International Journal of Technology Research and Management
ISSN (Online): 2348-9006
Vol 11 Issue 09 September 2024

Performance Analysis of Machine Learning for Card Fraud Detection

Suman Bharti[1], Dr. Sadhna K. Mishra[2]
Research Scholar[1], Head & Professor[2]
Department of CSE, LNCT, Bhopal[1,2]
sumanbhartisavings@gmailcom[1], sadhnak@lnct.ac.in[2]

Abstract: *Credit card being one of the most used financial products is designed to make purchases such as gas, groceries, TVs, traveling, shopping bills and so on because of non-availability of funds at that instance. Credit cards are of most value that provides various benefits in the form of points while using them for different types of transactions. machine learning is the semi-automated extraction of knowledge from data. Broadly speaking, machine learning (ML) deals with the question of how to build computer programs that learn from data and, as a result, can generate programs that generalize from that data in the form of a program that reflects concepts implicit in the underlying data. In this research work presents comparative study based on different machine learning techniques, and study suggest that proposed model gives better results than other models.*

Keywords: *Artificial intelligence, Classification, Supervised Machine learning, Credit Card, Fraud Detection, Online trnsactions*.

1. INTRODUCTION

Information technology advancements have significantly impacted the financial sector, leading to the broad adoption of electronic commerce (e-commerce) platforms. Also, the recent outbreak of the novel corona virus (COVID-19) pandemic has further shown the need for a more digital world and further expanded the e-commerce industry. One of the major issues associated with modern e-commerce is the high cases of credit card fraud. Also, in the last decade, there has been an increase in credit card fraud, which is a huge burden on financial institutions. The increased credit card fraud rate is associated with the expansion of e-commerce and increased online transactions. Artificial intelligence (AI) and machine learning applications in the financial sector can produce excellent results for companies, such as improved efficiency, reduced operational cost, and enhanced customer satisfaction. Several ML-based systems have been developed to detect credit card fraud.

E-commerce has flourished in the recent decades. As an increasing number of people are accustomed to online trans1actions, this has contributed to the prevalence of card payments. Unfortunately, the prevailing emergence of spending behavior has become an ideal condition for the increase in fraudulent activities. The Oxford Dictionary has defined fraud [1] as wrongful or criminal deception that results in financial or personal gain. Fraud detection is the process of identifying cardholders' unusual behaviors when compared to their prior card usage profile. Based on such differences, an alert is sent if the target transactions have a probability exceeding the threshold of being classified as fraud. Fraudulent transactions are typically performed via unauthorized access to card information, such as credit card numbers [2], email addresses, phone numbers [3], and many more to steal money.

Nowadays, advancement in computing technology has made it possible for internet users to speedy upload, retrieve and process huge volumes of data over remote locations through high-speed linked networks. There are various applications of the Internet which include sending and receiving voice email, searching for patterns from repositories, navigating driverless cars, performing financial transactions, audio, and video streaming. However, increased usage of technology has gained the attention of attackers in various domains like intrusion in smart devices including credit card fraud, click fraud, procurement fraud, and identity theft. Fraud is a general term that can be defined as deceiving someone by stealing sensitive information for harming them financially or degrading their reputation. Furthermore, innovation in e-commerce has also attracted users to shift to online platforms for the purchase of goods and services which requires a credit/debit card. In recent years, credit or debit card holders have been the victim of a number of notable crimes in fraudulent transactions or theft.

2. RELATED WORK

[1] The advance in technologies such as e-commerce and financial technology (FinTech) applications have sparked an increase in the number of online card transactions that occur on a daily basis. As a result, there has been a spike in credit card fraud that affects card issuing companies, merchants, and banks. It is therefore essential to develop mechanisms that ensure the security and integrity of credit card transactions. In this research, we implement a machine learning (ML) based framework for credit card fraud detection using a real world imbalanced datasets that were generated from European credit cardholders. To solve the issue of class imbalance, we re-sampled the dataset using the Synthetic Minority over-sampling Technique (SMOTE). This framework was evaluated using the following ML methods: Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Decision Tree (DT), and Extra Tree (ET). These ML algorithms were coupled with the Adaptive Boosting (AdaBoost) technique to increase their quality of classification. The models were evaluated using the accuracy, the recall, the precision, the Matthews Correlation Coefficient (MCC), and the Area Under the Curve (AUC). Moreover, the proposed framework was implemented on a highly skewed synthetic credit card fraud dataset to further validate the results that were obtained in this research.

[2] The development of effective fraud detection algorithms is vital in minimizing these losses, but it is challenging because most credit card datasets are highly imbalanced. Also, using conventional machine learning algorithms for credit card fraud detection is inefficient due to their design, which involves a static mapping of the input vector to output vectors. Therefore, they cannot adapt to the dynamic shopping behavior of credit card clients. This paper proposes an efficient approach to detect credit card fraud using a neural network ensemble classifier and a hybrid data re-sampling method. The ensemble classifier is obtained using a long short term memory (LSTM) neural network as the base learner in the adaptive boosting (AdaBoost) technique. Meanwhile, the hybrid re-sampling is achieved using the synthetic minority oversampling technique and edited nearest neighbor (SMOTE-ENN) method. The effectiveness of the proposed method is demonstrated using publicly available real-world credit card transaction datasets. The performance of the proposed approach is benchmarked against the following algorithms: support vector machine (SVM), multilayer perceptron (MLP), decision tree, traditional AdaBoost, and LSTM.

[3] Along with the significant increase in the number of credit cards, the number of credit card frauds worldwide is increasing day by day. At the same time, the development of Internet technology has led to the emergence of new fraud methods. The traditional credit card fraud detection methods can no longer meet the needs of the current credit card financial industry development. Identifying fraudulent credit card transactions effectively, quickly and accurately has become a major concern for banks. Methods combining expert rules and statistical analysis, decision tree methods, anomaly detection methods, and feature engineering methods are used in credit card fraud detection research. Among the many methods, deep learning is a new artificial intelligence method that has developed rapidly in recent years and is widely used in credit card fraud detection research. This paper uses a self-paced ensemble neural network (SP-ENN) model to learn credit card fraud transactions by dividing the datasets with different hardness, then identifying these transactions by neural networks, and finally performing a comprehensive evaluation.

[4] This paper proposes a method, called autoencoder with probabilistic random forest (AE-PRF), for detecting credit card frauds. The proposed AE-PRF method first utilizes the autoencoder to extract features of low-dimensionality from credit card transaction data features of high-dimensionality. It then relies on the random forest, an ensemble learning mechanism using the bootstrap aggregating (bagging) concept, with probabilistic classification to classify data as fraudulent or normal. The credit card fraud detection (CCFD) dataset is applied to AE-PRF for performance evaluation and comparison. The CCFD dataset contains large numbers of credit card transactions of European cardholders; it is highly imbalanced since its normal transactions far outnumber fraudulent transactions. Data re-sampling schemes like the synthetic minority oversampling technique (SMOTE), adaptive synthetic (ADASYN), and Tomek link (T-Link) are applied to the CCFD dataset to balance the numbers of normal and fraudulent transactions for improving AE-PRF performance. Experimental results show that the performance of AE-PRF

does not vary much whether re-sampling schemes are applied to the dataset or not.

[5] Machine learning has opened up new tools for financial fraud detection. Using a sample of annotated transactions, a machine learning classification algorithm learns to detect frauds. With growing credit card transaction volumes and rising fraud percentages there is growing interest in finding appropriate machine learning classifiers for detection. However, fraud data sets are diverse and exhibit inconsistent characteristics. As a result, a model effective on a given data set is not guaranteed to perform on another. Further, the possibility of temporal drift in data patterns and characteristics over time is high. Additionally, fraud data has massive and varying imbalance. In this work, we evaluate sampling methods as a viable pre-processing mechanism to handle imbalance and propose a data-driven classifier selection strategy for characteristic highly imbalanced fraud detection data sets.

[6] Every year there is an increasing loss of a huge amount of money due to fraudulent credit card transactions. Recently there is a focus on using machine learning algorithms to identify fraud transactions. The number of fraud cases to non-fraud transactions is very low. This creates a skewed or unbalanced data, which poses a challenge to training the machine learning models. The availability of a public dataset for this research problem is scarce. The dataset used for this work is obtained from Kaggle. In this paper, we explore different sampling techniques such as under-sampling, Synthetic Minority Oversampling Technique (SMOTE) and SMOTE-Tomek, to work on the unbalanced data. Classification models, such as k-Nearest Neighbour (KNN), logistic regression, random forest and Support Vector Machine (SVM), are trained on the sampled data to detect fraudulent credit card transactions. The performance of the various machine learning approaches are evaluated for its precision, recall and F1-score. The classification results obtained is promising and can be used for credit card fraud detection.

[7] In the world of finance, as the technology grown, new systems of business making came into picture. Credit card system is one among them. But because of lot of loop holes in this system, lot of problems are aroused in this system in the method of credit card scams. Due to this the industry and customers who are using credit cards are facing a huge loss. There is a deficiency of investigation lessons on examining practical credit card figures in arrears to privacy issues. In the manuscript an attempt has been made for finding the frauds in the credit card business by using the algorithms which adopted machine learning techniques. In this regard, two algorithms are used viz Fraud Detection in credit card using Decision Tree and Fraud Detection using Random Forest. The efficiency of the model can be decided by using some public data as sample. Then, an actual world credit card facts group from a financial institution is examined. Along with this, some clatter is supplemented to the data samples to auxiliary check the sturdiness of the systems. The significance of the methods used in the paper is the first method constructs a tree against the activities performed by the user and using this tree scams will be suspected. In the second method a user activity based forest will have constructed and using this forest an attempt will be made in identifying the suspect.

[8] It is crucial to actively detect the risks of transactions in a financial company to improve customer experience and minimize financial loss. In this study, we compare different machine learning algorithms to effectively and efficiently predict the legitimacy of financial transactions. The algorithms used in this study were: MLP Repressor, Random Forest Classifier, Complement NB, MLP Classifier, Gaussian NB, Bernoulli NB, LGBM Classifier, Ada Boost Classifier, K Neighbors Classifier, Logistic Regression, Bagging Classifier, Decision Tree Classifier and Deep Learning. The dataset was collected from Kaggle depository. It consists of 6362620 rows and 10 columns. The best classifier with unbalanced dataset was the Random Forest Classifier. The Accuracy 99.97%, precession 99.96%, Recall 99.97% and the F1-score 99.96%. However, the best classifier with balanced dataset was the Bagging Classifier. The Accuracy 99.96%, precession 99.95%, Recall 99.98% and the F1-score 99.96%.

3. PROPOSED WORK

Neural networks (ANNs) are computing systems vaguely inspired by biological neural networks. They “learn” to perform tasks using examples on its own. The model is based on connected units or nodes named “artificial neurons,” similar to neurons in the brain. The connection can transmit information, a “signal,” similar to the synapse in the brain. The signal is processed and is sent to other connected neurons. An edge is a connection between the neurons. These neurons have weights that can fluctuate based on the signal strength in a connection. Artificial neurons can send a signal only when the aggregate signal surpasses the threshold. The neurons are aggregated into layers performing various types of transformations based on their inputs. From the input layer to the output layer signals travel by traversing the layers many times. This was all to meet the original goal of the ANN similar to human brain function. This changed with time as it is now being employed in computer vision, speech recognition, and machine translation, filtering of social media content, playing games, and medical diagnosis. Similar to the human brain, DL processes light as well as sound and converts into vision and hearing. They are comprised like ANNs with multiple hidden layers.

Neurons may have states, represented by real numbers in the range from 0 to 1. The weights of neurons and synapses change as learning proceeds depending upon the strength of the signal. A neural network greater than three layers is said to be a DL algorithm or a deep neural network, and a neural network less than that is a basic neural network. The signals travel from the input transverse through multiple layers to reach the output layer. AI has helped in tasks difficult to express using the classical algorithm. As of 2017, neural networks possess thousand to a few million units with millions of connections. Though the order of magnitude is less when compared to the human brain, still they perform tasks beyond human brainpower.

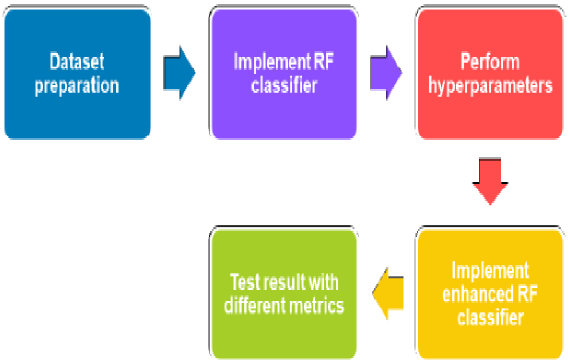
Dataset preparation
Implement RF classifier
Perform hyperparameters
Implement enhanced RF classifier
Test result with different metrics

Figure 1: Framework for predicting fraud in CC transactions.

Increase in online transactions using payment methods like credit card has also increased the fraudulent activities. Every year, a large amount of financial losses are caused by these illegal credit card transactions. No system is 100% secure and there is always a loophole in them. Therefore there is need to solve the issues of detecting fraud in transactions done by credit cards.

As shown in the block diagram, following are the modules of system:

a. Data Cleaning and Preprocessing
b. Alerts Classification
c. Alert Ranking
d. Performance Analysis

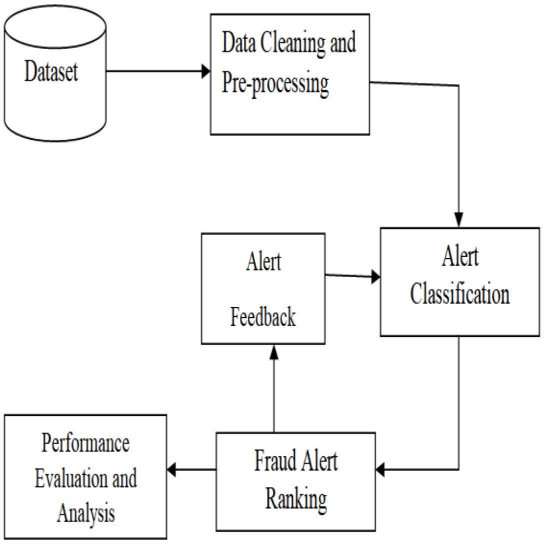
Dataset
Data Cleaning and Pre-processing
Alert Feedback
Alert Classification
Performance Evaluation and Analysis
Fraud Alert Ranking

Figure 2: Proposed system block diagram.

4. EXPERIMENTAL WORK

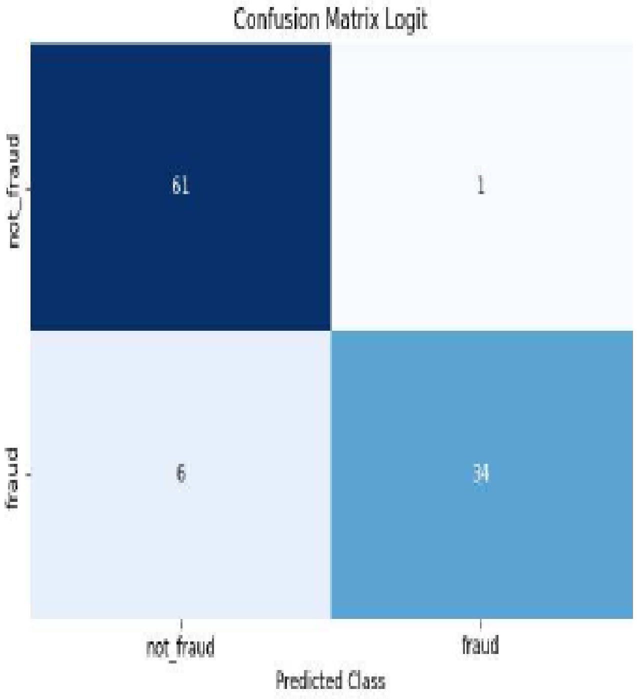

Figure 3: This picture shows that confusion matrix for artificial neural network classifier.

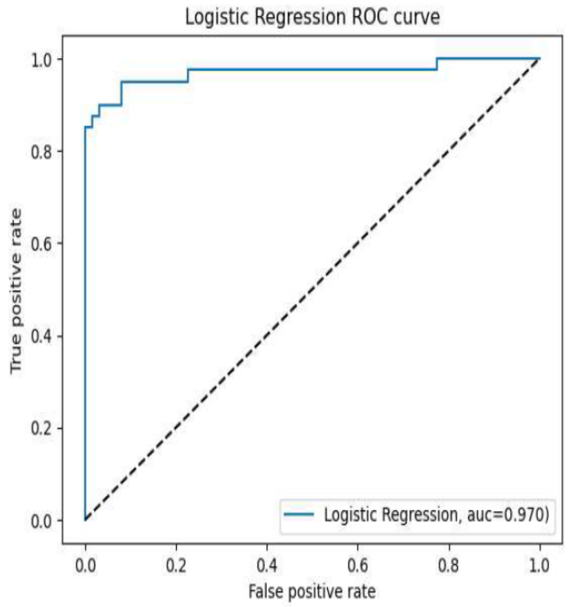

Figure 4: This picture shows that ROC curve for artificial neural network classifier.

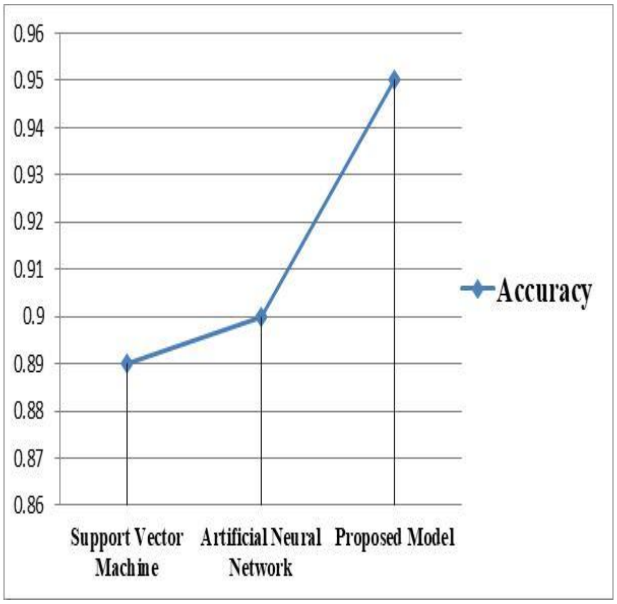

Figure 5: The above figure represents comparative study for different machine learning model and proposed model performance evaluation based on the parameters like accuracy.

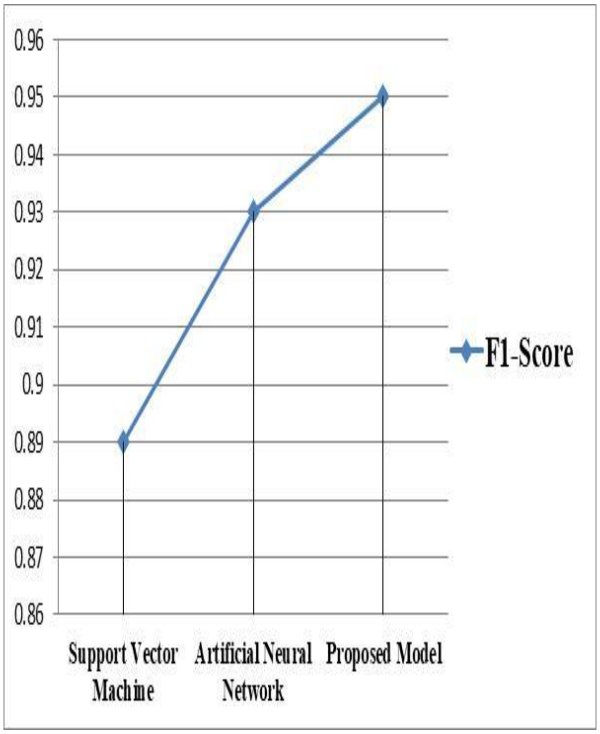

Figure 6: The above figure represents comparative study for different machine learning model and proposed model performance evaluation based on the parameters like F1-Score.

5. CONCLUSION

The advance in technologies such as e-commerce and financial technology (FinTech) applications has sparked an increase in the number of online card transactions that occur on a daily basis. As a result, there has been a spike in credit card fraud that affects card issuing companies, merchants, and banks. It is therefore essential to develop mechanisms that ensure the security and integrity of credit card transactions. In this research work, presents implement a machine learning (ML) based framework based comparative experimental work for credit card fraud detection. In future work also enhance the performance of this system using optimization techniques and deep learning approach.

REFERENCES

[1] Dilip Sharma, SandeepLavavanshi, “A Review On: Identifying Credit Card Fraud Using A Deep Learning Approach”, Journal of Data Acquisition and Processing, 2023, pp. 2907-2916.

[2] Biao Xua, Yao Wang, “E_cient Fraud Detection Using Deep Boosting Decision Trees”, Decision Support Systems, 2023, pp. 1-33.

[3] IbomoiyeDomorMienye, “A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection”, IEEE access, 2023, pp. 30628-30638.

[4] SeyedehKhadijehHashemi, SeyedehLeiliMirtaheri, “Fraud Detection in Banking Data by Machine Learning Techniques”, IEEE access, 2023, pp. 3034-3044.

[5] Wang Ning, Siliang Chen, “AMWSP-L Adaboost Credit Card Fraud Detection Method Based on Enhanced Base Classifier Diversity”, IEEE Access 2023, pp. 66488-66497.

[6] Nayyer, NadeemJavaid, “A New Framework for Fraud Detection in Bitcoin Transactions Through Ensemble Stacking Model in Smart Cities”, IEEE Access 2023, pp. 90916-90939.

[7] Yuanming Ding, Wei Kang, “Credit Card Fraud Detection Based on Improved VariationalAutoencoderGenerativeAdversarial Network”, IEEE Access 2023, pp. 83680-83692.

[8] Fuad A. Ghaleb, Faisal Saeed, “Ensemble Synthesized Minority Oversampling-Based Generative Adversarial Networks and Random Forest Algorithm for Credit Card Fraud Detection”, IEEE Access 2023, pp. 89694-89710.

[9] Oluwadare Samuel Adebayo, “Comparative Review of Credit Card Fraud Detection using Machine Learning and Concept Drift Techniques”, IJCSMC, 2023, 24-48.

[10] AnikMalaker, AbidHasanMiad, “An Approach to Detect Credit Card Fraud Utilizing Machine Learning”, Int. J. Advanced Networking and Applications, 2023, pp. 5619-5626.

[11] AltyebTaha, “A novel deep learning-based hybrid Harris hawks with sine cosine approach for credit card fraud detection”, Mathematics, 2023, pp. 23200-23217.

[12] ZhaoruiMeng, YanqiXie, Jinhua Sun, “Detecting Credit Card Fraud by Generative Adversarial Networks and Multi-head Attention Neural Networks”, IAENG International Journal of Computer Science, 2023, pp 1-7.

[13] Dhwanir Shah, Lokesh Kumar Sharma, “Credit Card Fraud Detection using Decision Tree and Random Forest”, ITM Web of Conferences, 2023, pp.1-10.

[14] Ahmed Hassan Butt, “A Review: Credit Card Fraud Detection in Banks using Machine Learning Algorithms”, 2023, pp. 1-7.

[15] BodundeOdunolaAkinyemi, DaudaAkinwuyiOlalere, “Performance Evaluation of Machine Learning Models for Cyber Threat Detection and Prevention in Mobile Money Services”, Informatica, 2023, pp.173-190.

[16] Kiran Jot Singh, Khushal Thakur, “Comparative Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection”, Springer, 2023, pp. 69-80.

[17] Biao Xua, Yao Wang, “Efficient Fraud Detection Using Deep Boosting Decision Trees”, 2023, pp. 1-34.

[18] AmmarahUroojAftab, “Fraud Detection of Credit Cards Using Supervised Machine Learning Techniques”, 2023, pp. 1-14.

[19] Ebenezer Esenogho, IbomoiyeDomorMienye, “A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection”, IEEE Access, 2022, pp. 16400-16408.

[20] Gayan K. Kulatilleke, “Credit Card Fraud Detection Classifier selection Strategy”, 2022, pp. 1-17.

[21] Wei Zhou, XiaoruiXue, “Credit card fraud detection based on self-paced ensemble neural Network”, ITCC 2022, pp. 92-99.