ISSN: 2581-3404 (Online) IF: 5.445 (SJIF)
IJIRTM, Volume-7, Issue-4, August-2023.

Credit Card Fraud Detection using Imbalance Re-Sampling with Machine Learning Based Approach

Rizwana Parveen[1], Dr. Harsh Lohiya[2]
Ph.d. Research Scholar, Department of Computer Science and Engineering[1]
Assistant Professor, Department of Computer Science and Engineering[2]
Sri Satya Sai University of Technology & Medical Sciences, Sehore (M.P.)[1, 2]

Abstract: *As a payment instrument that provides credit services, credit cards bring great convenience to modern life. Credit card and debit card transactions are divided into online spending and offline payments. With the popularity of online transactions and the development of Internet technology, credit/debit card fraud cases are rapidly increasing. The detection and identification of credit/debit card fraud can help maintain the normal development of the credit card financial industry system, which is of great significance to the national economic development and social stability. This study makes a major contribution to research on the detection of Credit Card fraud transactions through Machine Learning Algorithms such as random forest, support vector machines, and multi layer perceptron.*

Keywords: Credit card, Fraud detection, Artificial intelligence, Machine learning, Classification, Imbalance.

Introduction

Fraud in simple words can be termed as an unfair or fraudulent activity expected to result in personal and financial gain, or to injure another individual without actually contributing to clear legal impacts. The two key measures to eliminate frauds and damages due to the unethical activities are fraud avoidance and fraud detection systems. The constructive mechanism with the aim of blocking the phenomenon of fraud is fraud prevention.

The constructive method with the objective of preventing the incidence of fraud is fraud prevention. When scammers overtake the fraud prevention networks and initiate a fraudulent transaction, fraud detection systems come into consideration. No one can really recognize whether the prevention procedures have been activated by a fraudulent transaction. The intention of detection techniques is often to evaluate each transaction for the likelihood of fraud, irrespective of the prevention techniques, and to detect fraud ones as rapidly as possible after a fraudulent transaction has started to be executed by the fraudster. The most popular forms of cheating are fraud activities in credit card and e-commerce networks, laundering in financial systems, computer network cyber attacks, fraudulent conversations or utilization of some services in the field of healthcare and telecommunication structures.

Credit card typically refers to a card granted to the consumer (credit card issuer), generally enabling them to buy products or services or borrow cash in advance under the credit limits. The credit card gives the cardholder the benefit of the moment to pay the bills later in the next cycle. By bringing it through the next payment period, the credit card provides the cardholder with a benefit of time or that moment. As a very significant unique card number, each card's safety relies primarily on the physical safety of the card and the secrecy of the card number. Below figure shows the process of credit card fraud detection.

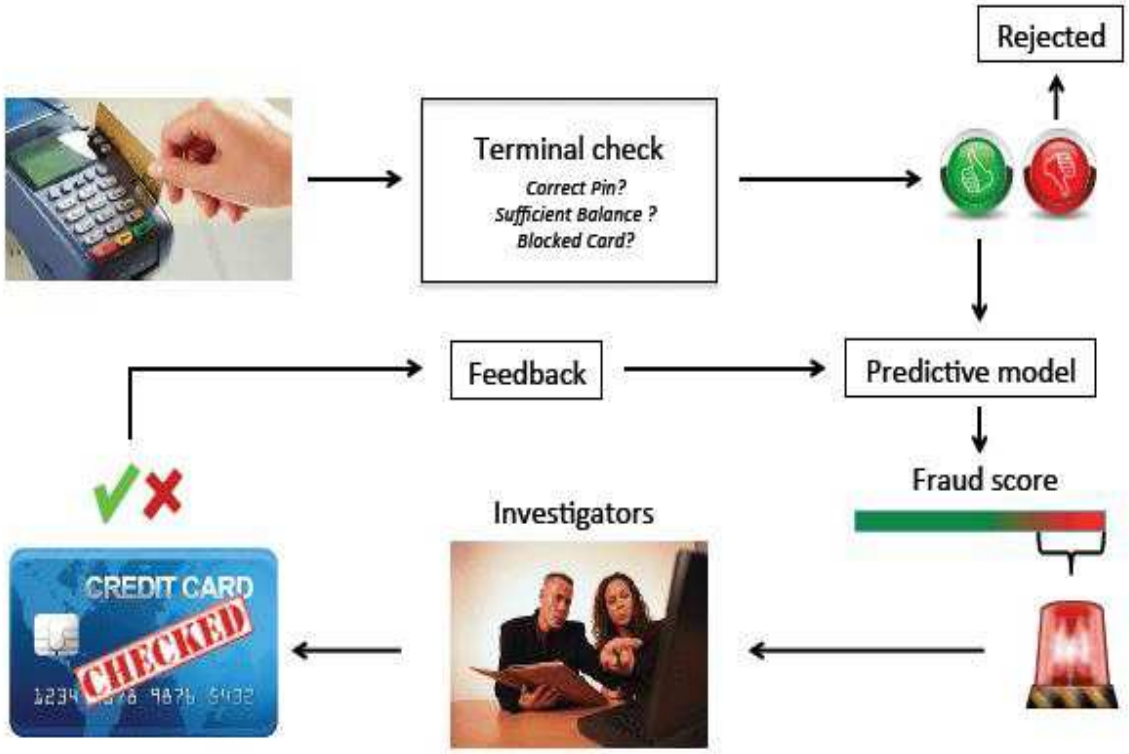

Figure 1: Credit Card Fraud Detection [14].

The rest of the paper is assembled in Sections. Section-II elaborates the proposed model using machine learning approach, Section-III represents the information of utilized dataset along with used performance parameter, Section-IV discuss the experimental work and comparative work for different machine learning approach, and finally Section-V consists of the generated conclusions of the shown study.

II. Proposed Model

Machine Learning is a branch of Artificial Intelligence that has become very popular, and useful, in the last 10 years. One definition of Machine Learning is that it 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 effect, with machine learning we have programs using data to create new programs. This is in contrast to the traditional way that programs have

been generated by human programmers in which they encode the rules that the computer follows in a programming language in order to produce a solution to a specified problem. Traditional or conventional writing of programs for a computer can be summarized as automating the procedures to be performed on input data in order to create output artifacts. Almost always, they are linear, procedural and logical.

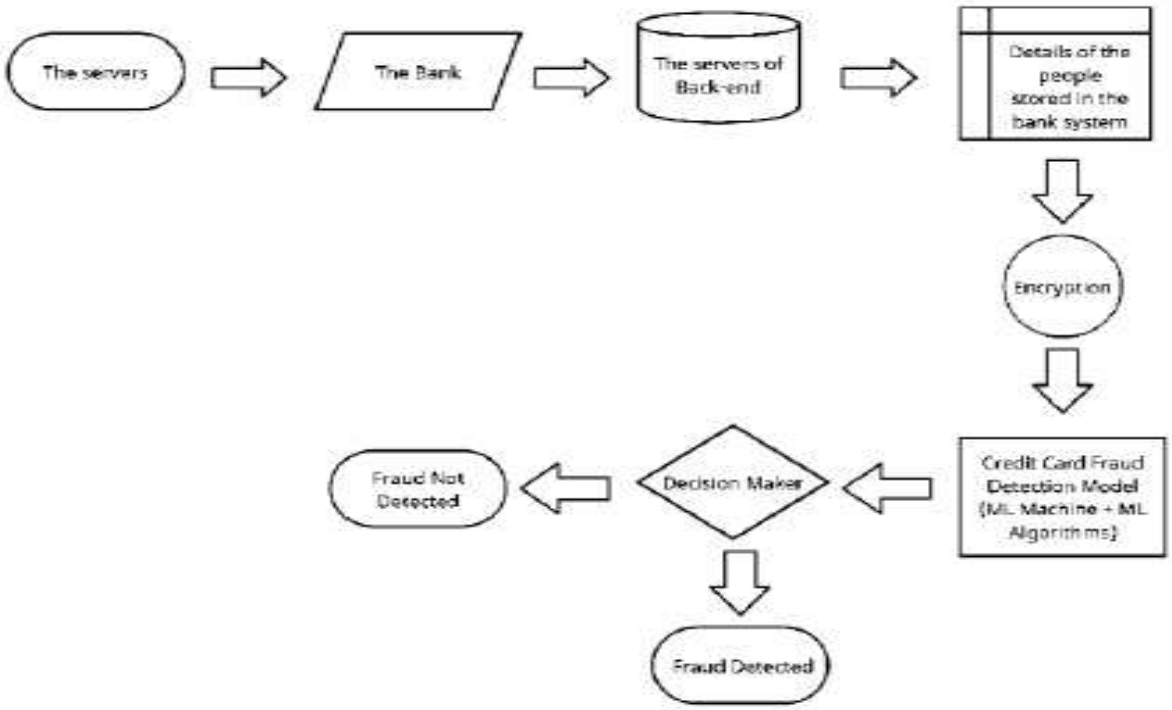

Figure 2: Architecture of Credit Card Fraud Detection.

In this paper, various Machine Learning approaches are implemented to verify that which algorithm of Machine Learning delivers the most efficient outcome and detects the fraud fast underneath whatever the situations and they are as follows:
1) Logistic Regression (LR)
2) Support Vector Machine (SVM)
3) Multi Layer perceptron (MLP)

III. Dataset and Performance Parameter

The dataset used in this research was generated from European cardholders in September 2013. This dataset is highly skewed and is publicly available through Kaggle [1]. Moreover, this dataset is not synthetic; therefore, the transactions found in it occurred over a period of time. Further, the dataset has 284807 card transactions in total whereby 99.828% are legitimate and 0.172% are fraudulent. All the features within the dataset are numerical. The class (label) is represented by the last column whereby the value of 0 represents a legitimate transaction and the value of 1 is a fraudulent activity.

Performance parameter

The performance of the models is evaluated using the following performance evaluation metrics: sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). Sensitivity, also called recall, indicates the proportion of fraud samples correctly predicted by the classifier. In contrast, specificity (true negative rate) is the proportion of legitimate transactions predicted correctly by the classifier. Meanwhile, the AUC is a measure of the classifier's ability to distinguish between legitimate and fraudulent transactions. An AUC value of 1 implies a perfect model, and the closer the AUC value is to 1, the better the classifier [2].

The F1-score is an index composed of accuracy and recall rate. It no longer pays attention to the performance of a certain aspect of the model. But it gives a comprehensive index according to the overall performance of the model. the higher the F1-score, the better the overall performance of the model.

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

where

- True positive (TP) represents an instance where a transaction is fraudulent, and the classifiers correctly classify it as fraudulent.
- True negative (TN) denotes an instance where a transaction is legitimate, and the classifiers correctly predict it as legitimate.
- False-positive (FP) represents a case where a transaction is legitimate, and the classifier classifies it as fraudulent.
- False-negative (FN) is an instance where a fraudulent transaction is wrongly classified as legitimate.

		Predicted Value	
		Non-Fraud (NO)	Fraud (YES)
Actual Value	Non-Fraud (NO)	TN	FP
	Fraud (YES)	FN	TP

Figure 3: Confusion Matrix.

IV. Experimental Result Analysis

Credit card scam finding is while a trade receipts steps to preclude whipped cash, merchandises, or amenities attained via an illegal credit card business. Credit card scam can occur together by the customer or by somebody else. To avoid happening such frauds, there are many techniques invented. If such frauds happen, then how to track the misused transactions are also improvised. Among the many methods, machine learning algorithms make accurate predictions by extracting some underlying information features based on large data samples of different dimensions. To reduce the bias caused by unbalanced data and improve the accuracy of credit card detection, scholars have mainly focused on studied in several aspects, such as data re-sampling, cost-sensitive learning, unbalanced regression, ensemble learning algorithms, and deep learning algorithms. Here we used three different machine learning model to compare the performance in the terms of performance parameter as discussed earlier, all the results are simulated with python framework.

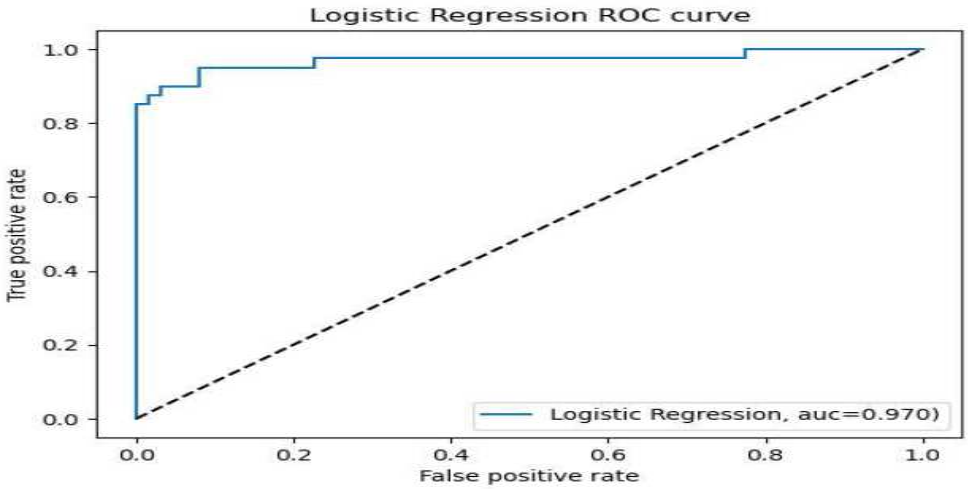

Figure 4: ROC of logistic regression.

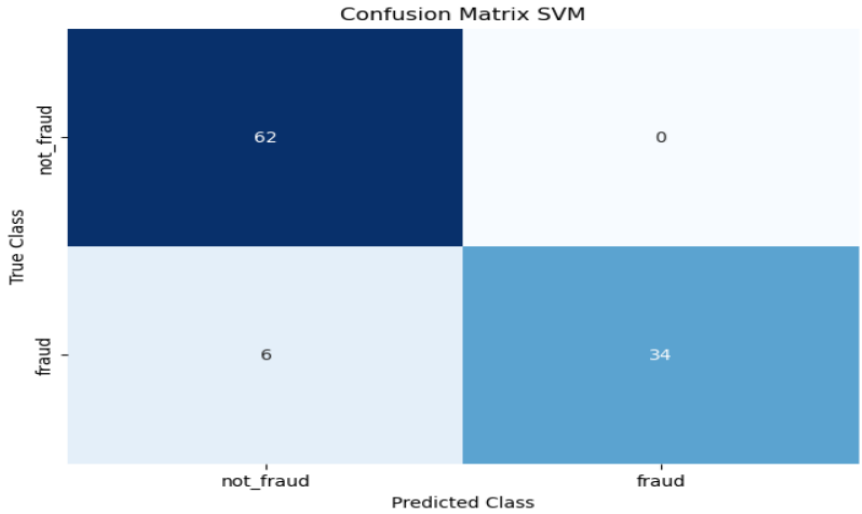

Figure 5: Confusion matrix of support vector machine.

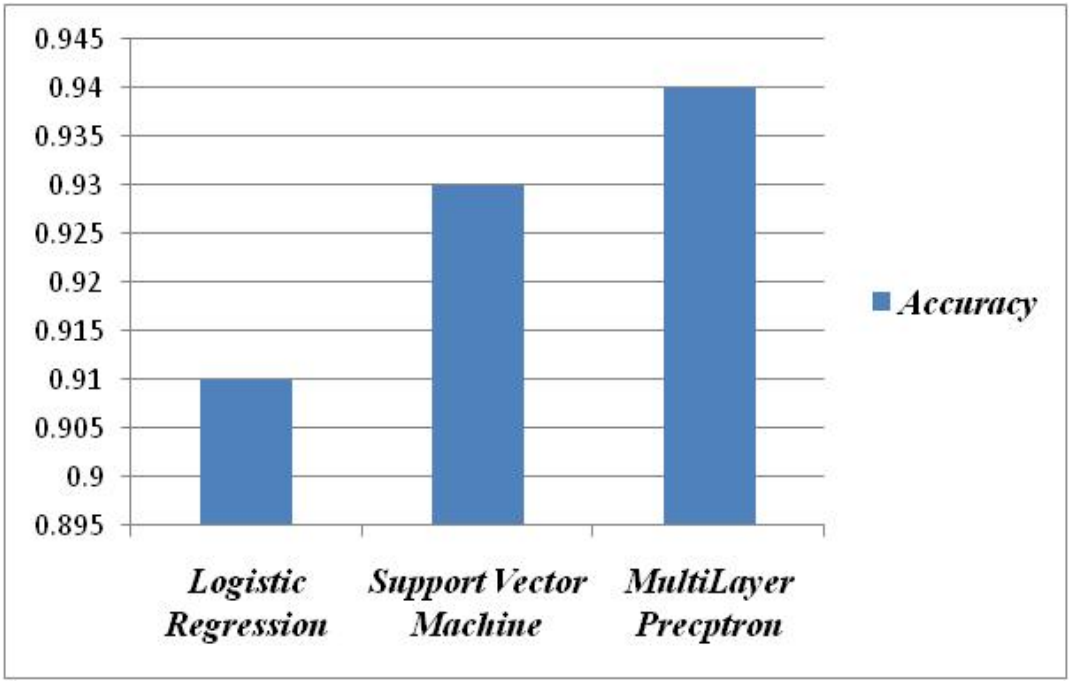

Figure 6: Accuracy using different machine learning approaches.

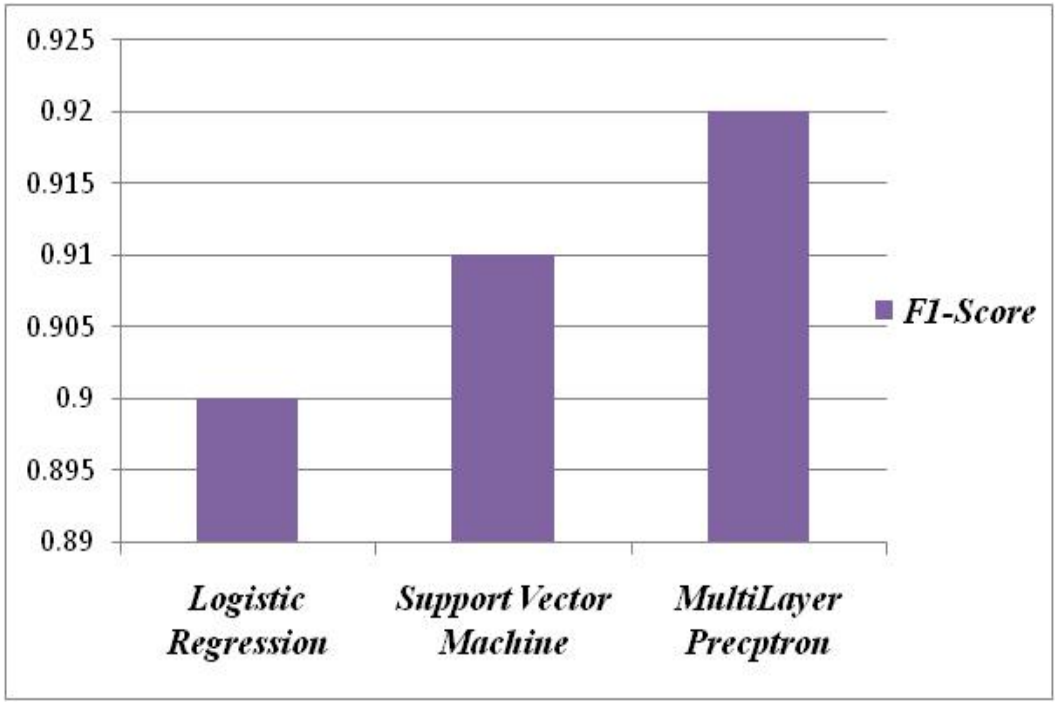

Figure 7: Precision using different machine learning approaches.

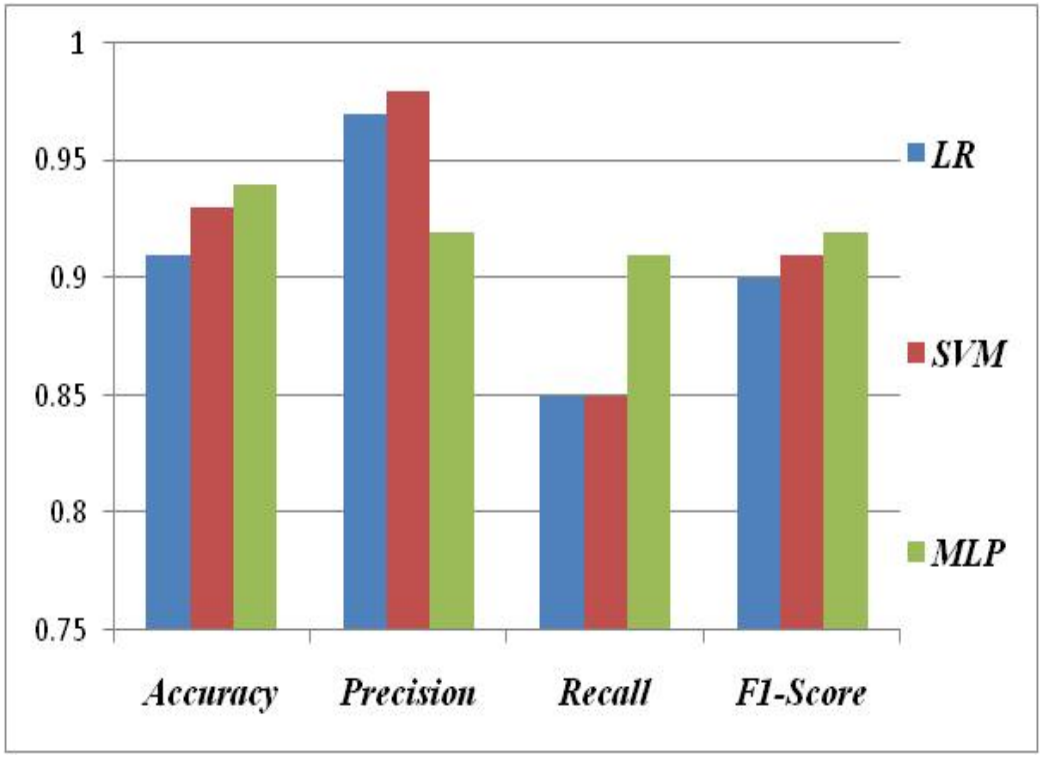

Figure 8: Accuracy, precision, recall, and F1 score using different machine learning approaches.

V. Conclusion

This paper implemented several ML algorithms for credit card fraud detection using the European credit card fraud dataset that was generated in 2013. The ML methods proposed in this work included the support vector machines, logistic regression, and multi layer perceptron. The proposed approach includes the comparison of support vector machines, logistic regression, and multi layer perceptron under sampling method which is used to handle the imbalanced data. It is shown that multi layer perceptron achieved better accuracy than support vector machines, and logistic regression. In future works, we intend to test and validate the proposed framework on additional credit card fraud datasets that will be sourced from some real time data or transactions.

References

[1] Emmanuel Ileberi, Yanxia Sun, Zenghui Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost", IEEE Access, 2021, pp. 165286-165295.

[2] Ebenezer Esenogho, Ibomoiye Domor Mienye, "A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection", IEEE Access, 2022, pp. 16400-16408.

[3] Wei Zhou, Xiaorui Xue, "Credit card fraud detection based on self-paced ensemble neural Network", ITCC 2022, pp. 92-99.

[4] Tzu-Hsuan Lin, Jehn-Ruey Jiang, "Credit Card Fraud Detection with Autoencoder and Probabilistic Random Forest", Mathematics 2021, pp. 1-16.

[5] Gayan K. Kulatilleke, "Credit Card Fraud Detection Classifier selection Strategy", 2022, pp. 1-17.

[6] Deepak Kumar Rathore, Dr. Praveen Kumar Mannepalli, "Recent Trends in Machine Learning for Health Care Sector ", International Journal of Innovative Research in Technology and Management, Vol-5, Issue-2, 2021.

[7] Konduri Praveen Mahesh, Shaik Ashar Afrouz, "Detection of fraudulent credit card transactions: A comparative analysis of data sampling and classification techniques", Journal of Physics: Conference Series, 2021, pp. 1-9.

[8] Dileep M R, Navaneeth A V, "A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms", IEEE, 2021, pp. 1025-1028.

[9] Mosa M. M. Megdad, Bassem S. Abu-Nasser, " Fraudulent Financial Transactions Detection Using Machine Learning", International Journal of Academic Information Systems Research, 2022, pp. 30-39.

[10] Shubham Shah, Dhairya Shah, "Credit Card Fraud Detection System using Machine Learning", International Journal of Research in Engineering and Science, 2022, pp. 9-14.

[11] Deepak Kumar Rathore, Dr. Praveen Kumar Mannepalli, "A Review of Machine Learning Techniques and Applications for Health Care ", International Conference on Advances in Technology, Management & Education, 2021, IEEE proceeding, 978-1-7281-8586-6/21.

[12] Appala Srinuvasu Muttipati, Sangeeta Viswanadham, "Recognizing Credit Card Fraud Using Machine Learning Methods", Turkish Journal of Computer and Mathematics Education, 2021, pp. 3271-3278.

[13] Akhil Songa, Sri Teja Kumar Reddy Tetali, Naga Sai Tanmai Raavi, "Credit Card Fraud Detection using Various Machine Learning Algorithms", International Journal for Research in Applied Science & Engineering Technology, 2022, pp. 1174-1185.

[14] G. Sudha Sadasivam, Mutyala Subrahmanyam and Dasaraju Himachalam, Bhanu Prasad Pinnamaneni, "Corporate governance fraud detection from annual reports using big data analytics", Int. J. Big Data Intelligence, Vol. 3, No. 1, 2016

[15] R Dubey, D Rathore, D Kushwaha, JP Maurya, "An empirical study of intrusion detection system using feature reduction based on evolutionary algorithms and swarm intelligence methods", International Journal of Applied Engineering Research 12 (19), 2017. pp. 8884-8889.

[16] Ophir Gottlieb, Curt Salisbury, Howard Shek, Vishal Vaidyanathan, "Detecting Corporate Fraud: An Application of Machine Learning", December 15, 2006

[17] Renjith, S. Detection of Fraudulent Sellers in Online Marketplaces using Support Vector Machine Approach. International Journal of Engineering Trends and Technology (2018).

[18] Deepak Kumar Rathore, Praveen Kumar Mannepalli, "Diseases Prediction and Classification Using Machine Learning Techniques", AIP Conference Proceedings 2424, 070001 (2022); https://doi.org/10.1063/5.0076768.

[19] Roy, Abhimanyu, et al. "Deep learning detecting fraud in credit card transactions." 2018 Systems and Information Engineering Design Symposium (SIEDS). IEEE, 2018.

[20] Pumsirirat, Apapan, and Liu Yan. "Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine." International Journal of advanced computer science and applications 9.1 (2018): 18-25.

[21] Appala Srinuvasu Muttipati, SangeetaViswanadham, "Recognizing Credit Card Fraud Using Machine Learning Methods", Turkish Journal of Computer and Mathematics Education, 2021, pp. 3271-3278.

[22] Akhil Songa, Sri Teja Kumar Reddy Tetali, Naga Sai Tanmai Raavi, "Credit Card Fraud Detection using Various Machine Learning Algorithms", International Journal for Research in Applied Science & Engineering Technology, 2022, pp. 1174-1185.