Inspira-Journal of Commerce, Economics & Computer Science (JCECS)
(A National bi-lingual Quarterly Double Blind Open Access Peer Reviewed Refereed Journal)
ISSN : 2395-7069(Print), Impact Factor: (Cosmos 7.650 & I2OR 5.120), Volume 12, No. 03, July-September, 2026, pp. 01-11

A Comparative Analysis of Machine Learning Classifiers for Detection of Credit Card Fraud

Surbhi Bansal[1*], Dr. Reena Hooda[2] & Rohit Yadav[3]

[1,2,3]The Department of Computer Science & Engineering, Indira Gandhi University, Meerpur, Rewari, India.

[*]Corresponding Author: surbhi.cse.rs@igu.ac.in

***Citation**: Bansal, S., Hooda, R. & Yadav, R. (2026). A Comparative Analysis of Machine Learning Classifiers for Detection of Credit Card Fraud. Journal of Commerce, Economics & Computer Science, 12(03), 01–11. https://doi.org/10.62823/jcecs/12.03.9233*

Abstract

Now the days world becomes digital, credit card customers have become common; it makes the payment hassle-free. With the ease of use of credit cards; Fraudulent use of credit cards is growing as a significant affair for financial institutions and consumers on an international level. Traditional rule-based detection algorithms are ineffective in determining a transaction's fraudulent nature. First and foremost, it is imperative to comprehend the pattern of fraudulent activities. The current study explores various supervised machine-learning algorithms to analyze patterns and predict the fraudulent nature of transactions in a large dataset used for training the model. The effectiveness of different methods is assessed by comparing their accuracy, precision, F1score, and recall. In the present paper, we discuss the techniques named KNN, SVM(Support Vector Machine), Logistic Regression, Gradient Boosting, Neural Network, XG Boost, Naïve Bayes, Ada Boost, Decision Forest, and Random Forest.

***Keywords**: Credit Card Fraud, Machine Learning Algorithms, Gradient Boosting, Neural Network, Naïve Bayes.*

Introduction

In the modern era online transaction plays an integral part in our daily routines. People believe in going outside without cash or say dependent on the use of online transactions. Unquestionably, credit card use is progressing in our more digital environment because it provides businesses and consumers with flexibility and convenience. But this advancement has also brought forth an alarming increase in fraudulent activity. These fraudulent instances have far-reaching effects on individual customers in addition to financial organizations and banks, which result in financial losses and enhanced security measures. Cardholders may experience identity theft, unauthorized transactions, and possible harm to their credit scores if their personal and financial information is exposed. To reduce the dangers of credit card theft, people are advised to exercise caution, keep a close eye on their accounts, and adhere to recommended online security practices. According to the CFPB (Consumer Financial Protection Bureau) data, there were a total of 548 million active accounts on general-purpose credit cards in the United States during the fourth quarter of 2022. This follows a trend observed over ten years with consistent increments and this depicts an increase of 41% from the 389 million accounts recorded in 2013(1). According to the PWD Global Economic Survey, over 48% of corporations suffer economic crimes. “Debit and credit card fraud online rises sharply in 2020, with 225% more cases than in 2019, based on the NCRB Report"(2). The United States could lose up to 12.5 billion dollars by 2025 as an outcome of credit card theft, according to the Nilson Report(3).

According to the report presented by the National Payments Corporation of India, 90 automated teller machines (ATMs) experienced financial losses, while 641 customers overall lost 13 million Indian Rupees (equivalent to 1.3 Crores) due to fraudulent transactions(4).

A bank is a particular kind of monetary institution where customers can deposit and make investments. Fraud is referred to as illegal conduct performed with the false intention of tricking the cardholder into gaining funds from him. Cyber theft is defined as the directed behaviour of Associated Certified Professionals that involves depriving another person's money or property by lying, deceit, or other unfair means. One develops critical incompetence upon being exposed to multiple frauds. It can be referred to as credit card fraud when a human uses a card of another one's without his knowing i.e. in an illegal manner in other words without the owner's knowledge. In other words, using credit by someone else other than the authorised person. People use a compact, rectangular metallic credit card to make payments for goods and services. Purchasing things, making payments, and withdrawing money all happen in the same instant. Credit card fraud can be done in two modes (a) offline mode, and (b)online mode. In our study following are the types of credit card fraud:

Different Kinds of Credit Card Theft

Fraud using credit cards are of various types and presents an ongoing risk to the global financial system. Identity theft and account takeover, in which unauthorized access is gained, are two common types of cybercrime. Identity theft occurs when criminals exploit stolen personal information. In addition, card-not-present fraud can occur during online transactions, highlighting the importance of implementing stringent security measures to protect consumers as well as financial institutions. Some common types of fraud are discussed below:

- **Point-Of-Sale Fraud:** Point-of-sale (POS) machines have small devices attached to them that can read data. As the customer swipes their card, these devices read and keep the information from the card. This indicates a vendor providing this data with malicious intent.
- **Skimming:** Credit card information is stolen in this type of fraud using a tiny electronic device known as a "skimmer. "A "skimmer" that scans the credit card's magnetic stripe, records, and stores the information is used to swipe the card. After that, the credit card's blank magnetic strip can be filled up with identical details once more.
- **Phishing:** Phishing is a form of credit card fraud in which a misleading email, claiming to be from a bank or another financial organization, is forwarded to the consumer. When the cardholder clicks on the email link, they are taken to a fraudulent website that asks for personal information. The majority of people fall victim to this scam because the URL they receive in the email appears to be authentic.
- **Lost/Stolen Cards:** If a user loses or steals their card; Once the fraudster uses the card's details to make purchases, the cardholder finds out about the purchases through the monthly statement.
- **CNP (Card Not Present):** The consumer does not physically present the card to the merchant, which is a sort of scam that is known as electronic fraud. Without actually possessing the card, a fraudster can use it to obtain benefits if they know details about it, such as the account number and expiration date.
- **Counterfeit Card Fraud:** Skimming is a frequently employed technique for endeavouring to achieve this objective. With the precise dimensions of the real card, a counterfeit magnetic card is produced. The fraudulent card is active and able to be used for transactions.
- **ID Card Theft:** Comparable to application deception is this type of falsification. The original information of the card was stolen by the fraudster, which they obtained through identity theft, to open a new account or to use the card. This type of fraud is the hardest to identify.

Machine Learning to Identify Fraudulent Credit Card Transaction

The rapid increase in availability of data, combined with significant developments in hardware capabilities, has led to the emergence of a new era characterized by powerful and affordable machine-learning algorithms. Present work investigates a significant impact of machine learning in addressing complicated social problems. This study specifically examines the utilization of ML and artificial intelligence (AI) within financial institutions and payment processors, emphasizing their capacity to analyse extensive datasets in realtime. Consequently, These technologies facilitate the finding of not normal patterns of transactions and the identification of suspicious behavior.

As machine learning techniques continue to progress, the financial sector will probably see even more substantial growth in the future. The utilization of advanced algorithms, diverse data sources, and continuous real-time monitoring will enhance the existing security protocols and safeguard financial systems against the dynamic array of risks.

Some issues that we face on real data sets are the high-dimensionality of feature vectors, class imbalance in credit card datasets, and difficulties dealing with complex patterns and relationships in fraud data are some of the problems that non-optimization machine learning methods have when it comes to finding credit card fraud. Since our dataset is pre-processed and correlated features are used from the Kaggle dataset. In the present study, we consider ten distinct machine learning techniques as express in the below image(in Figure 1):

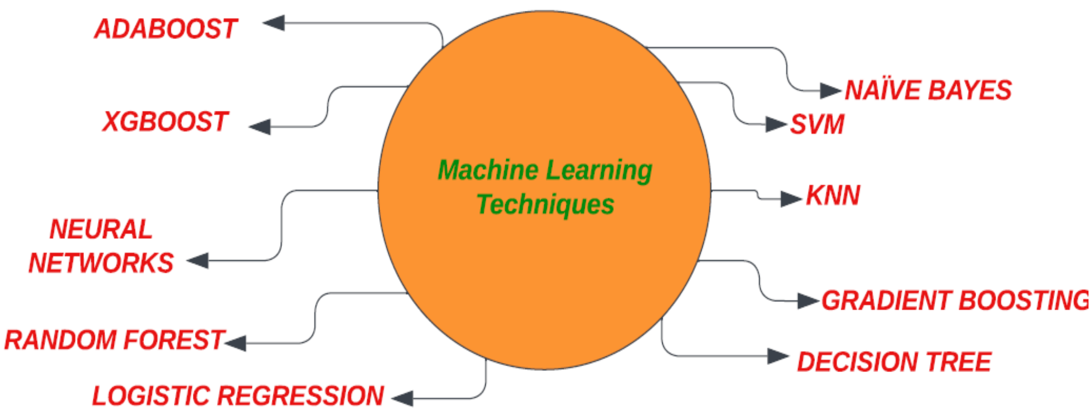

Fig. 1: Supervised Machine Learning Type

Existing Work using ML Techniques

Rtayli et al. illustrated the significance of feature selection and utilized the Recursive feature reduction technique which iteratively selects the attributes and checks the accuracy of the selected feature. Furthermore, the authors utilized hyper-plane-oriented feature optimization utilizing SVM.(5)

Preeti Rathi and Nipur Singh The research introduces a new method for identifying growing fraud using data mining techniques, particularly in credit card and web-based scenarios. A specialized fraud detection architecture is created, data is gathered, and preprocessing is done to eliminate unnecessary information. Important criteria for performance evaluation include amount of time spent searching, size of memory used, accuracy, and error rate. In addition to exploring various fraudster types, the study emphasizes the value of web mining, particularly when making utilization of datasets from phishing tanks. The report suggests that further studies should take precedence over improving performance through the integration of clustering and classification algorithms. Taking a combined approach, can expedite pattern detection and minimize search time(6).

Awoyemi et al. 2017 used highly skewed CCF data, this research analyses the performance of three different classifiers. A credit card transaction dataset has a set of transactions that were collected from cardholders of Europe. Regarding the false information of dataset, a fusion of techniques; under-sampling approach integrated with the oversampling had been employed. Further, the processed data had been an emphasis to categorize using different procedures. The work of implementation was done under the environment of python to determine the performance metrics such as coefficient correlated, neutral, and correlation factor using the Matthews function. The outcomes show that these classifiers have optimal accuracy result of Naïve Bayes was found 98 percent, in the same manner result for K-nearest neighbour about 98 percent, whereas, 55 percent found in case of Logistic Regression. The outcome tells that the k-nearest neighbor outperforms than other two classifiers (7).

Bagga et al. represent multiple classifiers to identify credit card fraud, and a method of detection is utilised tocheck the best suitable class for the provided dataset. The authors miss out on the optimization that could have increased the classification accuracy(8).

Asha et al. 2021 aimed to utilize the classifiers for the identification of frauds during transaction in real time named; K-NN, SVM, ANN. Further, to differentiate between legitimate and fraudulent transactions, a classification process was employed. The differentiation process ensures that fewer fraud activities have been done in the future (9).

Alharbi et al. 2022 present a model utilising the Kaggle dataset, this study employs a deep learning technique that deals with the issue of text data. A novel method for converting text to images, which creates small graphics to fraud in the case of credit cards, is proposed. To address class imbalance, images are fed from side to side a Convolutional Neural Network (CNN) framework, with specific weights allocated to each class. We employed deep learning and machine learning techniques to calculate the precision, efficacy and potency of the presented system. The Coarse-KNN algorithm employed deep features derived from a Convolutional Neural Network (CNN), which results an outstanding accuracy score of 99.87%.(10).

Rajab Mohsen et al.2023, presents a technique for identifying credit card fraud by implementing SVM, , random forests, ANN and logistic regression. Feature selection is prioritized to identify the primary variables impacting numerous transaction types.. Then, recall, precision, f-measure, accuracy, and confusion matrix are used to apply and assess the machine learning model. Tests conducted on the set of data which are 284,807 transactions confirms effect of method to detect and prevent fraud of credit cards(11).

Dheepa et al.2012 present that with its efficient performance and scalability, SVM is a potent instrument for fraud detection. Its kernel function and special solution provide adjustable threshold separation, facilitating effective classification in complex domains. This approach can handle high transaction volumes since it has reduced mistake rates and increased accuracy(12).

Singh et al., 2022 The study presents a new strategy, namely FFSVM, to identify credit card fraud. The proposed technique combines a support vector machine and an optimization algorithm inspired by biology. The FFSVM consists of two distinct phases. The feature subset has been optimized from the beginning by using the Firefly algorithm (FFA) and the CfsSubset Eval strategy to select the feature. The next step is to design an algorithm that can detect fraudulent cases made with credit cards, employing the support vector machine classifier (13).

The Problem Definition

Credit cards are a fundamental financial instrument that lets their users make purchases and put off interest-free payments over a while. while also providing the ease of interest-free payments over some time. Credit card customers have the benefit of deferring payment until a specified period has elapsed. This puts credit cards vulnerable to exploitation by scammers. Fraudsters can withdraw a significant sum of money without the owner's awareness and create the illusion that the legitimate cardholders made the withdrawal. The fraudsters perform this very cautiously and secretly, making it challenging to detect and apprehend them. In 2017, there were numerous instances of data breaches, resulting in a compromise of around 179 million pieces of information. Among these breaches, credit card fraud emerged as the prevailing kind of illegal activity. This is a serious problem to look into because numerous frauds are occurring worldwide, fraud with metallic card known as credit crad; is the most widespread form of fraud. Subsequently. The credit card dataset is mostly unbalanced(14) because the number of genuine records are more and the number fraud cases are less. EMV cards, which use integrated circuits to store data and improve the security of some card payments, are currently being accepted by banks. However, this does not eliminate the risk of non-card payment fraud, which continues to occur at higher rates. Based on the 2017 US Payments Forum report, criminals have shifted their attention away from conditioning associated with card-not-present (CNP) transactions due to the enhanced security measures implemented for chip cards.

Method

- **Collection of Data and Pre-Processing**

However, we will employ a dataset that comprises European cardholders executed in the month of September for a period of 48 hours. The Kaggle repository is the source of the dataset[5][6][7]our database has 284,807 transactions of which 493 transactions were identified as fraudulent(18).The dataset is significantly imbalanced which means 0.17309% of the deals belong to the positive class, which indicates that they are fraudulent, the remaining classes are not fraudulent i.e. genuine transactions. Due to privacy reasons, we can not provide information regarding the authentic characteristics with additional more context details of the facts. The dataset comprises 30 columns names which are features of our dataset, namely V1 to V28, Time, and Amount. Each and every variable in the set of data are numeric. Feature "Class" is the outcome variable; its value is 1 where fraud is present and 0 where fraud is not present, The data has already been pre-processed to train the model using numerous machine learning techniques. At the initial stage we will do pre processing in the data set so that we get the better result for the evaluation or analysis of different supervised machine learning models.

- **Evaluation of Models**

In the next step, we will discuss supervised machine learning algorithms. Numerous algorithms can be employed. A few algorithms have been taken into consideration, counting Naïve Bayes, AdaBoost, SVM (Support Vector Machine), Gradient Boosting, Decision-Tree, XG Boost, KNN, Logistic Regression, Random Forest, and Neural Networks(19). To compute performance matrices, we divide our dataset In a ratio of 70:30, we partition the dataset into components. Seventy percent of the dataset will be used for training, while the remaining thirty percent will be assigned for testing purposes.

  - **NAÏVE BAYES:** Utilizing probability theory, the Naïve Bayes classifier[9][10] detects fraudulent transactions for credit card users. The system determines how probable a transaction is to be fraudulent based on the presence of certain features including transaction amount, location, and frequency. The simplification of calculations is achieved by assuming independence among features. if the predetermined threshold is greater than the computed probability then the transaction is marked as fraudulent. The term "naive" is used as it implies that all features, or variables, are independent of each other. However, this assumption is seldom accurate in real-life circumstances. However, this assumption makes the computation easier and generally works well in practice.
  - **ADABOOST:** AdaBoost is a classifier that improves accuracy by designing a strong model out of several weak classifiers for the credit card fraud identification process, usually decision trees. At every iteration, it gives misclassified case weights so that the next classifier can concentrate on reducing those errors. Through the use of adaptive boosting, the entire algorithm's capacity to differentiate between authentic and theft transactions is strengthened, ultimately leading to a machine-learning solution that increases credit card fraud detection's dependability and accuracy (22).
  - **SVM(Support Vector Machine):** To distinguish between data points, we use a hyperplane with N dimensions, where N is the quantity of features. To efficiently divide the two sets of data points, many hyperplanes might be used. For instance, by increasing the margin distance, we can better identify a plane that has the largest separation between data points from different categories. It enhances accuracy by enabling precise classification of future data points.(4)The SVM algorithm smoothly classifies texts. The Support Vector Machine (SVM) effectively classifies positive and negative instances by generating a clear separation with significant margins. Based on vectors of support, training points are divided into two groups via a decision surface(23).
  - **GRADIENT BOOSTING:** Gradient Boosting classifiers are effective methods to detect fraud in credit cards because they forma robust predictive model by gradually training weak learners in common practice decision trees—to fix errors produced by the earlier ones. Every individual tree in the ensemble focuses on identifying and correcting misclassified occurrences, hence enhancing the overall accuracy of the model. Through this repeated process, a strong ensemble classifier that can identify patterns suggestive of fraudulent

transactions is created, providing a strong machine-learning method to raise the accuracy and efficiency of a financial card theft detection system(24).

- **DECISION TREE:** A predictive model developing assessments using a tree-like framework is called a decision tree classifier(23). By continually splitting the facts according on the way to features like transaction volume, location, or frequency, it develops a set of rules that result in classification (fraudulent or not). Decision trees are useful to detect probable credit card fraud based on numerous transaction variables since they are comprehensible and can efficiently identify patterns among the data(25). In supervised statistical produces good performance compare to others. Three factors that affect performance: the distance to identify the nearest neighbours, the criterion to establish the classification depends on the k-nearest neighbours, and the quantity of neighbours to assign a label to the new sample. If the algorithm deems the nearest neighbour with the smallest distance as fraudulent, it also labels the new transaction as fraudulent. In this case, using Euclidean distance to calculate the distances is a smart approach. Euclidean distance for the two dimensional data points A(a1 ,b1) and B(a2,b2) is as follows;

Distance=$\sqrt{(a2-a1)^2+(b2-b1)^2}$(26)

- **XGBOOST:** XGBoostrefers to Extreme Gradient Boosting. Mostly, this strong classifier is used for identifying credit card theft. It is an optimized technique that combines numerous weak learners, usually decision trees, in a certain sequence. In XGBoost, regularization terms are added and a better-optimizing method is implemented, which improves accuracy and speed. It can deal with missing data and avoid overfitting, which makes it a strong and effective machine-learning tool for finding patterns related to fraudulent credit card transactions(27).
- **KNN:** This classifier uses the distance calculation between transactions in a multi-dimensional feature space, to detect credit card fraud. The 'k' nearest neighbours' class labels are taken into consideration when classifying a transaction. Transaction amount, location, and frequency are examples of features in this context. Based on transaction patterns, KNN provides a non-parametric and versatile method for predicting possible frauds that occur in credit cards. It is successful in detecting anomalies by utilizing the similarity of transactions. For the present dataset, we calculate the results at different values of k from 1 to 8.(23)
- **Logistic Regression:** Commonly used to specify and train hypotheses about the relationship between Classification and predictive analytics frequently use logistic models. Considering independent factors, such as voting or not voting; logistic regression(28),(29) determines the probability of an event occurring. Given the unknown nature of the result, the dependent variable can only have values between 0 and 1. The basic mathematical idea guiding logistic regression is the logit—the logarithm of nature of an odd ratio. This algorithm have categorical outcome and one or more continuous or categorical predictor variables.[25] [30].
- **Random Forest:** Developed by Adele Cutler and Leo Breiman, the Random Forest (RF)(30)an autonomous learning method compiles to generate a single outcome from the multilple decision trees. Random Forest's adaptability and simplicity for application has helped it to be adopted since it addresses classification and regression challenges. The Random Forest (RF) produces a set of random trees to build a group of trees; in which the researcher or user specifies the number of trees. The final classification conclusion is determined by analysing all the trees that are constructed using a voting method in the resultant model. An ensemble of classification or regression models that are aggregated together and find the best result or prediction. The system utilizes forward-learning ensemble models, which provide prediction outcomes by iteratively refining guesses. Boosting The precision of the tree is enhanced by boosting . With a single splita decision tree is created by using Decision Stump (DS). It is applicable for categorizing heterogeneous data sets.(31)

- **Neural Networks:** A subfield of artificial intelligence inspired by biology is called “artificial neural networks” (ANNs)(17). The construction and function of the human brain, namely the biological neural network, serve as inspiration for the computational model known as a neural network. Numerous levels of interconnections between neurones comprise the networks. The human brain's neural connectivity model bears a striking resemblance(32).The current study highlights the positive and negative aspects of each approach(33). This summary aims to offer a complete laboratory study of the benefits and drawbacks associated with every technique, which will assist in the selection process based on particular data characteristics, interpretability requirements, computational resources, and optimization requirements[10][11].

Finding and Analysis

In the present study we discussed different methods like Naïve Bayes, AdaBoost, SVM (Support Vector Machine), Gradient Boosting, Decision Forest, XG Boost, KNN, Logistic Regression, Random Forest, and Neural Networks are implemented in Python environment. We discover these methods in order to determine accuracy, F1 score, recall, and precision [7][12] and MCC(34).Accuracy is a usually employed metric in machine learning to measure the model performance. It quantifies the fraction of accurately classified cases to the all transaction of the data of the data set. This parameter is generally used where our dataset is balanced and we have to just calculate the performance of the model straightforwardly. The accuracy of a model is; the ratio of (TP+TN) and (TP+TN+FP+FN).

Precision defines the level of correctness in a positive forecast. It provides the number of correctly predicted positive or true cases out of all instances predicted as positive. The metrics are especially used in context when a dataset is highly imbalanced and false positives have significant costs e.g. detection of fraud in credit cards. The precision formula can be described as the proportion of (TP) and (TP + FP).

The recall is also known as the true positive rate and sensitivity; the metrics analyse the model’s capacity to accurately detect each and every one positive cases in the set of data. Matrics evaluate the performance significantly in cases where a dataset is highly imbalanced and false negatives have significant costs.

The recall is calculated as the ratio (TP) to (TP + FN). The F1 Score is determined by following the equation given below

F1 Score = 2 * (Precision * Recall) / (Precision + Recall).

The score of F1 is a quantitative metric used to estimate a classification model's performance, especially when working with unbalanced datasets. It is particularly useful when trying to strike an appropriate equilibrium between two requirements since it provides the ideal equilibrium between recall and precision. The F1 score ranges from 0 to 1, where 0 denotes the lowest performance and 1 the greatest, demonstrating a better balance between precision and recall.The equation that follows is used to determine the F1 Score:

F1 Score = 2 * (Precision * Recall) / (Precision + Recall).

MCC: To estimate quality of binary classifier it is a immensely trustworthy statistical metric usually used in ML models. The Matthews Correlation Coefficient employs true positives , false positives , true negatives , and false negatives to compute the efficacy of various classifiers i.e. it takes all quadrant into accountof the confusion matrix. The MCC value ranges from -1 to 1, where 1 denoting that classifier shows a great outcome.

MCC = (TP * TN - FP * FN) / √((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)) (35)

TP, FP, TN and FN stands forTrue Positive, False Positive, True Negative, and False Negative, respectively.

The table (fig. 2) given below shows the calculation of MCC. It is regarded as a more informative and reliable measure than accuracy or the F1 score when evaluating classification models. The reason for this is that the MCC considers all four quadrants of the confusion matrix, resulting in a fair and unbiased measure, even when we compute on set of data that include imbalanced cases. The MCC is a numerical measure that varies from +1 to -1. The outcome -1 presents full disagreement, +1 represents perfect agreement, and 0 suggests random guesses. The continuous scale used in MCC makes it particularly effective for evaluating the performance of classifiers, especially in situations when the distribution of classes is unequal.

Technique	TP	TN	FP	FN	TP*TN	FP*FN	TP*TN – FP*FN	TP+FP	FN+TN	FP+TN	TP+FN	(TP+FP)*(FN+TN)*(FP+TN)*(TP+FN)	SQRT((TP+FP)*(FN+TN)*(FP+TN)*(TP+FN))	TP*TN –FP*FN /(TP+FP)*(FN+TN)*(FP+TN)*(TP+FN)
KNN	85287	0	152	4	0	608	-608	85439	4	152	85291	4430604071392	2104900	-0.00029
Logistic Regression	85243	44	63	93	3750692	5859	3744833	85306	137	107	85336	106712723809744	10330185	0.362514
Decision Tree	85249	38	44	112	3239462	4928	3234534	85293	150	82	85361	89552558007900	9463221	0.341801
Random Forest	85278	9	37	119	767502	4403	763099	85315	128	46	85397	42897878083840	6549647	0.11651
Gradient Boosting	85260	27	111	45	2302020	4995	2297025	85371	72	138	85305	72359646868080	8506447	0.270033
SVM	85287	0	156	0	0	0	0	85443	0	156	85287	0	0	Invalid
Neural Networks	85270	17	127	29	1449590	3683	1445907	85397	46	144	85299	48251062128672	6946298	0.208155
XGBoost	85282	5	35	121	426410	4235	422175	85317	126	40	85403	36723091865040	6059958	0.069666
Naive Bayes	84632	655	58	98	55433960	5684	55428276	84690	753	713	84730	3852599334909300	62069311	0.893006
AdaBoost	85260	27	48	108	2302020	5184	2296836	85308	135	75	85368	73736055108000	8586970	0.267479

Fig. 2: MCC Result(Calculated)

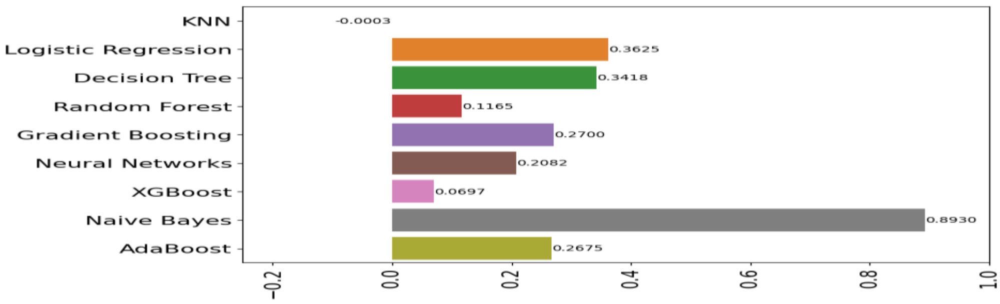

Fig. 3: MCC Result(Graph)

The bar graph shown in Figure 3 illustrates the performance of 10 supervised ML techniques in terms of the Matthews correlation coefficient (MCC). The Naive Bayes algorithm delivers finer performance relative to the other listed algorithms, while KNN exhibits the worst outcome.The table (fig 4) given below gives the result of 10 supervised learning algorithms, here in the given table we can easily compare the performance of techniques on the bases of their results individually; it calculates Recall, Precision, F1- score,Support and MCC for both the classes named 0 (non-fraud class) and 1(fraud class).

Technique	Training Accuracy	Testing Accuracy	Precision 0	Precision 1	Recall 0	Recall 1	F1-Score 0	F1-Score 1	Support 0	Support 1	Total Support	Accuracy	MCC
KNN	0.9983%	0.9982%	1.00	0.03	1.00	1.00	1.00	0.05	85439	4	–	1	-0.00029
Logistic Regression	0.9989%	0.9987%	1.00	0.60	1.00	0.68	1.00	0.63	85306	137	85443	1	0.362514
Decision Tree	100%	0.9990%	1.00	0.72	1.00	0.75	1.00	0.73	85293	150	85443	1	0.341801
Random Forest	100%	0.9994%	1.00	0.76	1.00	0.93	1.00	0.84	85315	128	85443	1	0.11651
Gradient Boosting	0.9987%	0.9983%	1.00	0.29	1.00	0.62	1.00	0.39	85371	72	85443	1	0.270033
SVM	0.9983%	0.9981%	1.00	0.00	1.00	0.00	1.00	0.00	85443	0	85443	1	Invalid
Neural Networks	0.9984%	0.9983%	1.00	0.19	1.00	0.63	1.00	0.29	85397	46	85443	1	0.208155
XGBoost	100%	0.9995%	1.00	0.78	1.00	0.96	1.00	0.86	85317	126	85443	1	0.069666
Naive Bayes	0.9922%	0.9916%	0.99	0.63	1.00	0.13	1.00	0.22	84690	753	85443	0.99	0.893006
AdaBoost	0.9993%	0.9991%	1.00	0.69	1.00	0.80	1.00	0.74	85308	135	85443	1	0.267479

Fig. 4: To compare the different parameters of the result

Conclusion

This research article investigates various supervised machine-learning approaches to spot occurrences of false transection with credit cards. It is abundantly clear that applying machine learning strategies in this area is critical. This work presents a thorough analysis of machine learning strategies, delving deeply into their complexity. In addition, the study analyses a variety of parametric quantifiers, incorporating accuracy, precision, recall, F1 score, support, and MCC to determine the value of these techniques for implementation. Through the process of comparing models, we acquire a comprehensive comprehension of their strengths and weaknesses. To enhance and streamline fraud detection systems, future research could investigate hybrid models or novel approaches.

References

1. NerdWallet [Internet]. 2024 [cited 2024 Jul 20]. Credit Card Data, Statistics and Research. Available from: https://www.nerdwallet.com/article/credit-cards/credit-card-data
2. Das P. EdexLive [Internet]. 2021 [cited 2024 Jul 20]. NCRB Report 2020: Debit, credit card fraud online climbs steeply, 225% more cases when compared to 2019. Available from: https://www.edexlive.com/news/2021/Sep/16/ncrb-report-2020-debit-credit-card-fraud-online-climbs-steeply-225-more-cases-when-compared-to-2-24053.html
3. Prajapati D, Tripathi A, Mehta J, Jhaveri K, Kelkar V. Credit Card Fraud Detection Using Machine Learning. In: 2021 International Conference on Advances in Computing, Communication, and Control (ICAC3) [Internet]. 2021 [cited 2024 Jul 15]. p. 1–6. Available from: https://ieeexplore.ieee.org/document/9697227 doi:10.1109/ICAC353642.2021.9697227
4. Singh A, Jain A. An Empirical Study of AML Approach for Credit Card Fraud Detection–Financial Transactions. Int J Comput Commun CONTROL. 2020 Feb 2;14(6):670. doi:10.15837/ijccc.2019.6.3498
5. Rtayli N, Enneya N. Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization. J Inf Secur Appl. 2020 Dec 1;55:102596. doi:10.1016/j.jisa.2020.102596
6. Rathi P, Singh N. A NOVEL APPROACH TO DETECTION OF EMERGING FRAUD USING MINING TECHNIQUES. ICTACT J SOFT Comput. 2020;11(01).

7. Awoyemi J, Adetunmbi A, Oluwadare S. Credit card fraud detection using machine learning techniques: A comparative analysis. In. 2017. p. 1–9. doi:10.1109/ICCNI.2017.8123782
8. Bagga S, Goyal A, Gupta N, Goyal A. Credit Card Fraud Detection using Pipeling and Ensemble Learning. Procedia Comput Sci. 2020 Jan 1;173:104–12. doi:10.1016/j.procs.2020.06.014
9. Rb A, Kr SK. Credit card fraud detection using artificial neural network. Glob Transit Proc. 2021 Jun 1;1st International Conference on Advances in Information, Computing and Trends in Data Engineering (AICDE - 2020)2(1):35–41. doi:10.1016/j.gltp.2021.01.006
10. Alharbi A, Alshammari M, Okon OD, Alabrah A, Rauf HT, Alyami H, et al. A Novel text2IMG Mechanism of Credit Card Fraud Detection: A Deep Learning Approach. Electronics. 2022 Jan;11(5):5. doi:10.3390/electronics11050756
11. Mohsen O, Nasserddine G, Massoud M. Credit Card Fraud Detector Based on Machine Learning Techniques. J Comput Sci Technol Stud. 2023 Jun 30;5. doi:10.32996/jcsts.2023.5.2.2
12. V. D, R. D. BEHAVIOR BASED CREDIT CARD FRAUD DETECTION USING SUPPORT VECTOR MACHINES. ICTACT J Soft Comput. 2012 Jul 1;02(04):391–7. doi:10.21917/ijsc.2012.0061
13. Singh A, Jain A, Biable SE. Financial Fraud Detection Approach Based on Firefly Optimization Algorithm and Support Vector Machine. Appl Comput Intell Soft Comput. 2022;2022(1):1468015. doi:10.1155/2022/1468015
14. Kennedy RKL, Villanustre F, Khoshgoftaar TM, Salekshahrezaee Z. Synthesizing class labels for highly imbalanced credit card fraud detection data. J Big Data. 2024 Mar 9;11(1):38. doi:10.1186/s40537-024-00897-7
15. Mienye ID, Sun Y. A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection. IEEE Access. 2023;11:30628–38. doi:10.1109/ACCESS.2023.3262020
16. Ileberi E, Sun Y, Wang Z. A machine learning based credit card fraud detection using the GA algorithm for feature selection | Journal of Big Data | Full Text [Internet]. [cited 2024 Jul 15]. Available from: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00573-8
17. Ileberi E, Sun Y, Wang Z. A machine learning based credit card fraud detection using the GA algorithm for feature selection. J Big Data. 2022 Dec;9(1):24. doi:10.1186/s40537-022-00573-8
18. Ndama O, Bensassi I, En-Naimi EM. Optimizing credit card fraud detection: a deep learning approach to imbalanced datasets. Int J Electr Comput Eng IJECE. 2024 Aug 1;14(4):4802. doi:10.11591/ijece.v14i4.pp4802-4814
19. Alarfaj FK, Malik I, Khan HU, Almusallam N, Ramzan M, Ahmed M. Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms. IEEE Access. 2022;10:39700–15. doi:10.1109/ACCESS.2022.3166891
20. Husejinovic A. Credit card fraud detection using naive Bayesian and C4.5 decision tree classifiers. Vol. 8. 2020 Jan 1;8:1–5. doi:10.21533/pen.v%25vi%25i.300
21. Gupta A, Lohani MC, Manchanda M. Financial fraud detection using naive bayes algorithm in highly imbalance data set. J Discrete Math Sci Cryptogr [Internet]. 2021 Jul 4 [cited 2024 Jul 15]. Located at: world. Available from: https://www.tandfonline.com/doi/abs/10.1080/09720529.2021.1969733
22. Gedela B, Karthikeyan PR. Credit Card Fraud Detection using AdaBoost Algorithm in Comparison with Various Machine Learning Algorithms to Measure Accuracy, Sensitivity, Specificity, Precision and F-score. In: 2022 International Conference on Business Analytics for Technology and Security (ICBATS) [Internet]. 2022 [cited 2024 Jul 30]. p. 1–6. Available from: https://ieeexplore.ieee.org/document/9759022 doi:10.1109/ICBATS54253.2022.9759022
23. Alarfaj FK, Malik I, Khan HU, Almusallam N, Ramzan M, Ahmed M. Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms. IEEE Access. 2022;10:39700–15. doi:10.1109/ACCESS.2022.3166891

24. Novakovic J, Markovic S. Classifier Ensembles for Credit Card Fraud Detection. In: 2020 24th International Conference on Information Technology (IT) [Internet]. 2020 [cited 2024 Jul 30]. p. 1–4. Available from: https://ieeexplore.ieee.org/document/9070534 doi:10.1109/ IT48810.2020.9070534

25. Mohsen OR, Nasserddine G, Massoud M fawaz. (PDF) Credit Card Fraud Detector Based on Machine Learning Techniques [Internet]. [cited 2024 Jul 15]. Available from: https://www.researchgate.net/publication/372101533_Credit_Card_Fraud_Detector_Based_on_Machine_Learning_Techniques

26. Liu Q, Hu D, Yan Q. Decision tree algorithm based on average Euclidean distance. In: 2010 2nd International Conference on Future Computer and Communication [Internet]. 2010 [cited 2024 Jul 23]. p. V1-507-V1-511. Available from: https://ieeexplore.ieee.org/abstract/document/5497736 doi:10.1109/ICFCC.2010.5497736

27. Du H, Lv L, Wang H, Guo A. A novel method for detecting credit card fraud problems. Alex SA, editor. PLOS ONE. 2024 Mar 6;19(3):e0294537. doi:10.1371/journal.pone.0294537

28. Wang T, Zhao Y. Credit Card Fraud Detection using Logistic Regression. In: 2022 International Conference on Big Data, Information and Computer Network (BDICN) [Internet]. 2022 [cited 2024 Jul 15]. p. 301–5. Available from: https://ieeexplore.ieee.org/document/9758405 doi:10.1109/BDICN55575.2022.00064

29. Mahajan AM, Baghel VS, Jayaraman R. Credit Card Fraud Detection using Logistic Regression with Imbalanced Dataset [Internet]. [cited 2024 Jul 15]. Available from: https://ieeexplore.ieee.org/document/10112302

30. Krishna MV, Praveenchandar J. Comparative Analysis of Credit Card Fraud Detection using Logistic regression with Random Forest towards an Increase in Accuracy of Prediction. In: 2022 International Conference on Edge Computing and Applications (ICECAA) [Internet]. 2022 [cited 2024 Jul 15]. p. 1097–101. Available from: https://ieeexplore.ieee.org/document/9936488 doi:10.1109/ICECAA55415.2022.9936488

31. Randhawa K, Loo CK, Seera M, Lim CP, Nandi AK. Credit Card Fraud Detection Using AdaBoost and Majority Voting. IEEE Access. 2018;6:14277–84. doi:10.1109/ ACCESS.2018.2806420

32. Seera M, Lim CP, Kumar A, Dhamotharan L, Tan KH. An intelligent payment card fraud detection system. Ann Oper Res. 2024 Mar 1;334(1):445–67. doi:10.1007/s10479-021-04149-2

33. Mniai A, Tarik M, Jebari K. A Novel Framework for Credit Card Fraud Detection. IEEE Access. 2023;11:112776–86. doi:10.1109/ACCESS.2023.3323842

34. Kalid SN, Khor KC, Ng KH, Tong GK. Detecting Frauds and Payment Defaults on Credit Card Data Inherited With Imbalanced Class Distribution and Overlapping Class Problems: A Systematic Review. IEEE Access. 2024;12:23636–52. doi:10.1109/ACCESS.2024.3362831

35. Dornadula VN, Geetha S. Credit Card Fraud Detection using Machine Learning Algorithms. Procedia Comput Sci. 2019;165:631–41. doi:10.1016/j.procs.2020.01.057

❑❑❑