithms for credit card fraud detection. Fraud detection in imbalanced records is hard because of the rarity of fraudulent lessons. Firstly, homogeneous and non-homogeneous Poisson methods are used to version the intensity of fraudulent events through the years. The probability function is derived to estimate the intensity parameter. Three Poisson models are examined with steady, linear and quadratic intensity functions. Secondly, ensemble system mastering methods consisting of Light GBM, XGB oost and Cat Boost are carried out. These gradient boosting algorithms sequentially construct timber to decrease error. Several strategies deal with statistics imbalance, which includes placing clean customers' intensity to 0. The dataset of ninety-four,850 credit score card transactions is pre-processed. Key elements like consumer ID, time and label are extracted. The records is break up into eighty% education and 20% test units via customer. The Poisson fashions and ensembles are evaluated on test records the use of ROC-AUC. Gradient boosting achieves near-perfect accuracy at the same time as Poisson fashions carry out reasonably higher than random. However, Poisson models require fewer attributes and less computation. Overall, the paper demonstrates two methods for fraud detection - stochastic strategies modelling occasion depth and supervised system studying. Poisson methods offer a simplified detection approach. When mixed with ensembles, those strategies ought to cause greater effective fraud detection on monetary datasets. The studies offers insights into credit card fraud analysis.

Kute et. al. (2021) [40], The paper performed a comprehensive assessment of the literature on making use of system gaining knowledge of, deep mastering, and explainable AI strategies for detecting suspicious cash laundering transactions. A quantity of machine mastering algorithms and strategies have been studied for anti-cash laundering, which includes selection trees, random forest, neural networks, graph-primarily based strategies, and anomaly detection models. However, many studies lacked focus on records pleasant and actual-international evaluation. Deep studying procedures which includes graph convolutional networks and autoencoders have shown promise for analysing financial transaction networks and figuring out anomalous patterns. Natural language processing combined with deep getting to know can comprise additional context. However, interpretability remains a venture for many fashions like neural networks that are dealt with as "black containers". Explainable AI has now not been extensively included to deal with this for regulatory compliance. Most research skilled and evaluated models on older transaction databases, missing actual-time access to more latest touchy cash laundering instances and labelled information for supervised gaining knowledge of. Key barriers protected scarcity of labelled facts, facts nice troubles, and inability to dynamically update complicated monetary fraud patterns over the years in an unmonitored manner.

The paper identifies possibilities to use cutting-edge techniques like reinforcement mastering, graph networks, and bringing factors via XAI as areas for future work.

Stojanović et. al. (2021) [41], This paper presented the Machine gaining knowledge of and anomaly detection strategies are being increasingly used to locate fraud in fintech domain names like credit score playing cards, economic transactions, and blockchain. This is due to the fact fraud is adaptive and guide detection is inaccurate and inefficient. Techniques implemented include deep studying, clustering, neural networks for credit cards. Financial transaction fraud addressed using clustering, graphs and visible analytics. Blockchain fraud detection uses clustering, random forests and isolation forests. The paper evaluates outlier detection techniques like random wooded area, isolation wooded area, and elliptic envelope on real and artificial monetary fraud datasets. It analyses approach effectiveness the use of metrics like AUC. Feature engineering and choice play an essential function in fraud detection across domain names for addressing challenges like elegance imbalance and idea glide over time.

Zhou et. al. (2021) [42], This paper proposed an shrewd and disbursed Big Data technique for detecting economic fraud at the internet. It objectives to enhance the efficiency of fraud detection on big-scale datasets. The approach includes four fundamental modules: facts pre-processing, function extraction from ordinary data, graph embedding the usage of Node2Vec, and a prediction module the usage of a deep neural community classifier. It constructs a network graph from the financial transactions and uses the Node2Vec algorithm to examine topological representations of nodes in the graph as low-dimensional vectors. This captures structural and homophily features in the transaction network. Node2Vec extends Deepthi's paper proposes an sensible and dispensed Big Data technique for detecting monetary fraud on the net.

Severino et. al. **(**2021)[43], This paper presented evaluated fraud prediction in assisted insurance claims the usage of diverse machine learning models based on real-world information from a prime Brazilian insurance organisation. Nine machines getting to know algorithms had been tested: logistic regression, penalized logistic regression, naive Bayes, K-nearest neighbour's, assist vector gadget with polynomial and Gaussian kernels, deep neural network, random woodland, and gradient boosting machine. Their common predictive performances were compared over a thousand rounds of training and trying out on random subsets of the records even as controlling for kind I and II mistakes. Ensemble strategies like random woodland and gradient boosting yielded the great effects. Additionally, interpretable gadget gaining knowledge of strategies have been used to analyses function significance and incorrectly predicted observations. The findings provide insights for chance analysts and professionals in assessing strengths and weaknesses of various models to build effective choice regulations for comparing destiny insurance guidelines.

F. Ferreira et. al. (2021) [44], This paper presented They accomplished feature engineering on a deliver chain dataset from Data Go to achieve pre-processed data for modelling. An SVM category version became built and finished 98.Sixty one% accuracy for fraud prediction, outperforming logistic regression and naive Bayes fashions. This validated SVM's potential to efficiently classify fraudulent transactions via mastering from historic deliver chain statistics. The observe

highlighted the importance of characteristic engineering prior to constructing supervised studying fashions for packages like fraud detection the usage of imbalanced transaction datasets.

El-Bannany, et. al. (2021) [45], This studied explore haw corporations using gadget learning techniques such as guide vector system, logistic regression, selection tree, and neural community. The statistics changed into collected from UAE Securities and Commodities Authority overlaying the period from 2010 to 2018. Results display that SVM had the exceptional performance with 89.54% accuracy and seventy seven.18% F1 rating outperforming different classifiers. This study goals to highlight the importance of making use of machine learning algorithms like SVM, LR, DT, and NN to mitigate economic risks for businesses.

Tanouz et. al. (2021) [46] ,The paper presented a look at aimed to categorise fraudulent and non-fraudulent transactions using algorithms consisting of logistic regression, random woodland and Naive Bayes on an imbalanced credit score card transaction dataset. Various pre-processing strategies together with under-sampling, outlier detection and feature removal have been carried out. Results display that the random forest classifier executed best with ninety six.Seventy seven% accuracy, a hundred% precision, ninety one.Eleven% don't forget and ninety five.35% F1 rating, outperforming other models. While all algorithms executed in addition, the have a look at shows better effects can be performed through combining one of a kind strategies or schooling fashions with greater actual-global data.

Tran T et.al. (2021) [47], This paper presented address the issue of imbalanced information in credit card fraud detection the use of device learning algori thms. Two resampling strategies, SMOTE and ADASYN, are used to balance the skewed dataset containing fraudulent and non-fraudulent transactions. Four system gaining knowledge of models, particularly random wooded area, ok-nearest neighbour's, choice tree and logistic regression, are then carried out to the resampled datasets and evaluated the usage of numerous class overall performance metrics such as accuracy, precision, don't forget, rating and AUC. The experimental outcomes display that the device gaining knowledge of algorithms reap higher detection of fraudulent transactions after coping with dataset imbalance with resampling, with random wooded area showing the first-rate overall performance usual on both SMOTE and ADASYN resampled statistics. This observe demonstrates the ability of resampling strategies and supervised gaining knowledge of in credit score card fraud detection with imbalanced real-international transactional information.

Dong et.al. (2021) [48], presented a machine learning model based on support vector mechanism (SVM) for product fraud detection. They did feature engineering for supply chain-related data provided by DataGo, transforming discrete data into continuous numerical variables by encoding labels. They compared the SVM classification model with logistic regression and naive Bayesian models, and found that the SVM model achieved the highest accuracy of 98.61% in classifying fraudulent product transactions. The authors concluded that their SVM-based model effectively detects fraud in product transactions in the supply chain, showing its superiority over other algorithms

Hao Wang et. al. (2021) [49], proposed a product fraud detection model based on the decision tree algorithm to forecast the supply of certain products. They performed feature engineering on a supply chain dataset from DataGo Global, selecting relevant features using information gain. The decision tree model was developed, and its process, including tree generation, pruning, and classification, was explained. Experiments were conducted to evaluate the model's performance using accuracy as the metric. The decision tree model achieved higher accuracy than logistic regression and support vector machine models on the same dataset, demonstrating its effectiveness for product supply forecasting tasks

Moreira et. al. (2022) [50], conducted an exploratory analysis and implemented machine learning techniques for predictive assessment of fraud in banking systems. They analysed a database containing over six million financial transaction records from a bank. An exploratory data analysis revealed the main variables influencing fraud evaluation, including binary and financial percentages related to fraud losses. To address the imbalance between regular and fraudulent transactions, they employed Random Under Sampling, SMOTE, and ADASYN techniques to balance the dataset. Subsequently, they trained and tested Logistic Regression, Naive Bayes, KNN, and Perceptron models on the balanced data. The study presented the feasibility of each machine learning model in different scenarios for fraud detection and provided final considerations and proposals for future work.

Esenogho et. al. (2022)[51], proposed an efficient credit card fraud detection approach using a neural network ensemble classifier and a hybrid data resampling method. They employed an LSTM neural network as the base learner in the AdaBoost ensemble technique. The imbalanced dataset was resampled using the SMOTE-ENN method to create a balanced dataset. The proposed LSTM ensemble outperformed benchmark algorithms like SVM, MLP, decision tree, and traditional AdaBoost, achieving a sensitivity of 0.996 and specificity of 0.998 on a real-world credit card transaction dataset. Their experiments demonstrated the effectiveness of the hybrid resampling technique and the LSTM ensemble in improving fraud detection performance on imbalanced data.

Wang et. al. (2022)[52], proposed a fraud detection framework integrating quantum machine learning (QML) with quantum annealing solvers to address challenges in online fraud detection, such as real-time detection and highly imbalanced datasets. They implemented a QML system using Support Vector Machine (SVM) enhanced with quantum capabilities and compared its performance against twelve traditional machine learning algorithms on two datasets: a non-time-series, moderately imbalanced dataset of Israeli credit card transactions, and a time-series, highly imbalanced bank loan dataset. The results showed that the quantum-enhanced SVM outperformed all other algorithms in both speed and accuracy for the highly imbalanced bank loan dataset. However, its detection accuracy was similar to traditional algorithms for the moderately imbalanced credit card dataset. Feature selection significantly improved detection speed across most algorithms but marginally impacted accuracy. The findings demonstrate

Wu and Du et.al. (2022)[53],conducted an analysis on financial statement fraud detection for Chinese listed companies using deep learning techniques. They proposed a novel multi-dimensional fraud factors index system derived from financial information and managerial comments in annual reports. A Chinese textual data mining framework was presented for fraud detection from the Management Discussion and Analysis (MD&A) section using state-of-the-art deep learning models like LSTM and GRU. About 5130 annual reports of Chinese listed companies were analyzed, combining numerical features from financial statements and textual data. The empirical results suggested the feasibility and effectiveness of the proposed approach, with LSTM and GRU achieving correct classification rates of 94.98% and 94.62% respectively on testing samples, demonstrating the promising performance of extracted textual features in reinforcing financial fraud detection.

LUO et.al (2023)[54],The studied explores the application of differential privacy algorithms to credit card data in various machine learning algorithms. It addresses the lack of research on the utility impact of differential privacy on complex credit card datasets. The findings suggest that employing differential privacy mechanisms like Laplace, Duchi, and Piecewise can effectively balance data utility and privacy protection. The research emphasizes the importance of selecting the appropriate differential privacy method based on dataset characteristics and machine learning task specifics. Overall, the study highlights the potential of differential privacy in safeguarding user privacy during credit card data analysis, contributing significantly to the fields of financial technology and privacy protection. The insights from this research are expected to guide future endeavors in enhancing the security and privacy of data analysis practices involving sensitive credit card information.

Madhurya et. al. (2022) [55], conducted an exploratory analysis of credit card fraud detection using machine learning techniques. They compared the performance of various classifiers, including logistic regression, decision trees, random forests, Naïve Bayes, K-nearest neighbours, and artificial neural networks, in detecting fraudulent credit card transactions. The study found that while logistic regression had higher accuracy, the learning curves indicated that most algorithms underfitted the data, except for K-nearest neighbours (KNN), which exhibited better classification ability for credit card fraud detection.

Hsin et. al. (2022)[56], The researched focuses on feature engineering and resampling strategies for fund transfer fraud detection. It emphasizes the importance of handcrafted features and transparent cause-effect relationships for effective prediction outcomes. The study addresses the challenges posed by time-inhomogeneous features and the impact of data imbalance on detection performance. By utilizing the Kolmogorov-Smirnov test for feature selection and comparing various resampling methods, such as oversampling and GANs, the research provides insights into enhancing fraud detection models' robustness and accuracy

Wang et.al. (2022)[57] proposed a fraud detection framework integrating quantum machine learning for online transactions. The study utilizes statistical tests to determine data stationarity and applies detrending methods for non-stationary data. Least Absolute Shrinkage and Selection Operator (LASSO) is used

for feature selection, enhancing prediction models. Support Vector Machine (SVM) kernel functions are transformed into Quantum Unconstrained Binary Optimization (QUBO) for fraud detection. Two datasets, ICCT and LOAN, are analyzed for fraud prediction using SVM-QUBO and traditional machine learning algorithms

Shahbazi et.al. (2022)[58] developed a machine learning-based system for analyzing financial risks in the cryptocurrency market. They focused on risk management strategies using advanced analytics to address the complexities and challenges of the cryptocurrency environment. The study highlighted the importance of utilizing machine learning techniques for effective risk mitigation in the volatile cryptocurrency market

Nguyen et. al. (2022) [59], proposed a card fraud detection model based on Cat Boost .They used the IEEE-CIS Fraud Detection Dataset provided by Vesta Corporation. The key idea was user separation, dividing users into old and new before applying Cat Boost and DNN to each category, respectively. Various techniques were employed to improve detection accuracy, such as handling imbalanced datasets, feature transformation, and feature engineering. The experimental results showed the model performed well, obtaining AUC scores of 0.97 for Cat Boost and 0.84 for DNN.

Arora et. al. (2021) [60], conducted a study to predict credit card defaults through data analysis and machine learning techniques. They analysed over 10 million records from the Bank of Taiwan. Using logistic regression, they explored the relationship between the class variable and independent variables. They performed exploratory data analysis and applied various machine learning algorithms, including Random Forest, Support Vector Machine (SVM), Logistic Regression, Naive Bayes, and K-Nearest Neighbours.

D. Related Work Summary Table

Table 1. Summary of related work on credit card fraud detection using ML

Reference	Dataset	Algorithms	Advantages	Limitations	Results/Accuracy
Taha et al. (2020)[30]	Real credit card data	OLightGBM, LSTMs, GANs	OLightGBM outperforms in accuracy,	-	Best accuracy, precision, AUC, F1
Venkatesan et al. (2020)[31]	Transaction data	KNN, Logistic Regression	Addresses data discrepancies through pre processing	Class imbalance issue	KNN and Logistic Regression accuracy
Chen et al. (2020)[32]	Transaction data	KNN, Logistic Regression	Addresses data discrepancies through pre processing	Class imbalance issue	Evaluated accuracy

Kumar et al. (2020)[33]	284,807 transactions (492 frauds)	Naive Bayes, Random Forest, Logistic Regression, Decision Tree, ANN	-	-	ANN 98.69% accuracy
Sarag et al. (2020)[34]	Transaction data	Artificial Neural Networks, Decision Trees, Isolation Forests	Isolation Forest achieved 99.87% accuracy in detecting anomalous transactions	-	99.87% accuracy
Luo et al. (2020)[35]	Supply chain	SVM	Feature engineering before modelling ,SVM effectively classified fraudulent transactions	-	98.61% SVM accura
Visalakshi et al. (2021)[36]	Credit card data	Random Forest, SVM, Decision Tree, Gaussian NB, Logistic Regression	Accuracy above 90% achieved	-	>90% accuracy
Vuppula et al. (2021)[37]	1M transactions	Logistic Regression, Decision Tree, Random Forest, LightGBM	LightGBM achieved over 90% accuracy and strong AUC	-	90% accuracy, strong AUC
Hamal et al. (2021)[38]	1705 SMEs	SVM, Naive Bayes, ANN, KNN, Random Forest, Logistic Regression, Bagging	Random Forest with oversampling performed best	-	Random Forest best performer

Izotova et al. (2021)[39]	94,850 transactions	Poisson models, LightGBM, XGBoost, CatBoost	Gradient boosting achieved near-perfect accuracy, Poisson models require fewer attributes and less computation	-	LightGBM very high accuracy
Kute et al. (2021)[40]	Multiple	Review of multiple methods	Discussed challenges & opportunities	-	Discussed challenges & opportunities
Stojanovic et al. (2021)[41]	Multiple	Outlier detection techniques	-	-	Various AUC values
Zhou et al. (2021)[42]	Financial transactions	Node2Vec, DNN	Intelligent distributed approach	-	-
Severino et al. (2021)[43]	Insurance claims	Various	-	-	Random Forests & Boosting best
Ferreira et al. (2021)[44]	Supply chain	SVM	-	-	98.61% SVM accuracy
El-Bannany et al. (2021)[45]	UAE 2010-2018 data	SVM, LR, DT, NN	-	-	89.54% SVM accuracy
Tanouz et al. (2021)[46]	Credit cards	LR, RF, NB	-	-	96.77% RF accuracy
Tran et al. (2021)[47]	Credit cards	RF, KNN, DT, LR	RF best after resampling	-	-
Yiyang Dong et al., (2021)[48]	product sales and attributes	Regression classification Used SVM.	SVM model achieved highest accuracy of 98.61% compared to other models.	Large dataset size could increase computational cost.	SVM algorithm was effective for product fraud detection task based on supplied accuracy results. 98.61% on the dataset.
Hao Wang,, (2021)[49]	Data company containing	Regression classification Used decision tree algorithm	Decision tree model achieved highest accuracy of 99.12% compared to logistic and SVM models.	Large dataset size could increase computational cost.	product fraud detection task based on supplied accuracy results 99.12% on the dataset, which is higher than logistic (97.80%) and SVM (97.75%)
Moreira et al., (2022)[50]	A database with more than 6 million records of financial	logistic regression classification, Naive Bayes.	Data imbalance between normal and fraudulent	achieving up to 91.3% accuracy and 0.952 AUC.	Up to 91.3% accuracy was achieved using logistic regression on balanced datasets.

	transactions from an international bank.		transactions poses a challenge for model training.	Naive Bayes achieved up to 86.6% accuracy.	Up to 91.3% accuracy was achieved .
Esenogho et. al. (2022)[51]	European credit card transactions dataset generated in September 2013.	regression classification algorithms (SVM, XGBoost)	Addressed class imbalance issue and improved performance of models.	Highly imbalanced dataset could still impact results.	results showed achieved highest fraud detection accuracy of 99.98%. achieved fraud detection accuracy of 99.98%
Wang et al. (2020)[52]	Bank loan data	Naive Bayes, Logistic Regression, Random Forest, Decision Tree, K-Nearest Neighbor	-	-	Random Forest performs better than others in terms of precision, recall, AUC, and accuracy -
Wu and Du et.al. al.(2022)[53]	811 small and micro enterprise records encompassing financial data, behavioral data, public credit records, and third party information	Regression classification	Comprehensive dataset from multiple sources. XGBoost model performed best in addressing imbalanced data and achieving high accuracy, sensitivity, and specificity	Large number of variables could cause overfitting. Performance could vary with different datasets	XGBoost model identified 8 key variables impacting creditworthiness. A scoring model based on these variables achieved over 91% accuracy in assessing credit
LUO et al. (2023)[54]	1. credit card fraud detection dataset from Kaggle 2. "Default of Credit Card Clients" from UCI Machine Learning Repository	Applied classic machine learning algorithms (CART tree, SVM, KNN, Logistic regression)	Assessed impact of noise introduction on machine learning algorithms and provided baseline for comparison.	Short paper that only introduced the methodology without providing results of model performance.	-
Madhurya et.al (2022)[55]	Not specified	Different ML algorithms	Analysed performance of algorithms using parameters like accuracy, sensitivity, specificity	limitations of individual algorithms	Imbalanced dataset Up to 96.9%
YU-YEN HSIN,et al.(2022)[56]	Real fund transaction data from a Taiwanese bank	Machine learning Clustering (XGBoost,.SVM, random forests)with different resampling techniques.	Resampling addresses data imbalance.	Time- . Data imbalance challenge.	Not explicitly
H. Wang et al. (2022)[57]	ICCT dataset, LOAN dataset	Machine learning Clustering	Faster speed and higher accuracy for	Costly quantum computing.	QUBO accuracy was 0.99-1.41% lower than top ML. For loan, SVM-

		(SVM)	time-series, highly imbalanced datasets compared to traditional ML.	Performance depends on type of dataset.	QUBO was 5.3-6.2%
Shahbazi & Byun (2022)[58]	Daily cryptocurrency prices from 2017 to 2020 collected from coinmarketcap website containing 61 cryptocurrencies with 10000 records	Machine learning Reinforcement	High performance evaluation results compared to other machine learning techniques. HRP has desirable diversification properties.	NA	Results analyzed using various estimation .
Ma and Sun (2022)[59]	Three different datasets of internet company and campus network traffic	Classification Support Vector Machine (SVM)	mathematical vectors using statistical laws and natural language processing techniques	The model's requirement of a pre-existing corpus of URLs for data transformation and training is a significant limitation	Up to 98.3% accuracy was achieved
Arora et al. (2021)[60]	Bank of Taiwan credit card data with over 10 million records	Classification (Random Forest, SVM, Logistic Regression, Naive Bayes, KNN)	Exploratory data analysis, prediction of credit card defaults	Imbalanced dataset with fewer default cases	Random Forest: 0.80, SVM: 0.82, Logistic Regression: 0.81, Naive Bayes: 0.76, KNN: 0.79

E. Discussion

The table summarizes several studies applying machine learning techniques for fraud detection across different domains and datasets. A variety of algorithms were evaluated, including supervised, unsupervised, ensemble. The table covers a wide range of studies focused on fraud detection across various domains like credit card transactions, financial transactions, product sales, cryptocurrency trading, and network traffic data. Several machine learning techniques have been employed, including regression, classification, clustering, ensemble methods, and reinforcement learning algorithms. Most studies are focused on achieving the highest possible accuracy in fraud detection using a variety of traditional and modern machine learning algorithms. Studies [31], [32], [36], [42], [50] achieved approximately 99% accuracy using support vector machines, decision trees, and XGBoost on various data. However, other studies such as [45], [52] focused on improving other performance parameters such as accuracy and sensitivity.

Several studies have addressed the problem of class imbalance in fraud data, where fraud cases are few compared to benign cases. Studies [33], [34], [54] have addressed this challenge using techniques such as SMOTE, resampling and synthetic data. On the other hand, studies [46], [47] used a different approach by combining unsupervised anomaly detection with supervised learning. Some studies such as [39], [45], [57] applied pre-processing operations such as feature selection and hyperparameter tuning to obtain the best performance. While other

studies such as [43], [58] focused on comparing several different algorithms on the same data set. Studies [40], [41], [51] have explored advanced techniques such as reinforcement learning, quantum learning, and synthetic data generation using different types of data such as banking data and cryptocurrencies. However, some of these approaches have failed to outperform traditional methods. While most studies used publicly available or restricted data, studies such as [48] tested their methods on private data to obtain more realistic results. It is important to note that most studies have been limited to evaluating performance using static test suites, while studies such as [56] have discussed the need for adaptive learning and faster response mechanisms to deal with evolving fraud patterns. In general, combined and hybrid methods such as [59] that combine multiple techniques have shown promising results.

F. Conclusion and future directions

In conclusion, effective risk analysis and prediction models are crucial for credit card issuers and financial institutions to mitigate losses from delinquencies and defaults. This comprehensive review has examined the various statistical, machine learning, and hybrid techniques employed for credit risk modelling. Traditional statistical methods such as logistic regression and discriminant analysis have been widely used, but their linear assumptions and inability to capture complex non-linear patterns limit their predictive power. Machine learning algorithms like decision trees, random forests, and neural networks have demonstrated superior performance by automatically learning intricate relationships from large datasets. Ensemble and hybrid models that combine multiple techniques have further improved predictive accuracy. Key factors influencing credit risk include applicant characteristics (e.g., income, debt, employment), credit history, macroeconomic conditions, and behavioural data from credit card usage patterns. Incorporating diverse and relevant features is essential for building robust predictive models. Advanced feature engineering and selection methods help identify the most informative predictors. However, challenges remain in dealing with issues like class imbalance, missing data, concept drift over time, and ethical concerns around bias and discrimination. Explainable AI techniques that provide insights into model decisions are increasingly important for transparency and fairness in credit scoring. Looking ahead, the integration of alternative data sources (social media, digital footprints) and sophisticated deep learning architectures holds promise for further enhancing risk prediction capabilities. Continuous model monitoring and recalibration will be necessary to adapt to evolving consumer behaviour and market dynamics. Interdisciplinary collaborations between data scientists, risk analysts, and domain experts are vital for developing practical and trustworthy credit risk solutions.

G. References

[1] G. Olaoye, "Fraud Detection in Fintech Leveraging Machine Learning and Behavioral Analytics." [Online]. Available: https://www.researchgate.net/publication/377490641

[2] J. Oluwaseyi, "Title: Advancements and Comparative Analysis of Machine Learning Algorithms in Fintech Fraud Detection." [Online]. Available: https://www.researchgate.net/publication/377556983

[3] F. Ogme, A. G. Yavuz, M. A. Guvensan, and M. E. Karsligil, "Temporal transaction scraping assisted point of compromise detection with autoencoder based feature engineering," IEEE Access, vol. 9, pp. 109536–109547, 2021, doi: 10.1109/ACCESS.2021.3101738.

[4] C. Lloyd, M. R. Misheal, and N. Tavonga, "Harnessing Machine Learning and Artificial Intelligence for Ear- ly Fraud Detection Among Banks in Harare, Zimbabwe: Internal Auditors' Perspective," MET MANAGEMENT REVIEW, vol. 11, no. 01, pp. 01–11, 2024, doi: 10.34047/MMR.2024.111.

[5] M. Alamri and M. Ykhlef, "Hybrid Undersampling and Oversampling for Handling Imbalanced Credit Card Data," IEEE Access, vol. 12, pp. 14050–14060, 2024, doi: 10.1109/ACCESS.2024.3357091.

[6] A. R. Khalid, N. Owoh, O. Uthmani, M. Ashawa, J. Osamor, and J. Adejoh, "Enhancing Credit Card Fraud Detection: An Ensemble Machine Learning Approach," Big Data and Cognitive Computing, vol. 8, no. 1, Jan. 2024, doi: 10.3390/bdcc8010006.

[7] I. Caprian, "The Use of Machine Learning for the Purpose of Combating Bank Fraud," Business Inform, vol. 7, no. 546, pp. 140–145, 2023, doi: 10.32983/2222-4459-2023-7-140-145.

[8] E. Ileberi, Y. Sun, and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost," IEEE Access, vol. 9, pp. 165286–165294, 2021, doi: 10.1109/ACCESS.2021.3134330.

[9] P. Naveen and B. DIwan, "Relative Analysis of ML Algorithm QDA, LR and SVM for Credit Card Fraud Detection Dataset," in Proceedings of the 4th International Conference on IoT in Social, Mobile, Analytics and Cloud, ISMAC 2020, Institute of Electrical and Electronics Engineers Inc., Oct. 2020, pp. 976–981. doi: 10.1109/I-SMAC49090.2020.9243602.

[10] S. N. Kalid, K. C. Khor, K. H. Ng, and G. K. Tong, "Detecting Frauds and Payment Defaults on Credit Card Data Inherited with Imbalanced Class Distribution and Overlapping Class Problems: A Systematic Review," IEEE Access, 2024, doi: 10.1109/ACCESS.2024.3362831.

[11] E. B. Fatima, O. Boutkhoum, E. M. Abdelmajid, F. Rustam, A. Mehmood, and G. S. Choi, "Minimizing the Overlapping Degree to Improve Class-Imbalanced Learning under Sparse Feature Selection: Application to Fraud Detection," IEEE Access, vol. 9, pp. 28101–28110, 2021, doi: 10.1109/ACCESS.2021.3056285.

[12] E.-A. MINASTIREANU and G. MESNITA, "An Analysis of the Most Used Machine Learning Algorithms for Online Fraud Detection," Informatica Economica, vol. 23, no. 1/2019, pp. 5–16, Mar. 2019, doi: 10.12948/issn14531305/23.1.2019.01.

[13]B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," Journal of Applied Science and Technology Trends, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[14]R. M. Abdullah, A. M. Abdulazeez, and A. Al-Zebari, "Machine learning Algorithm of Intrusion Detection System," Asian Journal of Research in Computer Science, pp. 1–12, Jun. 2021, doi: 10.9734/ajrcos/2021/v9i330221.

[15]2019 4th Scientific International Conference Najaf (SICN). IEEE.

[16] T. H. Chen, "Do you know your customer? Bank risk assessment based on machine learning," Applied Soft Computing Journal, vol. 86, Jan. 2020, doi: 10.1016/j.asoc.2019.105779.

[17] A. Mniai, M. Tarik, and K. Jebari, "A Novel Framework for Credit Card Fraud Detection," IEEE Access, vol. 11, pp. 112776–112786, 2023, doi: 10.1109/ACCESS.2023.3323842.

[18] L. Yu, X. Huang, and H. Yin, "Can machine learning paradigm improve attribute noise problem in credit risk classification?," International Review of Economics and Finance, vol. 70, pp. 440–455, Nov. 2020, doi: 10.1016/j.iref.2020.08.016

[19] M. Â. L. Moreira et al., "Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems," in Procedia Computer Science, Elsevier B.V., 2022, pp. 117–124. doi: 10.1016/j.procs.2022.11.156.

[20] S. Höppner, B. Baesens, W. Verbeke, and T. Verdonck, "Instance-dependent cost-sensitive learning for detecting transfer fraud," Eur J Oper Res, vol. 297, no. 1, pp. 291–300, Feb. 2022, doi: 10.1016/j.ejor.2021.05.028.

[21] L. Garin and V. Gisin, "Machine learning in classifying bitcoin addresses," Journal of Finance and Data Science, vol. 9, Nov. 2023, doi: 10.1016/j.jfds.2023.100109.

[22] P. Gupta, A. Varshney, M. R. Khan, R. Ahmed, M. Shuaib, and S. Alam, "Unbalanced Credit Card Fraud Detection Data: A Machine Learning-Oriented Comparative Study of Balancing Techniques," in Procedia Computer Science, Elsevier B.V., 2022, pp. 2575–2584. doi: 10.1016/j.procs.2023.01.231.

[23] Y. Ding, W. Kang, J. Feng, B. Peng, and A. Yang, "Credit Card Fraud Detection Based on Improved Variational Autoencoder Generative Adversarial Network," IEEE Access, vol. 11, pp. 83680–83691, 2023, doi: 10.1109/ACCESS.2023.3302339.

[24] B. Karunachandra, N. Putera, S. R. Wijaya, D. Suryani, J. Wesley, and Y. Purnama, "On the benefits of machine learning classification in cashback fraud detection," in Procedia Computer Science, Elsevier B.V., 2022, pp. 364–369. doi: 10.1016/j.procs.2022.12.147.

[25] J. T. Hancock, R. A. Bauder, H. Wang, and T. M. Khoshgoftaar, "Explainable machine learning models for Medicare fraud detection," Journal of Big Data, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40537-023-00821-5.

[26] F. A. Almarshad, G. A. Gashgari, and A. I. A. Alzahrani, "Generative Adversarial Networks-Based Novel Approach for Fraud Detection for the European Cardholders 2013 Dataset," IEEE Access, vol. 11, pp. 107348–107368, 2023, doi: 10.1109/ACCESS.2023.3320072.

[27] J. K. Afriyie et al., "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions," Decision Analytics Journal, vol. 6, Mar. 2023, doi: 10.1016/j.dajour.2023.100163.
[28] N. Rtayli and N. Enneya, "Selection features and support vector machine for credit card risk identification," in Procedia Manufacturing, Elsevier B.V., 2020, pp. 941–948. doi: 10.1016/j.promfg.2020.05.012.
[29] A. Aslam and A. Hussain, "A Performance Analysis of Machine Learning Techniques for Credit Card Fraud Detection," Journal on Artificial Intelligence, vol. 6, no. 1, pp. 1–21, 2024, doi: 10.32604/jai.2024.047226.
[30] A. A. Taha and S. J. Malebary, "An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine," IEEE Access, vol. 8, pp. 25579–25587, 2020, doi: 10.1109/ACCESS.2020.2971354.
[31] K. Vengatesan, A. Kumar, S. Yuvraj, V. D. Ambeth Kumar, and S. S. Sabnis, "Credit card fraud detection using data analytic techniques," Advances in Mathematics: Scientific Journal, vol. 9, no. 3, pp. 1185–1196, 2020, doi: 10.37418/amsj.9.3.43.
[32] L. Chen, B. Xiu, and Z. Ding, "Finding Misstatement Accounts in Financial Statements through Ontology Reasoning," IEEE Access, pp. 1–1, Aug. 2020, doi: 10.1109/access.2020.3014620.
[33] V. K. Kumar S, V. v Kumar G, and V. A. Shankar, "Credit Card Fraud Detection using Machine Learning Algorithms." [Online]. Available: www.ijert.org
[34] M. G. Saragih, J. Chin, R. Setyawasih, P. T. Nguyen, and K. Shankar, "Machine learning methods for analysis fraud credit card transaction," International Journal of Engineering and Advanced Technology, vol. 8, no. 6 Special issue, pp. 870–874, Aug. 2019, doi: 10.35940/ijeat.F1164.0886S19.
[35] G. Luo, W. Li, and Y. Peng, "Overview of Intelligent Online Banking System Based on HERCULES Architecture," IEEE Access, vol. 8, pp. 107685–107699, 2020, doi: 10.1109/ACCESS.2020.2997079.
[36] P. Visalakshi, K. v Madhuvani, S. Sunilraja, and A. Professor, "Detecting Credit Card Frauds Using Different Machine Learning Algorithms," 2021. [Online]. Available: http://annalsofrscb.ro
[37] K. Vuppula, "An advanced machine learning algorithm for fraud financial transaction detection," 2021. [Onlin e]. Available: www.jidps.com
[38] S. Hamal and O. Senvar, "Comparing performances and effectiveness of machine learning classifiers in detecting financial accounting fraud for turkish smes," International Journal of Computational Intelligence Systems, vol. 14, no. 1, pp. 769–782, 2021, doi: 10.2991/ijcis.d.210203.007.
[39] A. Izotova and A. Valiullin, "Comparison of Poisson process and machine learning algorithms approach for credit card fraud detection," in Procedia Computer Science, Elsevier B.V., 2021, pp. 721–726. doi: 10.1016/j.procs.2021.04.214.
[40] D. V. Kute, B. Pradhan, N. Shukla, and A. Alamri, "Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering-A Critical Review," IEEE Access, vol. 9. Institute of Electrical and Electronics Engineers Inc., pp. 82300–82317, 2021. doi: 10.1109/ACCESS.2021.3086230.

[41] B. Stojanović et al., "Follow the trail: Machine learning for fraud detection in fintech applications," Sensors, vol. 21, no. 5, pp. 1–43, Mar. 2021, doi: 10.3390/s21051594.

[42] H. Zhou, G. Sun, S. Fu, L. Wang, J. Hu, and Y. Gao, "Internet Financial Fraud Detection Based on a Distributed Big Data Approach with Node2vec," IEEE Access, vol. 9, pp. 43378–43386, 2021, doi: 10.1109/ACCESS.2021.3062467.

[43] M. K. Severino and Y. Peng, "Machine learning algorithms for fraud prediction in property insurance: Empirical evidence using real-world microdata," Machine Learning with Applications, vol. 5, p. 100074, Sep. 2021, doi: 10.1016/j.mlwa.2021.100074.

[44] F. Ferreira, N. Lourenco, B. Cabral, and J. P. Fernandes, "When Two are Better Than One: Synthesizing Heavily Unbalanced Data," IEEE Access, vol. 9, pp. 150459–150469, 2021, doi: 10.1109/ACCESS.2021.3126656.

[45] M. El-Bannany, A. H. Dehghan, and A. M. Khedr, "Prediction of Financial Statement Fraud using Machine Learning Techniques in UAE," in 18th IEEE International Multi-Conference on Systems, Signals and Devices, SSD 2021, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 649–654. doi: 10.1109/SSD52085.2021.9429297.

[46] D. Tanouz, R. R. Subramanian, D. Eswar, G. V. P. Reddy, A. R. Kumar, and C. H. V. N. M. Praneeth, "Credit card fraud detection using machine learning," in Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021, Institute of Electrical and Electronics Engineers Inc., May 2021, pp. 967–972. doi: 10.1109/ICICCS51141.2021.9432308.

[47] T. C. Tran and T. K. Dang, "Machine Learning for Prediction of Imbalanced Data: Credit Fraud Detection," in Proceedings of the 2021 15th International Conference on Ubiquitous Information Management and Communication, IMCOM 2021, Institute of Electrical and Electronics Engineers Inc., Jan. 2021. doi: 10.1109/IMCOM51814.2021.9377352.

[48] Y. Dong, K. Xie, Z. Bohan, and L. Lin, "A Machine Learning Model for Product Fraud Detection Based on SVM," in Proceedings - 2021 2nd International Conference on Education, Knowledge and Information Management, ICEKIM 2021, Institute of Electrical and Electronics Engineers Inc., Jan. 2021, pp. 385–388. doi: 10.1109/ICEKIM52309.2021.00091.

[49] H. Wang, F. Yang, and S. Shen, "Supply Fraud Forecasting using Decision Tree Algorithm," in 2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE 2021, Institute of Electrical and Electronics Engineers Inc., Jan. 2021, pp. 344–347. doi: 10.1109/ICCECE51280.2021.9342556.

[50] M. Â. L. Moreira et al., "Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems," in Procedia Computer Science, Elsevier B.V., 2022, pp. 117–124. doi: 10.1016/j.procs.2022.11.156.

[51] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido, "A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection," IEEE Access, vol. 10, pp. 16400–16407, 2022, doi: 10.1109/ACCESS.2022.3148298.

[52] Y. Wang, Y. Zhang, Y. Lu, and X. Yu, "A Comparative Assessment of Credit Risk Model Based on Machine Learning ——a case study of bank loan data," in Procedia Computer Science, Elsevier B.V., 2020, pp. 141–149. doi: 10.1016/j.procs.2020.06.069.

[53] W. Xiuguo and D. Shengyong, "An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies Using Deep Learning," IEEE Access, vol. 10, pp. 22516–22532, 2022, doi: 10.1109/ACCESS.2022.3153478.

[54] X. Luo, S. Wang, H. Chen, and Z. Luo, "The Utility Impact of Differential Privacy on Credit Card Data in Machine Learning Algorithms," in Procedia Computer Science, Elsevier B.V., 2023, pp. 664–672. doi: 10.1016/j.procs.2023.08.036.

[55] M. J. Madhurya, H. L. Gururaj, B. C. Soundarya, K. P. Vidyashree, and A. B. Rajendra, "Exploratory analysis of credit card fraud detection using machine learning techniques," Global Transitions Proceedings, vol. 3, no. 1, pp. 31–37, Jun. 2022, doi: 10.1016/j.gltp.2022.04.006.

[56] Y. Y. Hsin, T. S. Dai, Y. W. Ti, M. C. Huang, T. H. Chiang, and L. C. Liu, "Feature Engineering and Resampling Strategies for Fund Transfer Fraud with Limited Transaction Data and a Time-Inhomogeneous Modi Operandi," IEEE Access, vol. 10, pp. 86101–86116, 2022, doi: 10.1109/ACCESS.2022.3199425.

[57] H. Wang, W. Wang, Y. Liu, and B. Alidaee, "Integrating Machine Learning Algorithms With Quantum Annealing Solvers for Online Fraud Detection," IEEE Access, vol. 10, pp. 75908–75917, 2022, doi: 10.1109/ACCESS.2022.3190897.

[58] Z. Shahbazi and Y. C. Byun, "Machine Learning-Based Analysis of Cryptocurrency Market Financial Risk Management," IEEE Access, vol. 10, pp. 37848–37856, 2022, doi: 10.1109/ACCESS.2022.3162858.

[59] N. Nguyen et al., "A Proposed Model for Card Fraud Detection Based on CatBoost and Deep Neural Network," IEEE Access, vol. 10, pp. 96852–96861, 2022, doi: 10.1109/ACCESS.2022.3205416.

[60] S. Arora, S. Bindra, S. Singh, and V. Kumar Nassa, "Prediction of credit card defaults through data analysis and machine learning techniques," in Materials Today: Proceedings, Elsevier Ltd, 2021, pp. 110–117. doi: 10.1016/j.matpr.2021.04.588.