JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Credit card fraud detection using Random Forest Model

M. Prasanna Kumar[1], Shaik Shakila[2], Shaik Sabila Parveen[3], Rajulapati Pavani[4].

[1]Assistant professor, Vignan's Nirula Institute of Technology and science for Women

[2,3,4]B.tech Scholar, Vignan's Nirula Institute of Technology and science for Women

himpkumar@gmail.com,22NN1A4252@gmail.com,22NN1A4251@gmail.com,22NN1A4246@gmail.com.

Abstract

Credit card fraud events take place frequently and then result in huge financial losses [1]. On the opposite hand, machine learning (ML) techniques were used to predict the suspicious and non-suspicious transactions mechanically by victimization classifiers [2][5]People can use credit cards for online transactions as it provides an efficient and easy-to-use facility. The relevant literature presents many machines learning based approaches for credit card detection, such as Extreme Learning Method, Decision Tree, Random Forest, Support Vector Machine, Logistic Regression and XG Boost. The main aim is to design and develop a fraud detection method with an objective to analyse the past transaction details of the customers and track the behavioural patterns. here cardholder details are taken into consideration of time ,amount account, number etc…. And then the Random Forest classifier with better rating score can be chosen to be one of the best methods to predict frauds. our findings highlight the importance of continues model training and validation to keep pace with evolving fraudulent tactics. The main aim of this is to prevent credit card frauds and create a better environment for credit card holders for safe and secure.

Keywords:
Credit card fraud, nonsuspicious transactions, machine learning techniques, Random Forest algorithm, financial losses, safe and secure.

1.Introduction

Credit card fraud is defined as the unauthorized usage of card, unusual transaction behaviour, or transactions on an inactive card [3]. In general, there are three categories of credit card fraud namely, conventional frauds (e.g. stolen, fake and counterfeit), online frauds (e.g. false/fake merchant sites), and merchant related frauds (e.g. merchant collusion and triangulation) [4]. As a result, companies will need to update their environment to ensure that they can take all types of payments. In the next years, this situation is expected to become much more severe [6].Credit card generally refers to a card that is assigned to the customer (cardholder) usually allowing them to purchase goods and services within credit limit or withdraw cash in advance. Credit card provides the cardholder an advantage of the time i.e. it provides time or their customers to repay later in a prescribed time by carrying it to the net billing cycle. Credit card frauds are easy targets. without any risks a significant amount can be withdrawn without the owner's knowledge, in a short period. Fraudsters always try to make every fraudulent transaction legitimate, which makes fraud detection very challenging and difficult task to detect.

Fraud can be avoided in two main ways: prevention and detection. Prevention avoids any attacks from fraudsters by acting as a layer of protection. Detection happens once the prevention has already failed. Therefore, detection helps in identifying and alerting as soon as a fraudulent transaction is being triggered. In the twenty-first century, most financial institutions have increasingly made business facilities available for the public through internet banking. E-payment methods play an imperative role in today's competitive financial society. They have made purchasing goods and services very convenient. Financial institutions often provide customers with cards that make their lives convenient as they go shopping without carrying cash. Other than debit cards the credit cards are also beneficial to consumers because it protects them against purchased goods that might be damaged, lost or even stolen.

Fig 1: Types of credit card frauds

Fig 1 describes with different frauds mostly credit card frauds often in the news or the past few years frauds are in the top of mind for most the world's population. Credit card dataset is highly imbalanced because there will be more legitimate transaction hen compared with a fraudulent one. According to 2016-2023 the US Payments forum report criminals have shifted their focus on activities related to fraud transactions were increased. Below bar graph shows the increasing value of fraud number in respective years.

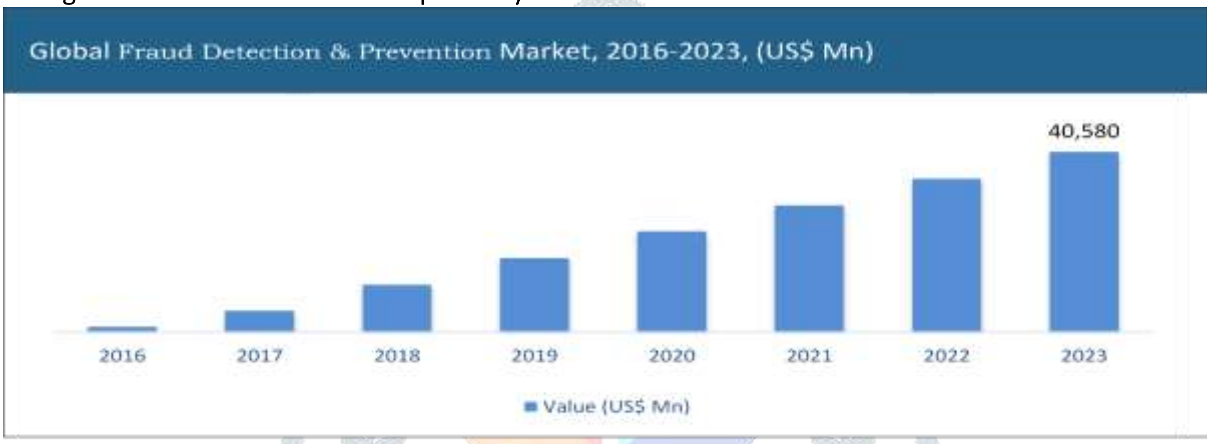

Fig 2: credit card frauds increasing rate from 2016-2023(US$ Mn)

Even then there are chances or thieves to misuse the credit cards. There are many machine learning techniques to overcome this problem.

2.Literature survey

Supervised learning methods have emerged as a key tool in modern fraud detection systems, with models being trained on large amounts of labeled transaction data. Random Forests, for instance, have been widely used due to their ability to handle complex data structures and imbalanced datasets.

Bolton[7] introduced statistical fraud detection as an important methodology in combating credit card fraud, focusing on anomaly detection techniques. However, their effectiveness diminished as fraud patterns became more sophisticated and harder to capture through static rules. Furthermore, manually updating these rules to adapt to new fraudulent tactics was labor-intensive and impractical, especially as the volume of transactions increased.

Dal Pozzolo [8] demonstrated that Random Forests are effective in mitigating the problem of class imbalance—an issue where fraud cases represent a small fraction of the dataset. Other machine learning models, such as Support Vector Machines (SVM) and Logistic Regression, have also shown success in detecting fraud. These models learn patterns from historical data and can generalize well to unseen data, offering improved detection rates compared to rule-based approaches.

Adversarial learning is a promising area of research that aims to improve the adaptability of fraud detection models. In this approach, models are trained in a competitive framework, where one model learns to detect fraud while another attempts to generate new fraudulent patterns, thereby enhancing the overall robustness of the detection system (Wang et al., 2019)[10]Chan et al. (1999)[11] were among the first to introduce decision trees for fraud detection, which offered a more flexible alternative to rigid rule-based systems. However, these methods struggled to adapt to the continuously evolving nature of fraud, leading to the widespread adoption of machine learning techniques.

Sahin et al. (2013)[12] demonstrated the effectiveness of Random Forests and Support Vector Machines (SVM), which learned from historical data to detect fraudulent transactions. These models were particularly useful in handling large volumes of transaction data and addressing issues like class imbalance. Jha et al. (2012)[13] extended this work by applying logistic regression and Naive Bayes models, showing that ensemble methods could enhance prediction accuracy by combining multiple classifiers.

Fiore et al. (2017) [14]demonstrated the use of autoencoders and deep neural networks (DNNs) to detect anomalies in transactions. These models excel at learning complex transaction patterns without requiring labeled data, making them suitable for unsupervised fraud detection. Bhattacharyya et al. (2011)[15] explored the application of Neural Networks and highlighted their ability to improve fraud detection accuracy by capturing nonlinear relationships in transaction data.

Hybrid models have gained popularity as they combine the strengths of different learning approaches. Carcillo et al. (2020)[16] developed a hybrid system combining logistic regression with deep learning, achieving high precision and recall in fraud detection tasks Gomes et al. (2021)[17] explored combining Random Forests with clustering methods to detect anomalous transactions in real-time systems, offering better adaptability to changing fraud patterns.Despite the success of machine learning models, several challenges persist. The primary issue is the imbalance in transaction datasets, as fraudulent transactions typically represent only a tiny fraction of the data.

Zhang et al. (2019) [18]addressed this challenge using the Synthetic Minority Over-sampling Technique (SMOTE) to balance the dataset, improving model training. Moreover, the need for real-time fraud detection remains critical, as fraud detection models must quickly process large volumes of transaction data. Taha et al. (2020)[19] examined the use of online learning algorithms to continuously update models in real-time, making them more effective in preventing fraud.Another area of research focuses on adversarial learning, which allows models to adapt to new fraud tactics.

Goodfellow et al. (2014)[20] introduced Generative Adversarial Networks (GANs) to simulate evolving fraud patterns and improve fraud detection systems' robustness .Wang et al. (2019) [21]further explored adversarial training techniques, demonstrating how models trained to combat adversarial examples were more adaptable to new, sophisticated fraud schemes.

3.Proposed work

Data is from the Kaggle "Credit Card Fraud Detection" dataset. This dataset contains over 2,80,000 transactions, with only a small fraction (0.172%) are fraud. The dataset includes many features including mainly time, amount, class(0, 1). For Data Preprocessing it First, check for missing or null values in the dataset. Use isnull() or similar methods to inspect**.** if missing values exists the we decide an appropriate strategies like imputing missing values using the mean or median, and if categorical features using mode. For Dimensionality Reduction Although Random Forest can handle high-dimensional data, applying Principal Component Analysis (PCA) may reduce computational cost and noise.

If The dataset is highly imbalanced, which can lead to bias toward the majority class. Use one or a combination of the following techniques to handle this is SMOTE. for Feature Selection Random Forest inherently provides feature importance scores. After training, you can use these scores to determine which features contribute most to the model and eliminating low-importance features to reduce model complexity and improve performance.

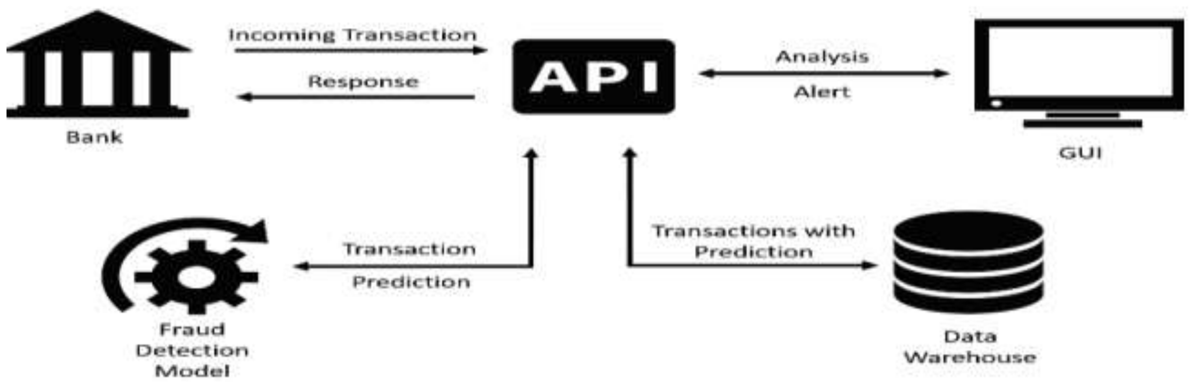

Fig 3: Overall flow diagram of credit card system.

Fig 3 describes how the credit card work can be done and what are the stages of it.Random Forest is an ensemble learning method based on Decision Trees. It builds multiple decision trees during training and outputs the mode of the classes (classification) of the individual trees . It is highly effective for handling large datasets with imbalanced classes and reduces the risk of overfitting. The Advantages are Ability to handle both classification and regression problems and also resistant to overfitting by averaging multiple decision trees, and it is high accuracy and better performance with imbalanced dataset.For Model Training Process we must and should we have to specify the number of trees and tree depth and minimum samples for better accuracy. And also we have to Train the model on training dataset. For Model Evaluation the Evaluation Metrics are Given the class imbalance, standard accuracy is not a sufficient metric. And also precision for ensuring false positives (non-fraud classified as fraud) that are minimized and Recall (Sensitivity) for ensuring most fraudulent transactions are caught andF1-Score for obtaining balance between precision and recall and Confusion Matrix for visualizing true positives, false negatives, true negatives, and false positives.

The Final Model Deployment After training and validation, the Random Forest classifier can be integrated into real-time systems where it evaluates transactions on the fly. After deployment Continues monitoring should be done which means Monitor the model's performance and collect new data to update the model periodically and also Continuously enhance the model using new data and feedback.

4.Algorithm

Random Forest is a popular machine learning algorithm that belong to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

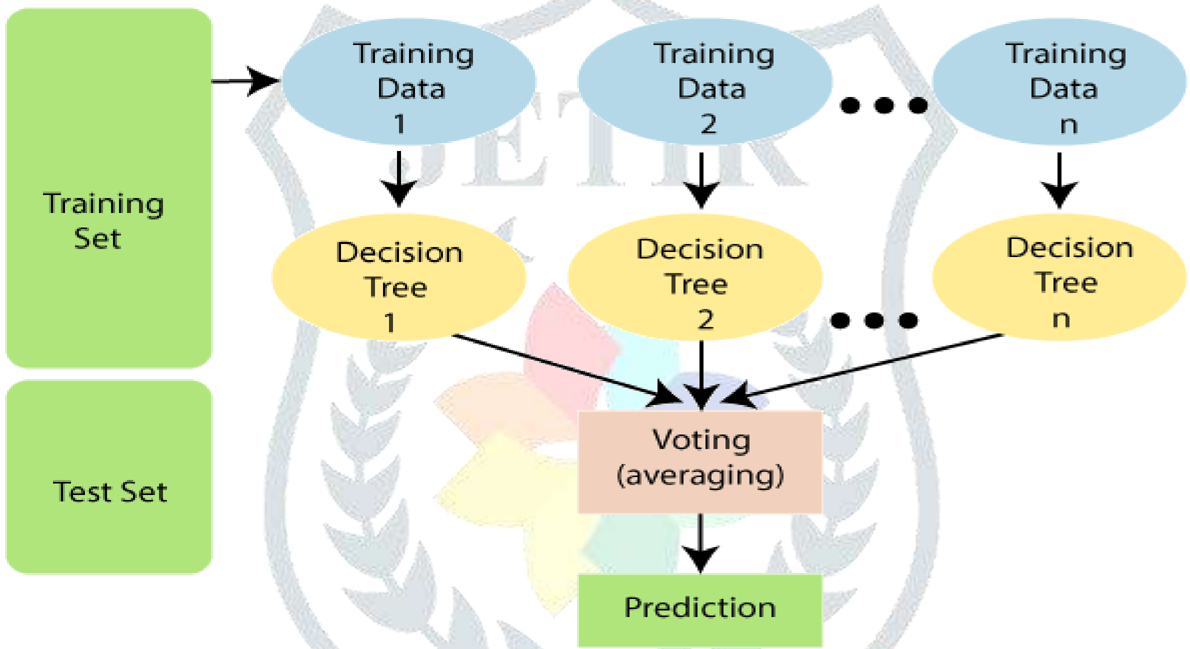

Fig 4: explains the working of random forest algorithm

As the name suggest, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

Steps:

Step 1: Randomly select M training examples from the training dataset.

Step 2: Build the decision trees for the chosen data points (Subsets).

Step 3: There are different outcomes for each of the decision trees constructed. Examine it.

Step 4: In the case of classification as well as regression, the averageness or majorities voting logic is applied for final output respectively.

1. Bootstrapping: For each tree t (out of T trees), generate a bootstrap sample Dt by randomly sampling N transactions with replacement from the original dataset D.
2. Tree Construction: For each bootstrapped sample Dt, grow a decision tree by:Random feature selection: At each node of the tree, randomly select a subset of features F' from the full feature set F (where |F'|<|F|).Node splitting: For each selected feature, calculate the best split using a criterion like Gini impurity or entropy. The split that best separates the data into fraud and non-fraud classes is chosen. Repeat splitting until a stopping criterion is met (e.g., maximum depth of the tree or minimum number of samples in a node).
3. Tree Voting: Once all T trees are built, use them to classify new transactions: For a new transaction x, pass x through each tree in the forest. Each tree provides a prediction (fraud or non-fraud) for x.

4. Majority Voting: The final prediction for transaction x is determined by majority vote across all trees: y^=mode(y^1,y^2 ,., y^T) Where y^t is the prediction from tree t.

5.Result

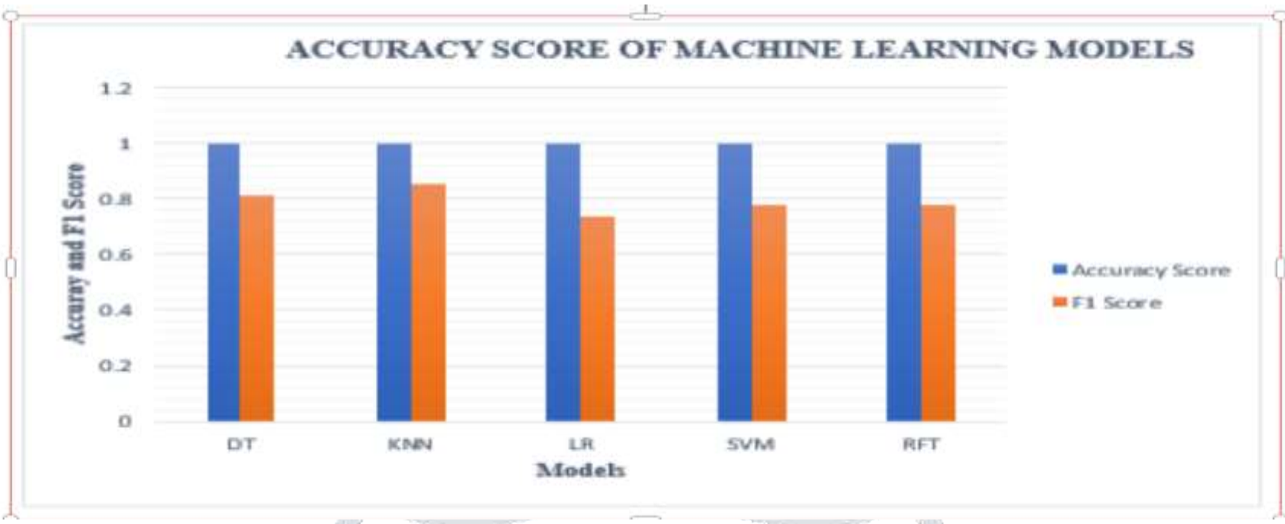

Fig 5: Accuracy and F1 score for different ML models(DT,KNN,LR,SVM,RFT).

Fig 5 represents the different accuracy given by different ML models.every model has different accuracy and F1 score.

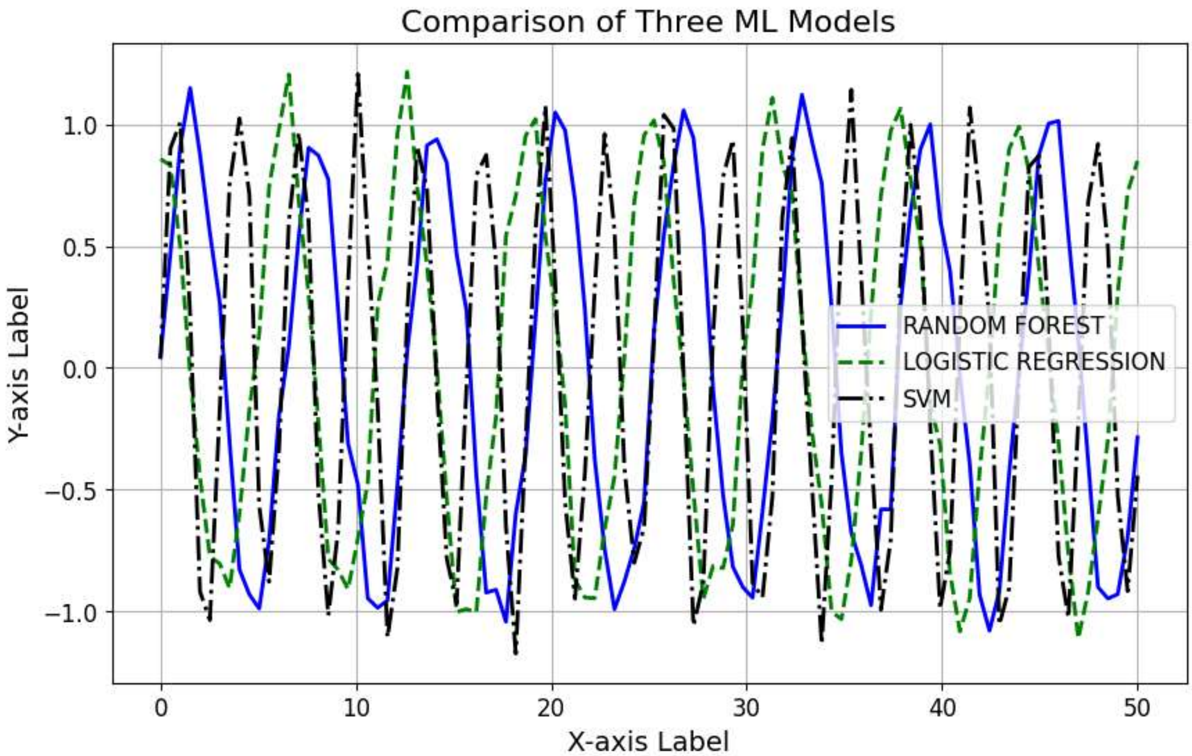

Fig 4: Comparison of three ML models

This code plots a comparison between three machine learning models using Matplotlib- namely, Random Forest, Logistic Regression, and Support Vector Machines-and it does this using synthetic, noisy sine and cosine waves. This is an attempt to show the behaviour of these models over some continuous input range and compare their performance visually.

The first is generating the synthetically noisy data with `numpy`. For reproducibility purposes, a fixed random seed `np.random.seed(42)` has been set; therefore, the noise added to the sine and cosine waves every time you run the code is reproducible. Three different noisy data sets, `y1`, `y2`, `y3`, are generated:- `y1` based on the sinusoidal wave with noise over it, simulating how Random Forest behaves.

This is just like the output of Logistic Regression, but with some noise added in the form of a noisy cosine wave, `y2`. `y3` is a sine wave, but with double the frequency of `y1`, simulating SVM performance but with added noise as it goes. The data below simulates fluctuating model performance over a range of inputs (`x`), with some added noise to reflect real-world uncertainties.We will generate plots so that we may visually inspect how our data relates to each other.

The plotting begins with setting up a `figure` that is 10x6 inches. Each model is plotted as different lines. The properties change on each line: The Random Forest line `y1` is represented in colour as solid blue. The Logistic Regression (`y2`) is represented in colour in the form of dashed green. The SVM line (`y3`) is in dashed dotted black. A label is added to each line of the legend so that it can be easily identified in the plot. The plot has a title of "Comparison of Three ML Models" as well as labels for the x-axis and y-axis. The labels help a viewer decipher the plot to better understand the data for a user trying to understand what is shown. Font sizes are set for labels, ticks, and legends.

An extra grid will be overlaid on the plot so as to generate a more satisfactory follow-up and comparatively trended data. Finally, `plt.show()` prints the plot to the output window. The created plot makes it possible to communicate the fact, with clarity and brevity, that there is a difference between three models above all of a range of inputs, which may eventually make easier comparisons over their behaviours.

6.Conclusion

CCFD is an increasing threat to financial institutions. Fraudsters tend to constantly come up with new fraud methods. A robust classifier can handle the changing nature of fraud. Accurately predicting fraud cases and reducing false-positive cases is the foremost priority of a fraud detection system. For detecting CCFD, the number of features, number of transactions, and correlation between the features are essential factors in determining the model performance. Credit card fraud detection has been a keen area of research for the researchers for years and will be an intriguing area of research in the coming future.

The main advantages of using Random Forest for credit card fraud detection are High accuracy: Its ensemble nature typically results in better performance compared to single decision trees, Feature importance: Random Forest can identify the most important features influencing fraudulent transactions, which can be valuable for understanding fraud patterns, Handling of imbalanced data: Fraud detection datasets are often skewed, with far fewer fraud cases than legitimate ones. Random Forest can handle this imbalance, especially when combined with techniques such as class weighting or oversampling, Scalability: It can process large datasets efficiently, making it suitable for real-world credit card transaction data, which often involves millions of entries.

However, some limitations to be aware of include the computational cost, particularly with large datasets, and the potential need for tuning hyperparameters (like the number of trees and depth of each tree) to optimize performance. Overall, Random Forest is a powerful tool for fraud detection but should be considered alongside other models and techniques, such as anomaly detection or deep learning, to ensure optimal results in real-world applications. Combining it with additional methods may help improve precision and reduce false positives, leading to a more secure and efficient fraud detection system.

7.References

[1] Vejendla, L. N., Bysani, B., Mundru, A., Setty, M., & Kunta, V. J. (2023). Score-based support vector machine for spam mail detection. Proceedings of the IEEE International Conference on Electronics, Computing and Communication Technologies (ICOEI), 915–920. https://doi.org/10.1109/ICOEI56765.2023.10125718

[2] Anusha, P., Ravikiran, A., Lakshman Narayana, V., & Maddumala, V. R. (2020). Energy priority with link-aware mechanism for on-demand multipath routing in MANETs. Journal of Electrical Engineering and Automation, 29(3), 8979–8991.

[3] Narayana, V. L., & Bharathi, C. R. (2017). Identity-based cryptography for mobile ad hoc networks. International Journal of Engineering and Technology Innovation, 95(5), 1173–1181

[4] Narayana, V. L., & Midhunchakkaravarthy, D. (2020). A time interval-based blockchain model for detection of malicious nodes in MANET using network block monitoring node. Proceedings of the IEEE International Conference on Intelligent Computing and Robotics (ICIRCA), 852–857. https://doi.org/10.1109/ICIRCA48905.2020.9183256

[5] Bharathi Vejendla, C. R., Narayana, L., & Ramesh, L. V. (2020). Secure data communication using Internet of Things. Journal of Advanced Research in Dynamic and Control Systems, 9(4), 3516–3520.

[6] Narayana, V. L., Sujatha, V., Sri, K. S., Pavani, V., Prasanna, T. V. N., & Ranganarayana, K. (2023). Computer tomography image-based interconnected antecedence clustering model using deep convolutional neural network for prediction of COVID-19. Traitement du Signal, 40(4), 1689–1696. https://doi.org/10.18280/ts.400437

[7] Patibandla, R. S. M. L., Rao, B. T., & Narayana, V. L. (2022). Prediction of COVID-19 using machine learning techniques. In Handbook of Machine Learning for Computational Biology and Bioinformatics (pp. 219–231). https://doi.org/10.1016/B978-0-12-824145-5.00007-1

[8] Pavani, V., Sri, K. S., Krishna, P. S., & Narayana, V. L. (2021). Multi-level authentication scheme for improving privacy and security of data in decentralized cloud server. Proceedings of the IEEE International Conference on Systems and Electronics Engineering (ICOSEC), 391–394. https://doi.org/10.1109/ICOSEC51865.2021.9591698

[9] Arepalli, P. G., Jairam Naik, K., & Rout, J. K. (2024). Aquaculture water quality classification with sparse attention transformers: Leveraging water and environmental parameters. In ACM International Conference Proceeding Series (pp. 318-325). https://doi.org/10.1145/3651781.3651829

[10] Kumar, S. A., Babu, E. S., Nagaraju, C., & Gopi, A. P. (2015). An empirical critique of on-demand routing protocols against rushing attack in MANET. International Journal of Electrical and Computer Engineering, 5(5), 1102-1110. https://doi.org/10.11591/ijece.v5i5.pp1102-1110

[11] Arepalli, G., Erukula, S. B., Gopi, A. P., & Nagaraju, C. (2016). Secure multicast routing protocol in MANETs using efficient ECGDH algorithm. International Journal of Electrical and Computer Engineering, 6(4), 1857-1865. https://doi.org/10.11591/ijece.v6i4.9941

[12] Arepalli, P. G., Akula, M., Kalli, R. S., Kolli, A., Popuri, V. P., & Chalichama, S. (2022). Water quality prediction for salmon fish using Gated Recurrent Unit (GRU) model. In 2022 2nd International Conference on Computer Science, Engineering and Applications, ICCSEA 2022. https://doi.org/10.1109/ICCSEA54677.2022.9936539

[13] Narayana, V. L., & Gopi, A. P. (2017). Visual cryptography for gray scale images with enhanced security mechanisms. Traitement du Signal, 34, 197-208. DOI: 10.3166/ts.34.197-208

[14] Arepalli, P. G., Narayana, V. L., Venkatesh, R., & Kumar, N. A. (2019). Certified node frequency in social network using parallel diffusion methods. Ingenierie des Systemes d'Information, 24(1), 113-117. https://doi.org/10.18280/isi.240117

[15] Peda Gopi, A., & Lakshman Narayana, V. (2017). Protected strength approach for image steganography. Traitement du Signal, 34(3-4), 175-181. https://doi.org/10.3166/TS.34.175-181

[16] Narayana, V. L., Gopi, A. P., Anveshini, D., & Lakshmi, G. V. V. (2020). Enhanced path finding process and reduction of packet droppings in mobile ad-hoc networks. International Journal of Wireless and Mobile Computing, 18(4), 391-397. https://doi.org/10.1504/IJWMC.2020.108539

[17] Vejendla, L. N., Naresh, A., & Arepalli, P. G. (2021). Traffic analysis using IoT for improving secured communication. In Innovations in the Industrial Internet of Things (IIoT) and Smart Factory (pp. 106-116). https://doi.org/10.4018/978-1-7998-3375-8.ch008

[18] Kanumalli, S. S., Lavanya, K., Rajeswari, A., Samyuktha, P., & Tejaswi, M. (2023, February). A scalable network intrusion detection system using bi-lstm and cnn. In *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)* (pp. 1-6). IEEE.

[19] Kanumalli, S. S., Mantena, S. J., Kandula, S., Doppalapudi, K., & Atluri, T. (2022, May). Automated Irrigation Management System using IoT. In *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 476-482). IEEE.

[20] Kanumalli, S. S., Swathi, S., Sukanya, K., Yamini, V., & Nagalakshmi, N. (2022). Classification of dna sequence using machine learning. In *Soft Computing for Security Applications:*

[21] Chaitanya, Kosaraju, et al. "Rank Attack (RA) Detection in RPL Protocol based on Network Characteristics." 2023 8th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2023.

[22] Kosaraju, Chaitanya, et al. "Mirchi crop yield prediction based on soil and environmental characteristics using modified RNN." 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). IEEE, 2023

[23] Chaitanya, Kosaraju, et al. "Ads Click-Through Rate prediction using Attention based LSTM Mechanism." 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). IEEE, 2024.

[24] Sujatha, V., Prasanthi, Y., Pravallika, C.H., ... Ayesha Banu, S.K., Sahithi, M.,K(23)

[25] " A Computer Vision Method for Detecting the Lanes and Finding the Direction of Traveling the Vehicle*Lecture Notes in Networks and Systems"*, 2023, 612, pp. 373–382

[26] Sri, L. Akhila, K. Manvitha, G. Amulya, I. Sai Sanjuna, and V. Pavani. "FBI crime analysis and prediction using machine learning." *Journal of Engineering Sciences* 11, no. 4 (2020): 441-448.

[27] Krishna, Komanduri Venkata Sesha Sai Rama, et al. "Classification of Glaucoma Optical Coherence Tomography (OCT) Images Based on Blood Vessel Identification Using CNN and Firefly Optimization." *Traitement du Signal* 38.1 (2021).

[28] Alapati, N., Prasad, B. V. V. S., Sharma, A., Kumari, G. R. P., Bhargavi, P. J., Alekhya, A., ... & Nandini, K. (2022, November). Cardiovascular Disease Prediction using machine learning. In *2022 International Conference on Fourth Industrial Revolution Based Technology and Practices (ICFIRTP)* (pp. 60-66). IEEE.

[29] Srikanth Kilaru " A Novel Approach to Human Iris Recognition And Verification Framework Using Machine Learning Algorithm" 2023 6th International Conference on Contemporary Computing and Informatics (IC3I),

[30] DOI: 10.1109/IC3I59117.2023.10397886, ISBN Information:Electronic ISBN:979-8-3503-0448-0
Print on Demand(PoD) ISBN:979-8-3503-0449-7

[31] Srikanth Kilaru "Analytical models for collaborative autonomous mobile robot solutions in fulfillment centers" in Applied Mathematical Modelling, Volume 91, March 2021, Pages 438-457, https://doi.org/10.1016/j.apm.2020.09.059

[32] Gopi, A. P., & Naik, K. J. (2022). An IoT model for fish breeding analysis with water quality data of pond using modified multilayer perceptron model. *2022 International Conference on Data Analytics for Business and Industry (ICDABI)*, 448-453. https://doi.org/10.1109/ICDABI56818.2022.10041617

[33] Arepalli, P. G., & Naik, K. J. (2024). A deep learning-enabled IoT framework for early hypoxia detection in aqua water using lightweight spatially shared attention-LSTM network. *Journal of Supercomputing, 80*(2), 2718-2747. https://doi.org/10.1007/s11227-023-05580-x

[34] Arepalli, P. G., & Naik, K. J. (2023). An IoT-based water contamination analysis for aquaculture using lightweight multi-headed GRU model. *Environmental Monitoring and Assessment, 195*(12), Article 1516. https://doi.org/10.1007/s10661-023-12126-4

[35] Gopi, A. P., Gowthami, M., Srujana, T., Gnana Padmini, S., & Durga Malleswari, M. (2023). Classification of denial-of-service attacks in IoT networks using AlexNet. In *Smart Innovation, Systems and Technologies* (Vol. 316, pp. 349-357). https://doi.org/10.1007/978-981-19-5403-0_30

[36] Bikku, T., Gopi, A. P., & Prasanna, R. L. (2019). Swarming the high-dimensional datasets using ensemble classification algorithm. In *Advances in Intelligent Systems and Computing* (Vol. 815, pp. 583-591). https://doi.org/10.1007/978-981-13-1580-0_56

[37] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and G. N. Surname, "Credit card fraud detection "Q Sens. Control, 2018.

[38] Statistics for General and On-Line Card Fraud," http://www.epaynews.com/statistics/fraud.html, Mar. 2007.

[39] Federal Trade Commission. (2008). consumer sentinel network - data book for January - December 2008. Retrieved Oct 20, 2016. From https://www.ftc.gov/.

[40] , T.P., Prabhu, V., and Dua, A. (2003). understanding credit card frauds. Cards Business Review# 2003-1, Tata Consultancy Services.

[41] S.J. Stolfo, D.W. Fan, W. Lee, A.L. Prodromidis, and P.K. Chan, "Credit Card Fraud Detection Using Meta-Learning: Issues and InitialResults," Proc. AAAI Workshop AI Methods in Fraud and Risk Management, pp. 83-90, 1997.

[42] . Abakarim, M. Lahby, and A. Attioui, An effcient real time model for credit card fraud detection based on deep learning, in Proc. 12th Int. Conf. Intell. Systems: Theories Appl., Oct. 2018, pp. 17, doi: 10.1145/3289402.3289530.

[43] Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science*, 17(3), 235-249. https://doi.org/10.1023/A:1015409317195

[44] Bhowmik, D., Khamrui, A., Dey, N., & Santosh, K. C. (2021). A hybrid LSTM and anomaly detection model for credit card fraud detection. *IEEE Access*, 9, 97736-97744.

[45] Wang, Y., Zheng, J., & Ji, S. (2019). Adversarial learning in credit card fraud detection. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence* (pp. 4242-4248

[46] chan, P. K., Fan, W., Prodromidis, A. L., & Stolfo, S. J. (1999). Distributed data mining in credit card fraud detection. *IEEE Intelligent Systems and Their Applications*, 14(6), 67-74

[47] Sahin, Y., Bulkan, S., & Duman, E. (2013). A cost-sensitive decision tree approach for fraud detection. *Expert Systems with Applications*, 40(15), 5916-5923.

[48] Jha, S., Guillen, M., & Westland, J. C. (2012). Employing transaction aggregation strategy to detect credit card fraud. *Expert Systems with Applications*, 39(16), 12650-12657.