International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

# IDENTIFYING FRAUDULENT CREDIT CARD TRANSACTIONS USING ENSEMBLE LEARNING

**SUVARNA SUNIL KUMAR[1], CH. LIKITHA[2], CH. VEDA SRI[3], FATHIMA ANJUM[4]**
[1]Assistant Professor, Department of IT, Mallareddy College of Engineering For Women
[2,3,4]UG Scholar, Department of IT, Mallareddy College of Engineering For Women

## ABSTRACT

Recognizing fraudulent credit card transactions is one of the main issues facing banking institutions. Since each transaction that completes the authentication procedure must be authorized by financial institutions, a hacker might pose as the actual cardholder and execute a fraudulent transaction. In this paper, we investigated the capacity of ensemble learning methods to identify credit card frauds on two distinct data sets: the Sparkov synthetic dataset and the real dataset of consumers in the European Union. XGBoost models, random forests, and naive Bayes classifiers are applied and assessed on both datasets. Accuracy, precision, recall, and F1 score are used to measure performance. According to the results, most ensemble classifiers perform exceptionally well on the real-world dataset, but significantly poorly on the simulated dataset. This study showed that, unlike in simulated environments, credit card transaction management scripts are quickly learned in deterministic settings. It is discussed that a larger danger of card information leakage results from strict determinism and lack of randomness.

## INTRODUCTION

Credit Card Fraud (CCF) detection problem consists of spotting credit card transaction anomalies. In 2020, Visa and MasterCard had more than 2.183 million cardholders[1]. Approximately 1.4 million identity theft reports were reported in 2020, including 393.207 cases of CCF [2].The loss due to fraudulent usage of credit cards reached $28.6 billion in 2020, where it was $23.97 billion. In 2017 with an increase of 19.3 %, it is expected to reach$408 billion in the next decade according to Nilson Report of 2022 [3]. This amount shows the big loss and how it is continuously increasing despite the efforts of banks to limit the effects of credit card misuse. It inevitable to detect CCF in this era of digital payments. To address credit card misuse, several approaches have been developed belonging mainly to two classes [2]: statistical approaches and machine learning-based approaches. From a statistical point of view, detecting a fraudulent transaction is equivalent to detecting an outlier from a dataset. Consequently, a variety of statistical techniques have been developed and used namely: box and whisker plots, normal distribution-based, cluster-based, etc. The second class of credit card fraud detection techniques is composed of machine learning classifiers. A classifier is designed, developed, fitted, and tuned on a training data set to separate authentic and fraudulent transactions. Classifiers aim to learn to detect correctly the type of transaction beforehand. Machine learning techniques applied to CCF problems include decision trees, support vector machines, neural networks, regression methods, etc. [1], [4], [5]. The performance of such methods varies depending on the used data set. To improve the quality of the classification step, ensemble learning was proposed. The

basic idea of ensemble learning is to build an enhanced classifier from a set of naive/basic classifiers. The objective is then to create a strong metal earner(ensemble model) from a list of basic learners (naïve classifiers). Many ensemble models have been proposed in the literature depending on the technique used in combining basic classifiers. For instance, bagging methods create many samples with replacements from the training set and use them to fit in parallel basic classifiers. An aggregation technique is also used to build the final strong classifier like voting. The second type of ensemble learning is comprisedof boosting techniques. Boosting consists of creating sequentially basic classifiers, where the prediction error is propagated from one model to the subsequent. This process will help in boosting the performance of the last classifiers. Many variants of boosting models have been developed including Adaboost, Gradient boosting, and XG Boost among others [6]. Stacking the third type of ensemble method(known also as stacked generalization) combines classifiers to find an improved model [7]. In this paper, a set of ensemble-based classifiers is developed for the detection and prevention of abnormal transactions on credit cards. Three techniques are proposed, a naive Bayes classifier and one from each class of ensemble methods: bagging and boosting. Based, on a data set of credit card transactions of European Union consumers available in the Kaggle repository [8] and the synthetic Sparkov dataset [9]. An attempt to find the best ensemble-based model for solving the credit card detection problem is made in this work. The main objective of this paper is to compare the performance of ensemble models in learning real and synthetic datasets and not equate ensemble methods in general to other machine learning techniques .

**LITERATURE REVIEW**

**Strong valid inequalities for Boolean logical pattern generation**

- Kedong Yan, H. Ryoo
- Published in Journal of Global… 20 March 2017

0–1 multilinear programming (MP) captures the essence of pattern generation in logical analysis of data (LAD). This paper utilizes graph theoretic analysis of data to discover useful neighborhood properties among data for data reduction and multi-term linearization of the common constraint of an MP pattern generation model in a small number of stronger valid inequalities. This means that, with a systematic way to more efficiently generating Boolean logical patterns, LAD can be used for more effective analysis of data in practice. Mathematical properties and the utility of the new valid inequalities are illustrated on small examples and demonstrated through extensive experiments on 12 real-life data mining datasets.

**A multi-term, polyhedral relaxation of a 0–1 multilinear function for Boolean logical pattern generation**

- Kedong Yan, H. Ryoo
- Published in Journal of Global… 25 June 2018

0–1 multilinear program (MP) holds a unifying theory to LAD pattern generation. This paper studies a multi-term relaxation of the objective function of the pattern generation MP for a tight polyhedral relaxation in terms of a small number of stronger 0–1 linear inequalities. Toward this goal, we analyze data in a graph to

discover useful neighborhood properties among a set of objective terms around a single constraint term. In brief, they yield a set of facet-defining inequalities for the 0–1 multilinear polytope associated with the McCormick inequalities that they replace. The construction and practical utility of the new inequalities are illustrated on a small example and thoroughly demonstrated through numerical experiments with 12 public machine learning datasets.

**Generating maximum prime patterns using Benders decomposition and Apriori algorithm**

- Hany Osman
- Published in Proceedings of the… 7 March 2021

Incorporating data mining tasks in different levels of planning has become an essential tactic in business, industry, and other sectors. The rationales for implementing a data mining task, such as classification, significantly increase if the techniques used in classification provide optimal results. Logical Analysis of Data (LAD) is a classification approach known for its promising accuracy in classification and its capabilities in providing interpretable pattens. The main challenge in implementing LAD is the pattern generation problem. In this study, the pattern generation problem is solved to optimality to find maximum prime patterns. The proposed approach incorporates Benders decomposition and Apriori algorithm to generate prime patterns with high coverage from past observations. These patterns are then employed to build LAD classifiers that are used to assign class labels to unseen observations. Computational experiments conducted on seven public datasets show that results of LAD classifiers, established by using the proposed pattern generation algorithm, surpassed results of six machine learning algorithms implemented in IBM SPSS Modeler.

**EXISTING SYSTEM** Nowadays, electronic payment (e-payment) constitutes one of the main trending Fintech solutions. E-payment is defined simply as the transfer of funds via electronic channels like digital wallets, credit or debit cards, or mobile banking. Such technology offers several advantages, like reduced costs in terms of time and resources, efficiency, a cashless economy, transparency of transactions, etc. However, e-payments are not always secure and the loss of credit card information will lead surely to fund loss [10]. The number of payments completed using credit cards is continuously increasing. Consequently, the risk of fraud on credit cards will grow and the lost funds will follow the same pattern. CCF detection and prevention is then a priority for financial services providers like banks, insurance, and card payment networks (Visa, MasterCard). The problem is defined as the detection of doubtful transactions being completed by fake cardholders [10], [11]. Typically, the fraudulent transaction should be spotted before its completion to avoid fund loss. If a hacker gets access to the card information, he can easily steal the card funds. The first remedy to CCF is then to secure more card information [1]. Even though the measures to avoid the leakage of credit card details are continuously improved and implemented by e-payment providers, still, the number and amounts of fraudulent transactions are still increasing as stated in [3]. Such a fact led to the design of innovative techniques to prevent CCF using statistical and machine learning

approaches. That's by studying the data collected from previous transactions Statistical inference approaches for CCF look for outliers detection in the transactions data sets [11]. An outlier is a data point lying out of the normal range of the distribution. In CCF, an outlier is a transaction completed by a hacker that led to stolen funds. Data visualization via some charts like box plots, and probability distribution helps to the pattern of fraudulent transactions [12]. Hybrid approaches employing fuzzy logic and neural networks have been designed also to handle CCF detection [13]. [14], [15] stated that graph analysis and unsupervised clustering analyses are widely used to detect fraudulent transactions. In a wider context, graph, and network analytics can be used also to spot criminal relationships in money laundering transactions [14]. Other statistical-based approaches have been also proposed to solve the CCF detection, like the Dempster-Shafer theory and Bayesian learning, BLAST-SSAHA hybridization, Hidden Markov Model (HMM), Fuzzy Darwinian logic [11]. Machine learning constitutes the second class of approaches developed to handle the CCF. One can see that almost all machine learning techniques have been used. Depending on the problem's representation as a clustering or classification, machine learning models have been designed accordingly. Therefore, in Table 1, the techniques proposed under each class are summarised. The before-mentioned classification or clustering methods for solving the CCF are based on the use of CCF data sets. The European customers' data set [8] is the most studied data set in the CCF literature.

**Disadvantages:**

• The complexity of data: Most of the existing machine learning models must be able to accurately interpret large and complex datasets to detect and Identifying Fraudulent Credit Card Transactions.

• Data availability: Most machine learning models require large amounts of data to create accurate predictions. If data is unavailable in sufficient quantities, then model accuracy may suffer.

• Incorrect labeling: The existing machine learning models are only as accurate as the data trained using the input dataset. If the data has been incorrectly labeled, the model cannot make accurate predictions..

**PROPOSED SYSTEM** The usage of online payment technologies is continuously increasing via credit cards, digital wallets, cryptocurrencies, etc. Such a fact will increase consequently the number of fraudulent transactions and misappropriated funds. A fraudulent transaction is an instance of a failure in the security system of the service provider. The aims/contributions of the research paper are to:

1) Study the learnability level of ensemble methods from real-life and synthetic data sets.

2) Identify the source of the failure of the credit card fraud prevention systems.

3) Assess the vulnerability of the credit card transactions requests processing scripts.

4) Proposal of remedial actions to improve the process of approving/declining a transaction on credit cards.

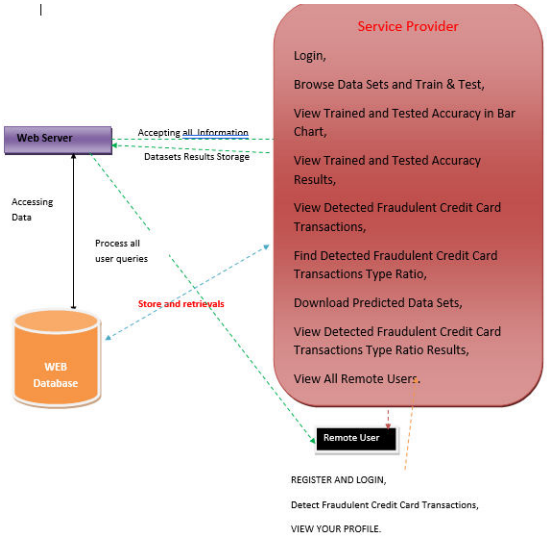
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

**Advantages :**

- The main advantage of the proposed system is to study the performance of ensemble methods on real and simulated CCF data sets. That's to analyze the outputs of business scripts implemented to process credit card transactions and how their determinism nature constitutes a vulnerability and a source of security systems failure.
- Implemented an effective models called Naive Bayes classifiers and Random forests

## IMPLEMENTATION

## SYSTEM ARCHITECTURE




## MODULES

- **SERVICE PROVIDER**

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Detected Fraudulent Credit Card Transactions, Find Detected Fraudulent Credit Card Transactions Type Ratio, Download Predicted Data Sets, View Detected Fraudulent Credit Card Transactions Type Ratio Results, View All Remote Users.

- **VIEW AND AUTHORIZE USERS**

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

- **REMOTE USER**

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, Detect Fraudulent Credit Card Transactions, VIEW YOUR PROFILE..

## RESULT


International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362


## CONCLUSION

The detection of malicious transactions on credit cards helps in avoiding a big loss of money for financial institutions. That loss is continuously increasing despite the efforts deployed by financial stakeholders. In this paper, a comparative study of ensemble methods' performance is discussed in classifying credit card transactions as authentic or malicious. It is found that XG Boost and bagging methods outperform basic classification techniques like the naive Bayes. However, they lack over fitting on real data sets. On simulated datasets, the performance of all classifiers decreases since data generation did not follow an a priori script or distribution. The high performance of ensemble methods on real data can be explained by the fact that the approval of credit card transactions by bankers follows a strict script easily discovered by transfer learning. Such a finding can be seen as a credit card transaction processing script vulnerability. In future works, multiplying, varying, and randomizing the factors should be used during the authentication phase. From a technical point of view, addressing the explain ability and interpretability of the algorithms and understanding their decision-making process will be an interesting perspective for future studies.

## REFERENCES

[1] Z. Faraji, ''A review of machine learning applications for credit card fraud detection with a case study,'' *SEISENSE J. Manage.*, vol. 5, no. 1, pp. 49–59, Feb. 2022.

[2] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, ''Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms,'' *IEEE Access*, vol. 10, pp. 39700–39715, 2022.

[3] Nilson Report. *Card Fraud Worldwide*. Accessed: May 2023. [Online]. Available: https://nilsonreport.com/

[4] N. S. Alfaiz and S. M. Fati, ''Enhanced credit card fraud detection model using machine learning,'' *Electronics*, vol. 11, no. 4, p. 662, Feb. 2022.

[5] B. Arora and Sourabh, ''A review of credit card fraud detection techniques,'' *Recent Innov. Comput.*, pp. 485–496, 2022.

[6] S. Srinidhi, K. Sowmya, and S. Karthika, ''Automatic credit fraud detection using ensemble model,'' in *ICT Analysis and Applications*. Springer, 2022, pp. 211–224.

[7] M. Sabih and D. K. Vishwakarma, ''A novel framework for detection of motion and appearance-based anomaly using ensemble learning and LSTMs,'' *Exp. Syst. Appl.*, vol. 192, Apr. 2022, Art. no. 116394.

[8] Kaggle. (2022). *European Cardholders Dataset*. Accessed: May 2023. [Online]. Available: https://www.kaggle.com/datasets/mlgulb/creditcardfraud

[9] *Sparkov Data Generation on Github*, Sparkv simulator, 2020.

[10] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic, and A. Anderla, ‘‘Credit card fraud detection–machine learning methods,’’ in *Proc. 18th Int. Symp. INFOTEH-JAHORINA (INFOTEH)*, Mar. 2019, pp. 1–5.

[11] S B. E. Raj and A A. Portia, ‘‘Analysis on credit card fraud detection methods,’’ in *Proc. Int. Conf. Comput., Commun. Electr. Technol. (ICCCET)*, 2011, pp. 152–156.

[12] C. Phua, V. Lee, K. Smith, and R. Gayler, ‘‘A comprehensive survey of data mining-based fraud detection research,’’ 2010, *arXiv:1009.6119*

[13] T. Razooqi, P. Khurana, K. Raahemifar, and A. Abhari, ‘‘Credit card fraud detection using fuzzy logic and neural network,’’ in *Proc. 19th Commun. Netw. Symp.*, 2016, pp. 1–5.

[14] M. E. Lokanan, ‘‘Financial fraud detection: The use of visualization techniques in credit card fraud and money laundering domains,’’ *J. Money Laundering Control*, vol. 26, no. 3, pp. 436–444, Apr. 2023.

[15] B. Lebichot, F. Braun, O. Caelen, and M. Saerens, ‘‘A graph-based, semisupervised, credit card fraud detection system,’’ in *Proc. 5th Int. Workshop Complex Netw. Appl. (COMPLEX Network)*. Springer, 2017, pp. 721–733.