e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704008 |

Fraudulent Transaction Detection System using Random Forest Classifier Algorithm

R.Arthi[1], Santhosh K [2]

[1]Assistant Professor, PG& Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu India

[2]UG Student, PG& Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu India

ABSTRACT: This paper investigates the effectiveness of a Random Forest classifier in detecting fraudulent transactions within a simulated dataset. The dataset is designed to reflect real-world scenarios, encompassing characteristics of both legitimate and fraudulent transactions. These characteristics include customer behavior (consistent email addresses, IP addresses, device types), transaction details (amount, location), and indicators of potential fraud (unusual location, high amount, multiple CVV attempts). The paper explores techniques to generate a balanced dataset with a controllable percentage of fraudulent transactions. This allows for evaluating the model's performance under different fraud prevalence conditions. The Random Forest classifier is chosen for its ability to handle complex relationships within the data and its robustness to Overfitting. The paper assesses the model's performance using metrics like precision, recall, and F1-score. These metrics evaluate the model's ability to correctly identify fraudulent transactions while minimizing false positives (legitimate transactions flagged as fraud). The results will provide insights into the effectiveness of the Random Forest classifier for fraud detection and highlight potential areas for further optimization.

KEYWORDS: Dall-E, audio, emotion intelligence, Transformers, hyper-parameters

I. INTRODUCTION

Fraudulent transactions are a major concern for financial institutions and online businesses. Detecting these transactions early and accurately is essential to minimize financial losses and maintain customer trust. This project investigates the effectiveness of a machine learning approach, specifically a Random Forest classifier, in identifying fraudulent transactions within a simulated dataset. The simulated dataset is designed to represent real-world scenarios by incorporating characteristics of both legitimate and fraudulent transactions. This includes details like customer behavior (consistent email addresses, IP addresses, and device types), transaction details (amount, location), and indicators of potential fraud (unusual location compared to billing address, high transaction amount, multiple CVV attempts).

This paper aims to assess the effectiveness of a Random Forest classifier in identifying fraudulent transactions amidst legitimate transactions. We will evaluate the model's performance using metrics like precision, recall, and F1-score. These metrics will provide insights into the model's ability to accurately detect fraud while minimizing false positives (legitimate transactions flagged as fraud). The findings of this project will contribute to the ongoing effort to develop robust and efficient fraud detection systems. By exploring the capabilities of Random Forest classifiers, this project can provide valuable insights for improving fraud detection accuracy and protecting financial systems from fraudulent activities.

In today's digital age, online transactions have become increasingly prevalent, offering convenience and efficiency for consumers and businesses alike. However, with the rise of online transactions comes the heightened risk of fraudulent activities, posing significant challenges for financial institutions, businesses, and consumers. The problem statement for the "Fraud Detection for Online Transactions for Improved Reliability" project revolves around addressing these challenges and developing a robust system to detect and prevent fraudulent activities in online transactions.

Online transactions encompass a wide range of activities, including e-commerce purchases, online banking, digital payments, and more. While these transactions offer numerous benefits, they also present opportunities for malicious actors to engage in fraudulent behaviors such as identity theft, credit card fraud, account takeover, and payment fraud. Fraudulent activities not only result in financial losses for businesses and consumers but also undermine trust in online platforms and financial systems.

II. RELATED WORKS

The existing system often includes predefined rules or filters that flag transactions based on specific criteria.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704008 |

For example, transactions above a certain monetary threshold, transactions from suspicious geographic locations, or transactions involving high-risk products or services may trigger alerts. Behavior analysis involves monitoring and analyzing patterns of user behavior to identify anomalies. This may include deviations from typical spending patterns, unusual transaction times or frequencies, or unexpected changes in user account activity. The system may perform checks on IP addresses and Geolocation data associated with transactions to identify potentially fraudulent activity. Transactions originating from known high-risk IP addresses or locations may be flagged for further investigation. Suspicious transactions flagged by the automated system may undergo manual
review by fraud analysts. Analysts assess additional contextual information, transaction details, and user behavior to determine the legitimacy of flagged transactions.

III. PROPOSED METHODOLOGY

The core functionality relies on a pre-trained Random Forest model for fraud detection. However, some offline data processing steps are involved:

Data Pre-processing: The paper assumes a historical dataset containing transaction information is available for model training. This stage involves cleaning and preparing the data, including:

- Handling missing values and outliers
- Encoding categorical variables
- Feature engineering (creating new features from existing ones)

Model Training: The preprocessed data is used to train the Random Forest model. This involves training the model to identify patterns that differentiate fraudulent and legitimate transactions.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704008 |

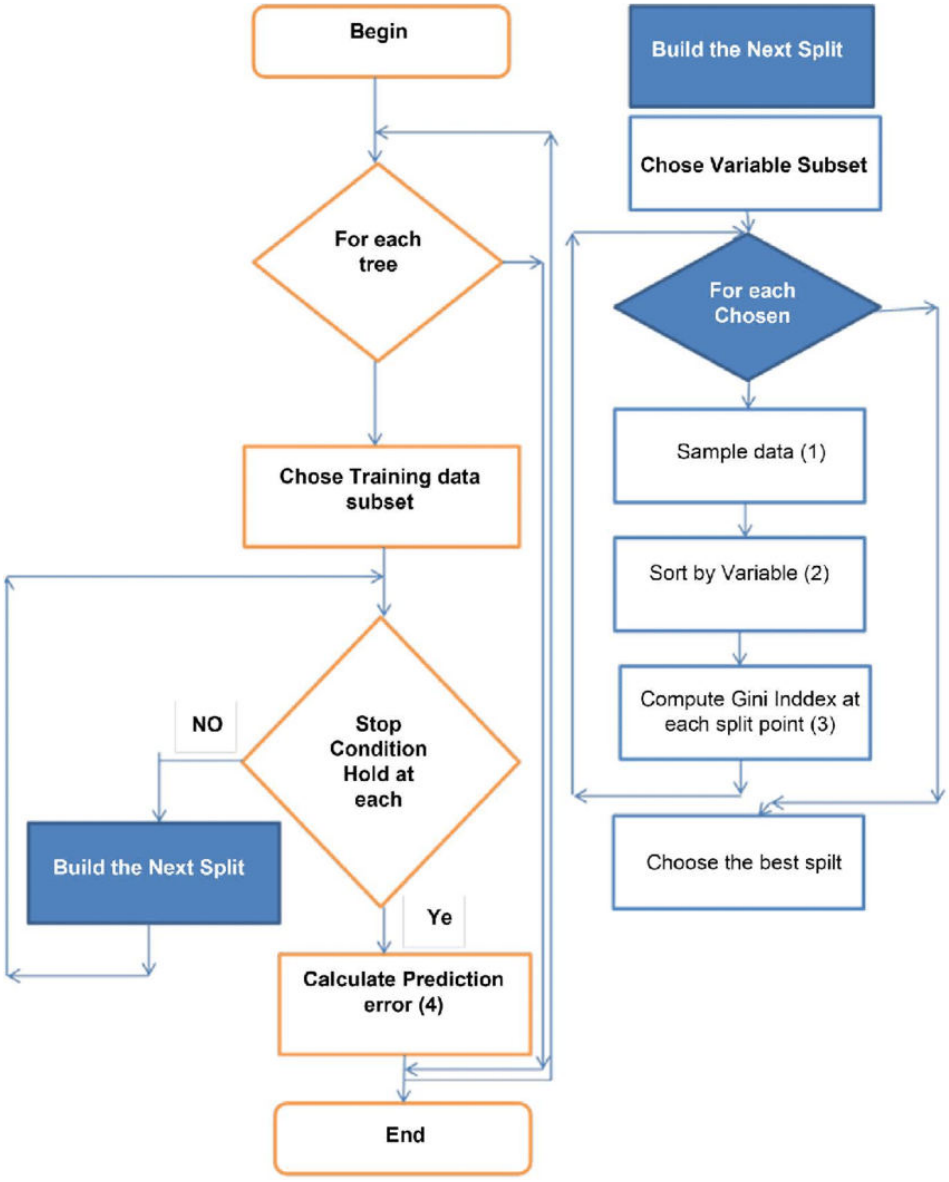

Figure – 1 System Architecture

Model Evaluation: The trained model's performance is evaluated using metrics like precision, recall, and F1-score. This assessment helps gauge the model's effectiveness in detecting fraud. The pre-trained Random Forest model is saved and deployed as part of the Flask API. This methodology leverages a pre-trained model for efficient real-time fraud analysis through a user-friendly API interface. The Flutter web application provides a convenient platform for user interaction.

The fraud detection system utilizes a Random Forest classifier as the machine learning model. It is trained on historical transaction data to identify patterns indicative of fraudulent activities. The model is trained using historical transaction data collected from diverse sources. The training data includes features such as transaction amount, device type, location, and customer details. Once trained, the Random Forest classifier model is deployed on the Flask server for real-time inference. The model accepts transaction data as input and outputs the probability of fraudulence for each transaction.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704008 |

IV. RESULT & DISCUSSION

The primary metric for evaluating the effectiveness of the fraud detection system is its accuracy in correctly identifying fraudulent and non-fraudulent transactions. The accuracy of the machine learning model, trained using a Random Forest classifier is assessed based on its ability to classify transactions accurately. The results demonstrate the accuracy achieved by the system in detecting fraudulent transactions, thereby reducing the risk of financial losses for businesses and users.
Performance metrics such as precision, recall, and F1-score are calculated to assess the overall performance of the fraud detection system.

Precision: Precision measures the ratio of correctly identified fraudulent transactions to the total number of transactions flagged as fraudulent. A higher precision indicates fewer false positives.
Recall: Recall, also known as sensitivity, measures the ratio of correctly identified fraudulent transactions to the total number of actual fraudulent transactions. A higher recall indicates fewer false negatives.
F1-Score: The F1-score, which is the harmonic mean of precision and recall, provides a balanced measure of the model's performance.

Overall, the results and discussions presented in this section underscore the importance of leveraging advanced technologies, such as machine learning and data analytics, to develop sophisticated fraud detection systems capable of safeguarding online transactions and mitigating financial risks. The insights gained from the evaluation of the system's performance provide valuable guidance for further optimization and refinement, ultimately leading to enhanced security and reliability in online transaction environments. Additionally, the discussion emphasizes the need for continuous monitoring, evaluation, and improvement of fraud detection systems to adapt to evolving threats and ensure ongoing effectiveness in combating financial fraud.

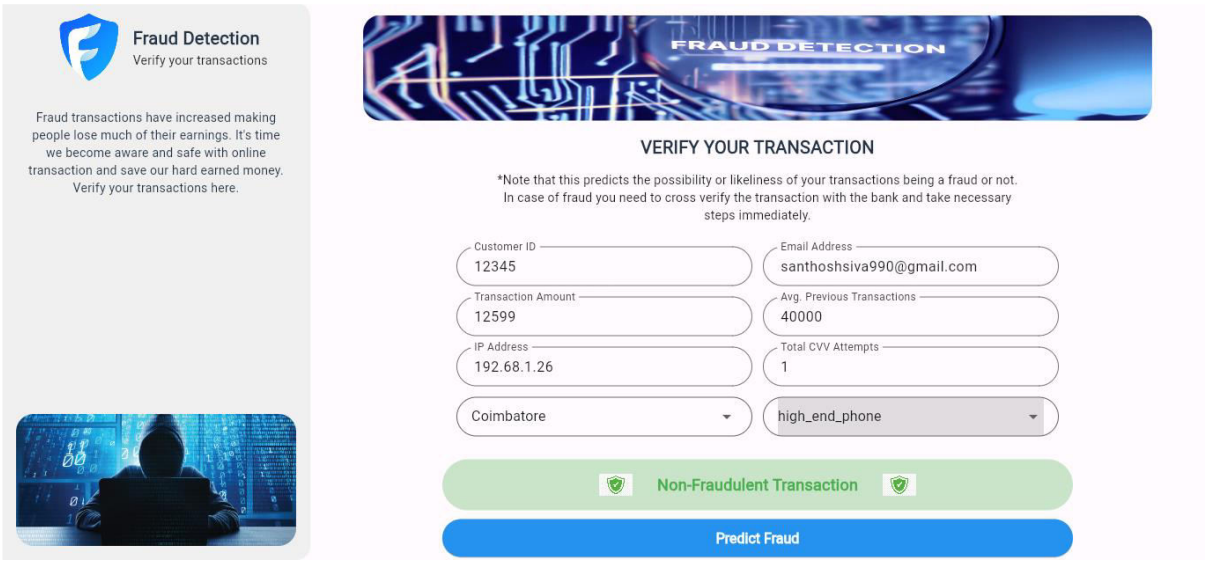

Figure- 2 Non-fraudulent Transaction results

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704008 |

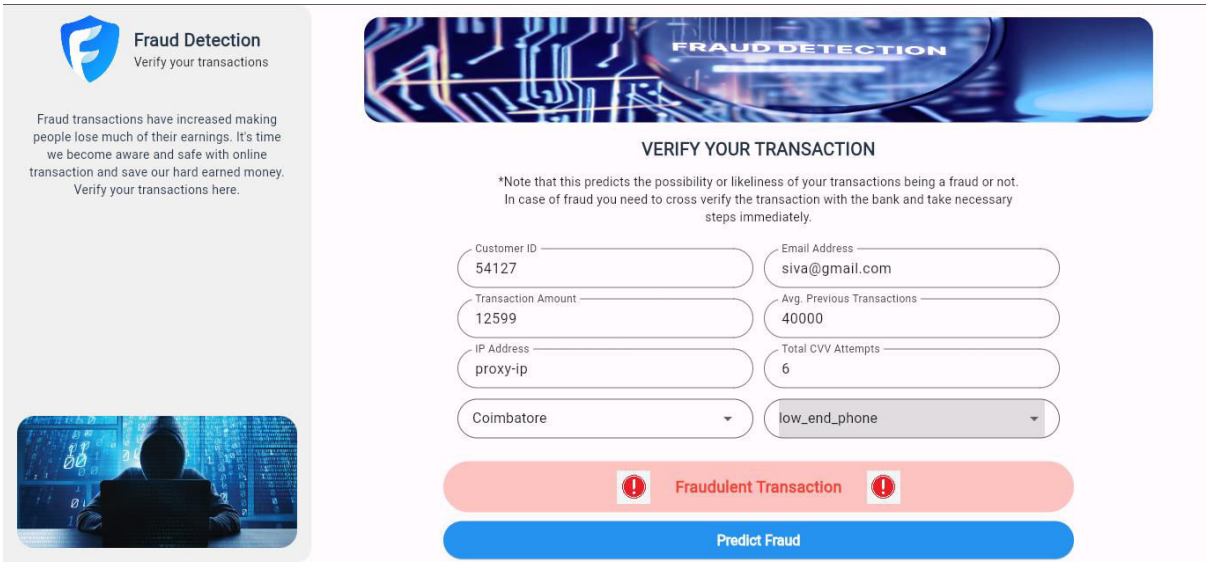

Figure- 2 fraudulent Transaction results

IV. CONCLUSION

The fraud detection project has developed a robust system leveraging the Random Forest classifier for real-time identification and prevention of fraudulent transactions in online transactions. The Random Forest classifier demonstrates high precision, recall, and F1-score metrics, showcasing its effectiveness in detecting fraudulent activities. In Conclusion, the fraud detection project represents a significant advancement in combating financial fraud in online transactions. Continued refinement and advancement of fraud detection techniques are essential to ensure the continued safety and integrity of online transactions globally. Leveraging machine learning techniques is crucial for maintaining trust, security, and reliability in the digital landscape.

REFERENCES

[1] S. Vimala, K.C. Sharmili, ―Prediction of Loan Risk using NB and Support Vector Machine‖, International Conference on Advancements in Computing Technologies (ICACT 2018), vol. 4, no. 2, pp. 110-113, 2018.
[2] X. Francis Jency, V.P.Sumathi, Janani Shiva Sri, ―An Exploratory Data Analysis for Loan Prediction Based on Nature of the Clients‖, International Journal of Recent Technology and Engineering (IJRTE), Vol. 7, No. 48, pp. 176-179, 2018.
[3] Anchal Goyal, Ranpreet Kaur, ―Loan Prediction Using Ensemble Technique‖, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 3, pp. 523 – 526, March 2016.
[4] Aboobyda Jafar Hamid and Tarig Mohammed Ahmed, ―Developing Prediction Model of Loan Risk in Banks using Data Mining‖, Machine Learning andApplications: An International Journal (MLAIJ), Vol.3, No.1, pp. 1-9, March 2016.
[5] Aditi Kacheria, Nidhi Shivakumar, Shreya Sawkar, Archana Gupta,Loan Sanctioning Prediction System, International Journal of Soft Computing and Engineering (IJSCE), vol. 6, no. 4, pp. 50-53, 2016.
[6] Anchal Goyal, Ranpreet Kaur, ―Accuracy Prediction for Loan Risk using Machine Learning Models‖, International Journal of Computer Science Trends and Technology (I JCST), Vol. 4, Issue 1, pp. 52-57, Jan- Feb 2016.
[7] Aboobyda Jafar Hamid and Tarig Mohammed Ahmed, ―Developing Prediction Model of Loan Risk in Banks using Data Mining‖, Machine Learning and Applications: An International Journal (MLAIJ), Vol.3, No.1, pp. 1-9, March 2016.
[8] S. Vimala, K.C. Sharmili, ―Prediction of Loan Risk using NB and Support Vector Machine‖, International Conference on Advancements in Computing Technologies (ICACT 2018), vol. 4, no. 2, pp. 110-113, 2018.
[9]Aditi Kacheria, Nidhi Shivakumar, Shreya Sawkar, Archana Gupta, ―Loan Sanctioning Prediction System‖, International Journal of Soft Computing and Engineering (IJSCE), vol. 6, no. 4, pp. 50-53, 2016.
[10] E.Chandra Blessie, K R Vineetha, An Effectual GA based Association Rule Generation and Fuzzy SVM Classification algorithm for Predicting students performance, International Journal of Engineering and Advanced Technology (IJEAT), Volume 8, Year 2019, Pages 2915-2920

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 4, April 2024 |

| DOI:10.15680/IJMRSET.2024.0704008 |

[11] X. Francis Jency, V.P.Sumathi, Janani Shiva Sri, ―An Exploratory Data Analysis for Loan Prediction Based on Nature of the Clients‖, International Journal of Recent Technology and Engineering (IJRTE), Vol. 7, No. 48, pp. 176-179, 2018.
[12]Vaidya, "Predictive andprobabilistic approach using logistic regression: Application to prediction of loan approval,"2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, 2017, pp. 1-6.doi: 10.1109/ICCCNT.2017.8203946
[13] Praneesh, M. "An Analysis of Gaussian Kernel Density Estimation for Feature Selection of Gene Expression." Excel International Journal of Technology, Engineering and Management, 2 (2), 34 40 (2015).
[14] Xin-She Yang and Suash Deb. Cuckoo search via levy flights. In ´ 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), pages 210–214. IEEE, 2009.
[15] Xin-She Yang and Suash Deb. Cuckoo search: recent advances and applications. Neural Computing and Applications, 24(1):169–174, 2014.
[16] E.Chandra Blessie, E.Karthikeyan, Roc Curve Assessment on Cancer Diagnostic Performance by a Classifier, , European Journal of Scientific Research, Volume 95, Year 2013, Pages 400-407
[17] J. Lohokare, R. Dani and S. Sontakke, "Automated data collection for credit score calculation based on financial transactions and social media,"2017 International Conference on Emerging Trends & Innovation in ICT (ICEI), Pune, 2017, pp. 134-138.doi: 10.1109/ETIICT.2017.7977024
[18]M. Bayraktar, M. S. Aktaş, O. Kalıpsız, O. Susuz and S. Bayracı, "Credit risk analysis with classification Restricted Boltzmann Machine,"2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4.doi: 10.1109/SIU.2018.8404397
[19] E.Chandra Blessie, K.R.Vineetha, Probability based Student Performance Prediction using Naive Bayesian Algorithm, , American International Journal of Research in Science, Technology, Engineering & Mathematics (AIJRSTEM), Volume , Year 2018, Pages 82-86
[20] S. Yadav andS. Thakur, "Bank loan analysis using customer usage data: A big data approach using Hadoop,"2017 2nd International Conference on Telecommunication and Networks (TEL-NET), Noida, 2017, pp. 1-8.doi: 10.1109/TEL-NET.2017.8343582
[21] X. Y. Zhang, "Click prediction for P2P loan ads based on support vector machine," *Journal of Physics*, vol. 1168, no. 3, Article ID 032042, 2019.
[22]M. Liu, M. Qu, and B. Zhao, "Research and citation analysis of data mining technology based on Bayes algorithm," *Mobile Networks and Applications*, vol. 22, no. 3, pp. 418–426, 2017. [23]C. X. Ling, J. Huang, and H. Zhang, "AUC: a better measure than accuracy in comparing learning algorithms," *Advances in Artificial Intelligence*, vol. 2671, pp. 329–341, 2003

[24]X. Ye, L.-A. Dong, and D. Ma, "Loan evaluation in P2P lending based on random forest optimized by genetic algorithm with profit score," *Electronic Commerce Research and Applications*, vol. 32, pp. 23–36, 2018.

[25]J. Luan, C. L. Zhanga, B. D. Xu et al., "The predictive performances of random forest models with limited sample size and different species traits," *Fisheries Research*, vol. 227, Article ID 105534, 2020.

[26]F. Lv, J. H. Huang, W. Wang et al., "A two-route CNN model for bank account classification with heterogeneous data," *PLoS One*, vol. 14, no. 8, Article ID 0220631, 2019.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com