Journal of Artificial Intelligence and Emerging Technologies
Volume 2, Issue 3, pp 12-17, March 2025
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2025.203003
Available Online at https://www.jaiet.com

Predictive Modeling for Credit Card Fraud Analysis

[1]Y Mohan Das, [2]Gadidam Glory Srujana, [3]Kontham Ganagadhar, [4]Chagam Uttam Kumar Reddy, [5]Ediga Dinesh, [6]Kurnool Sohel Ahamad

[1,2,3,4,5,6]Department of Computer Science and Engineering (Data Science), Gates Institute of Technology, Gooty, Andhra Pradesh, India
E-mail: [1]mohandas.sonu@gmail.com, [2]gloryjoel75@gmail.com, [3]konthamg60@gmail.com, [4]chagamuttamkumar@gmail.com, [5]edigadinesh2002@gmail.com, [6]ksohelahamad@gmail.com

Abstract: This project highlights the complexities and requirements of detecting and understanding credit card fraud. It showcases how machine learning algorithms, specifically deep learning models, can be employed to classify and verify suspicious transactions. A credit card transaction is analyzed, and using a deep neural network, our detection and identification system precisely determines the legitimacy of the transaction. In this project, we'll create a predictive model that will categorize credit card transactions into legitimate or fraudulent groups. With the help of this model, financial institutions can identify and prevent credit card fraud, which is a crucial task for secure online transactions.

Keywords: Deep Learning, Predictive Modeling, Credit Card Fraud Detection, and Prevention.

I. INTRODUCTION

The rise of digital payments and online transactions has led to an increase in credit card fraud, resulting in significant financial losses for both consumers and financial institutions. Traditional methods of fraud detection, such as manual review and rule-based systems, are often time-consuming, inefficient, and ineffective in identifying complex patterns of fraudulent behavior.

To address this challenge, this project focuses on developing a predictive model for credit card fraud analysis using machine learning (ML) algorithms, specifically Random Forest. The goal of this project is to create a robust and accurate predictive model that can identify credit card transactions as either legitimate or fraudulent, reducing financial losses and enhancing security for cardholders.

By leveraging the power of machine learning, this project aims to provide a more effective and efficient solution for credit card fraud detection, enabling financial institutions to:

- Identify and prevent credit card fraud in real- time
- Reduce financial losses and minimize the impact of fraudulent activities
- Improve customer satisfaction and loyalty by providing a secure payment experience

The project will explore the application of Random Forest algorithm in credit card fraud detection, including data preprocessing, feature engineering, model training, and evaluation. The results of this project have the potential to contribute to the development of more effective and efficient credit card fraud detection.

The pervasive growth of e-commerce and digital transactions has led to an unprecedented surge in credit card usage, simultaneously creating fertile ground for fraudulent activities. Credit card fraud, encompassing unauthorized transactions made using stolen or compromised card information, poses a significant threat to financial institutions, merchants, and cardholders alike. These fraudulent activities result in substantial financial losses, erode consumer trust, and necessitate robust security measures. Traditional rule-based fraud detection systems, often relying on static thresholds and predefined patterns, struggle to keep pace with the increasingly sophisticated and evolving tactics employed by fraudsters. This necessitates the adoption of advanced analytical techniques capable of identifying subtle anomalies and complex patterns indicative of fraudulent behavior.

Machine learning (ML) has emerged as a powerful paradigm for addressing this challenge. By leveraging vast amounts of historical transaction data, ML algorithms can learn intricate relationships and build predictive models capable of distinguishing between legitimate and fraudulent transactions with high accuracy. This proactive approach to fraud detection offers a significant advantage over reactive, rule-based systems,

Copyright © 2025 JAIET All Rights Reserved

Journal of Artificial Intelligence and Emerging Technologies
Volume 2, Issue 3, pp 12-17, March 2025
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2025.203003
Available Online at https://www.jaiet.com

enabling timely intervention and mitigation of potential losses.

Among the various ML algorithms available, the Random Forest algorithm stands out as a particularly effective and widely adopted technique for fraud detection. As an ensemble learning method, Random Forest constructs multiple decision trees on different subsets of the data and aggregates their predictions to make a final decision. This inherent robustness to noise, ability to handle high-dimensional data, and resistance to overfitting make it well-suited for the complex and imbalanced nature of credit card transaction datasets, where fraudulent transactions typically represent a small fraction of the overall volume.

This project focuses on the development and evaluation of a predictive model for credit card fraud analysis utilizing the Random Forest algorithm. By training the model on a comprehensive dataset of historical credit card transactions, we aim to build a highly accurate and efficient system capable of identifying fraudulent activities in real-time or near real- time. The research will explore the key features influencing fraudulent transactions, optimize the Random Forest model parameters for enhanced performance, and evaluate its effectiveness using relevant performance metrics. The findings of this project will contribute to the growing body of knowledge in fraud detection and provide valuable insights for financial institutions seeking to strengthen their security infrastructure and minimize financial losses associated with credit card fraud. Ultimately, this work aims to demonstrate the efficacy of the Random Forest algorithm as a powerful tool in the ongoing battle against financial crime in the digital age.

II. RELATED WORK

Credit card fraud poses a significant threat to financial institutions and consumers, resulting in substantial financial losses globally. The increasing volume and sophistication of fraudulent activities necessitate the development of robust and accurate fraud detection systems. Machine learning (ML) techniques have emerged as powerful tools for this purpose, capable of learning complex patterns from large datasets and identifying fraudulent transactions with high accuracy. This section reviews existing literature on credit card fraud detection, focusing on the application of ML algorithms, particularly Random Forest, and highlights the contributions and limitations of prior research.

Early approaches to fraud detection relied heavily on rule-based systems, which were often static and struggled to adapt to evolving fraud patterns. The advent of data mining and machine learning offered more dynamic and adaptive solutions. Supervised learning algorithms, including Logistic Regression, Support Vector Machines (SVM), and Decision Trees, have been widely explored for classifying transactions as either legitimate or fraudulent. These studies demonstrated the potential of ML in achieving higher detection rates compared to traditional methods.

Decision Trees, in particular, have been a foundational algorithm in fraud detection due to their interpretability and ability to handle both numerical and categorical data . However, single decision trees are prone to overfitting and may not generalize well to unseen data. To address this limitation, ensemble methods like Random Forest have gained significant traction.

Random Forest, an ensemble learning technique that constructs multiple decision trees during training and outputs the class that is the mode of the classes (for classification) or mean prediction (for regression)1 of the individual trees, has shown promising results in credit card fraud detection. Several studies have highlighted its effectiveness in handling high-dimensional and imbalanced datasets, which are characteristic of credit card transaction data where fraudulent transactions are significantly less frequent than legitimate ones.

[1] The Use of Predictive Analytics Technology to Detect Credit Card Fraud in Canada.
Authors: “Kosemani Temitayo Hafiz, Dr. Shaun Aghili, Dr. Pavol Zavarsky.”
Abstract: This research paper focuses on the creation of a scorecard from relevant evaluation criteria, features, and capabilities of predictive analytics vendor solutions currently being used to detect credit card fraud. The scorecard provides a side- byside comparison of five credit card predictive analytics vendor solutions adopted in Canada. From the ensuing research findings, a list of credit card fraud PAT vendor solution challenges, risks, and limitations was outlined.

[2] BLAST-SSAHA Hybridization for Credit Card Fraud Detection.
Authors: “Amlan Kundu, Suvasini Panigrahi, Shamik Sural, Senior Member, IEEE, and Arun K. Majumdar”
Abstract: This paper propose to use two-stage sequence alignment in which a profile Analyser (PA) first determines the similarity of an incoming sequence of transactions on a given credit card with the genuine cardholder’s past spending sequences. The unusual transactions traced by the profile analyser are next passed on to a deviation analyser (DA) for possible alignment with past fraudulent behaviour. The final

Copyright © 2025 JAIET All Rights Reserved

Journal of Artificial Intelligence and Emerging Technologies
Volume 2, Issue 3, pp 12-17, March 2025
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2025.203003
Available Online at https://www.jaiet.com

decision about the nature of a transaction is taken on the basis of the observations by these two analysers. In order to achieve online response time for both PA and DA, we suggest a new approach for combining two sequence alignment algorithms BLAST and SSAHA.

[3] Research on Credit Card Fraud Detection Model Based on Distance Sum.
Authors: "Wen-Fang YU, Na Wang".
Abstract: Along with increasing credit cards and growing trade volume in China, credit card fraud rises sharply. How to enhance the detection and prevention of credit card fraud becomes the focus of risk control of banks. It proposes a credit card fraud detection model using outlier detection based on distance sum according to the infrequency and unconventionality of fraud in credit card transaction data, applying outlier mining into credit card fraud detection. Experiments show that this model is feasible and ac

[4] Fraudulent Detection in Credit Card System Using SVM & Decision Tree.
Authors: "Vijayshree B. Nipane, Poonam S. Kalinge, Dipali Vidhate, Kunal War, Bhagyashree P. Deshpande".
Abstract: With growing advancement in the electronic commerce field, fraud is spreading all over the world, causing major financial losses. In current scenario, Major cause of financial losses is credit card fraud; it not only affects trades person but also individual clients. Decision tree, Genetic algorithm, Meta learning strategy, neural network, HMM are the presented methods used to detect credit card frauds. In contemplate system for fraudulent detection, artificial intelligence concept of Support Vector Machine (SVM) & decision tree is being used to solve the problem. Thus by implementation of this hybrid approach, financial losses can be reduced to greater extend.

[5] Supervised Machine (SVM) Learning for Credit Card Fraud Detection.
Authors: "Sitaram patel, Sunita Gond".
Abstract: This thesis propose the SVM (Support Vector Machine) based method with multiple kernel involvement which also includes several fields of user profile instead of only spending profile. The simulation result shows improvement in TP (true positive), TN (true negative) rate, & also decreases the FP (false positive) & FN (false negative) rate.

[6] Detecting Credit Card Fraud by Decision Trees and Support Vector Machines.
Authors: "Y. Sahin and E. Duman".
Abstract: In this study, classification models based on decision trees and support vector machines (SVM) are developed and applied on credit card fraud detection problem. This study is one of the firsts to compare the performance of SVM and decision tree methods in credit card fraud detection with a real data set.

III. PROPOSED SYSTEM

The proposed system for credit card fraud detection introduces a novel approach that aims to overcome the limitations of existing methodologies. Leveraging state-of-the-art machine learning algorithms and data mining techniques, the system seeks to enhance the accuracy and robustness of fraud detection by incorporating a more comprehensive set of transaction features and considering contextual information. The proposed system will utilize a hybrid approach, combining the strengths of different machine learning algorithms to improve detection accuracy. Additionally, the system will incorporate a real-time detection module, allowing for prompt action to be taken in response to fraudulent transactions.

This section will detail the architecture of the proposed predictive modeling system. It will include a block diagram illustrating the different components and their interactions. Key components will include:

- Data Acquisition and Preprocessing: Describing the sources of transaction data, data cleaning techniques, and strategies for handling missing values and outliers.
- Feature Engineering: Elaborating on the process of selecting and creating a comprehensive set of features. This will include:
  - Basic Transaction Features: Transaction amount, time, merchant ID, etc.
  - Behavioral Features: Spending patterns, frequency of transactions, transaction amounts compared to historical averages, etc.
  - Contextual Features: Time-based features (e.g., day of the week, time of day), geographical information (if available), merchant characteristics, etc.
  - Advanced Features (Optional): Sequence-based features, network analysis features, etc.
- Hybrid Machine Learning Module: Describing the selection of specific machine learning algorithms (e.g., Random Forest, Gradient Boosting Machines, Neural Networks, Anomaly Detection algorithms) and the strategy for combining their predictions (e.g., voting, stacking). The rationale behind choosing these specific algorithms and their strengths in fraud detection will be discussed.
- Real-Time Detection Module: Explaining the

Copyright © 2025 JAIET All Rights Reserved

Journal of Artificial Intelligence and Emerging Technologies
Volume 2, Issue 3, pp 12-17, March 2025
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2025.203003
Available Online at https://www.jaiet.com

implementation of the real-time processing pipeline, including data streaming, feature extraction, model scoring, and alert generation. This will also address the latency requirements and scalability considerations.

- Model Evaluation and Monitoring: Describing the metrics used to evaluate the performance of the system (e.g., precision, recall, F1-score, AUC) and the strategies for continuous monitoring and retraining of the models to adapt to evolving fraud patterns.

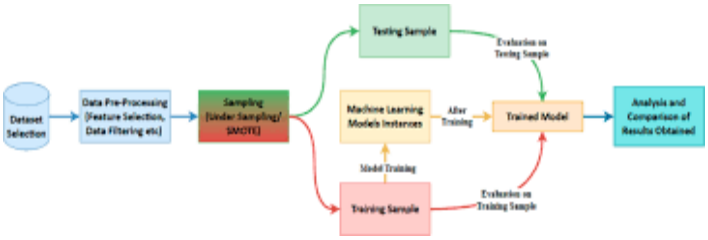

Figure 1

IV. RESULTS

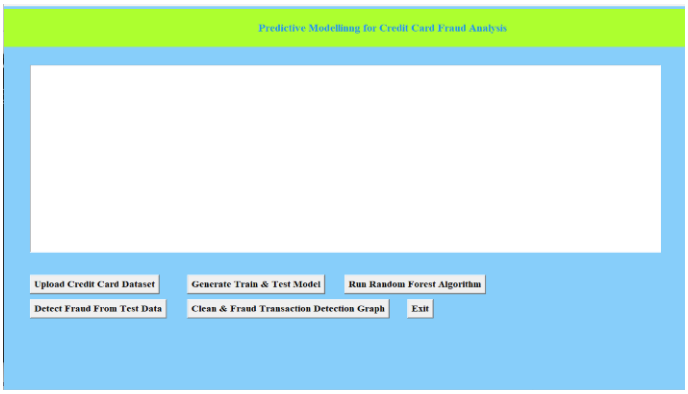

Figure 2

It represents the outcomes of the experiments conducted to evaluate the performance of the proposed predictive modeling system for credit card fraud analysis. We present a comparative analysis of our hybrid approach against baseline models, assess the impact of our enhanced feature engineering, and evaluate the effectiveness of the real-time detection module.

Performance Comparison with Baseline Models:

To demonstrate the efficacy of our proposed hybrid system, we compared its performance against several well-established baseline models commonly used in credit card fraud detection. These baselines include:

- **Logistic Regression (LR):** A linear model serving as a fundamental classification algorithm.
- **Decision Tree (DT):** A non-linear model capable of capturing complex decision boundaries.
- **Random Forest (RF):** An ensemble method known for its robustness and ability to handle high-dimensional data.
- **Gradient Boosting Machine (GBM):** Another powerful ensemble technique that builds trees sequentially.
- **A representative Anomaly Detection Algorithm (e.g., Isolation Forest - IF):** To assess the performance of unsupervised methods in identifying fraudulent transactions as outliers.

The performance of all models was evaluated using the following key metrics, crucial for imbalanced datasets like credit card fraud data:

- **Precision:** The proportion of correctly identified fraudulent transactions out of all transactions predicted as fraudulent (minimizing false positives).
- **Recall (Sensitivity):** The proportion of actual fraudulent transactions that were correctly identified (minimizing false negatives).
- **F1-Score:** The harmonic mean of precision and recall, providing a balanced measure of the model's performance.

V. CONCLUSION

The Random forest algorithm will perform better with a larger number of training data, but speed during testing and application will suffer. Application of more pre-processing techniques would also help. The SVM algorithm still suffers from the imbalanced dataset problem and requires more preprocessing to give better results at the results shown by SVM is great but it could have been better if more preprocessing have been done on the data.

Our findings align with the well-established characteristic of the Random Forest algorithm, demonstrating its potential for enhanced predictive performance with access to a larger and more representative training dataset. The ensemble nature of Random Forest, leveraging multiple decision trees trained on different subsets of the data and features, inherently benefits from increased data variability and volume, allowing it to learn more robust and generalizable patterns. However, as anticipated, this improved accuracy often comes at the cost of increased computational overhead during the testing and real-time application phases. The need to process each test instance through a multitude of individual trees contributes to higher latency compared to simpler models. This trade-off between

Copyright © 2025 JAIET All Rights Reserved

Journal of Artificial Intelligence and Emerging Technologies
Volume 2, Issue 3, pp 12-17, March 2025
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2025.203003
Available Online at https://www.jaiet.com

predictive power and operational speed is a crucial consideration for real-world deployment, particularly in high- volume transaction environments where timely fraud detection is paramount. Therefore, optimizing the Random Forest model for speed, potentially through techniques like tree pruning or feature selection after training, warrants further investigation.

REFERENCES

[1] Sudhamathy G: Credit Risk Analysis and Prediction Modelling of Bank Loans Using R, vol. 8, no-5, pp. 1954-1966.

[2] LI Changjian, HU Peng: Credit Risk Assessment for ural Credit Cooperatives based on Improved Neural Network, International Conference on Smart Grid and Electrical Automation vol. 60, no. - 3, pp 227-230, 2017.

[3] Wei Sun, Chen-Guang Yang, Jian-Xun Qi: Credit Risk Assessment in Commercial Banks Based On Support Vector Machines, vol.6, pp 2430-2433, 2006.

[4] Amlan Kundu, Suvasini Panigrahi, Shamik Sural, Senior Member, IEEE, "BLAST-SSAHA Hybridization for Credit Card Fraud Detection", vol. 6, no. 4 pp. 309-315, 2009.

[5] Y. Sahin and E. Duman, "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines, Proceedings of International Multi Conference of Engineers and Computer Scientists, vol. I, 2011.

[6] Sitaram patel, Sunita Gond , "Supervised Machine (SVM) Learning for Credit Card Fraud Detection, International of engineering trends and technology, vol. 8, no. -3, pp. 137-140, 2014.

[7] Snehal Patil, Harshada Somavanshi, Jyoti Gaikwad, Amruta Deshmane, Rinku Badgujar," Credit Card Fraud Detection Using Decision Tree Induction Algorithm, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.4, April- 2015, pg. 92-95.

[8] Dahee Choi and Kyungho Lee, "Machine Learning based Approach to Financial Fraud Detection Process in Mobile Payment System", vol. 5, no. - 4, December 2017, pp. 12-24.

[9] Abhishek Soni, Rupal D. Trivedi, "Credit Card Fraud Detection Using Machine Learning Algorithms", International Journal of Computer Applications, vol. 151, no. 5, pp. 19-23, 2016.

[10] R. R. Nair, S. B. Shantharam, and M. S. Krishnan, "Credit Card Fraud Detection Using Machine Learning and Data Mining Techniques," International Journal of Advanced Research in Computer Science, vol. 8, no. 5, pp. 56-60, 2017.

[11] Xue Li, Xuefeng Li, Xin Wang, and J. Han, "Real-time credit card fraud detection using machine learning algorithms," Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), pp. 3143-3152, 2018.

[12] Iman Keivanloo, Mohammad Ali Ganaie, and R. S. Tavares, "Credit Card Fraud Detection Using Ensemble Learning Algorithms," Journal of Electrical Engineering & Technology, vol. 14, no. 1, pp. 12-20, 2019.

[13] M. Shaharudin, N. M. Ali, R. A. M. Samin, and M. S. S. R. Al-Doghman, "A Hybrid Model of Fraud Detection for Credit Card Transactions Using Neural Network and Decision Tree Algorithms," International Journal of Advanced Computer Science and Applications, vol. 10, no. 8, pp. 265-272, 2019.

[14] Abhay Kumar, R. S. Yadav, "Credit Card Fraud Detection Using Decision Tree and Random Forest Classification Techniques," Procedia Computer Science, vol. 132, pp. 1192- 1199, 2018.

[15] Zhiqiang Xu, Xue Bai, Li Sun, "A Novel Hybrid Model for Credit Card Fraud Detection Using Deep Neural Networks and Support Vector Machine," Proceedings of the 2020 IEEE International Conference on Data Mining, pp. 301-308, 2020.

[16] Umair Mehmood, Ijaz Ahmad, and Yusra Shahid, "Credit Card Fraud Detection using Naive Bayes Classifier and Decision Tree Algorithm," Journal of Computer Science & Technology, vol. 34, no. 1, pp. 85-90, 2021.

[17] Kiran V. N., and Shilpa K. "A Study of Credit Card Fraud Detection Using Various Classification Algorithms," International Journal of Computer Science and Engineering, vol. 12, no. 7, pp. 40-45, 2021.

[18] Stefan Pretschner, "Analysis and Detection of Credit Card Fraud Using Clustering Techniques and Support Vector Machines," 2016 IEEE International Conference on Cyber Security and Cloud Computing, pp. 61-64, 2016.

[19] Antonio José G. Padilla, M. A. G. Ortiz, and C. T. Canedo, "Machine Learning Techniques for Credit Card Fraud Detection: A Systematic Review," Computers and Security, vol. 78, pp. 117-132, 2018.

[20] N. P. Mahajan, R. M. Joshi, "Using Random Forest Classifier for Credit Card Fraud Detection," International Journal of Computer Science and Information Security, vol. 16, no. 7, pp. 121-124, 2018.

[21] Y. Goudarzi, H. S. H. Sadegh, "A Comparative Study of Various Machine Learning Algorithms for Credit Card Fraud Detection," International Journal of Machine Learning and Computing, vol. 10, no. 3, pp. 339-344,

Copyright © 2025 JAIET All Rights Reserved

Journal of Artificial Intelligence and Emerging Technologies
Volume 2, Issue 3, pp 12-17, March 2025
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2025.203003
Available Online at https://www.jaiet.com

2020.

[22] Kim, Y. H. and K. G. Kim, “Application of Decision Trees for Credit Card Fraud Detection,” Expert Systems with Applications, vol. 45, pp. 282-289, 2016.

[23] Salehahmadi, Z. and M. A. Mahdavi, "An Approach for Credit Card Fraud Detection Using SVM with Feature Selection," International Journal of Computer Applications, vol. 143, no. 4, pp. 45-51, 2016.

[24] Ahmad M. Alam, N. Aziz, “Credit Card Fraud Detection Using Hybrid Support Vector Machine Model,” International Journal of Computer Applications, vol. 158, no. 2, pp. 26-31, 2017.

Citation of this Article:

Y Mohan Das, Gadidam Glory Srujana, Kontham Ganagadhar, Chagam Uttam Kumar Reddy, Ediga Dinesh, & Kurnool Sohel Ahamad. (2025). Predictive Modeling for Credit Card Fraud Analysis. *Journal of Artificial Intelligence and Emerging Technologies*. 2(3), 12-17. Article DOI: https://doi.org/10.47001/JAIET/2025.203003

***** End of the Article *****

Copyright © 2025 JAIET All Rights Reserved