International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

UPI Transaction Fraud Detection Using Machine Learning

[1] Manjunath S N, [2] Dr Geetha M

[1]Sem MCA Student, Department of MCA, BIET, Davanagere

[2]Associate Professor, Department of MCA, BIET, Davanagere

Abstract – Nowadays, people conduct practically all their business online. While there are many benefits to online transactions, like viability, speedier payments, and ease of use, there are drawbacks as well, including fraud, phishing, and data theft. The increasing volume of internet transactions raises the possibility of fraud and dishonest business practices that compromise an individual's privacy. When a criminal can take control of an account and move money out of a person's online bank account. To minimize anticipated financial losses, it is also necessary to enhance traditional machine learning techniques. A feature-engineered machine learning-based model that is Random Forest and Gradient Boosting algorithm which can improve its performance, reinforce its stability, and gain experience by processing as much as data it can. Finally, understanding the costs and risks associated with payment methods is essential to fighting fraud in a methodical and cost-effective manner. We define three models to address these issues: a risk model to forecast fraud risk while taking counter measures into account; machine learning-based fraud detection; and economic optimization of machine learning outcomes. Real data is used to test the models.

***Key Words*: Machine Learning, Random Forest, Gradient Boosting, Reinforce, Fraud Detection**

1. INTRODUCTION

The world is heading quickly toward a cashless society. The number of people making purchases online has increased, according to numerous polls and studies, and it is expected that this trend will continue in the years to come. While this might sound like good news, there is also an increase in fraudulent transactions on the other hand. Despite the deployment of multiple security measures, fraudulent transactions nonetheless result in the loss of a substantial amount of money. When someone uses another person's credit card for unauthorized personal purchases online without the cardholder's or the card issuer's knowledge, this is known as online fraud.

The process of monitoring user activity to assess, spot, or stop unwanted behaviour, such as fraud, intrusion, and defaults, is known as fraud detection. A person who has fallen for one of these scams frequently does not recognize it until it is too late. Real-world examples show how fast automated systems evaluate the enormous amount of payment requests to select which transactions to authorize. Every authorized transaction is examined, and any suspicious activity is found, using machine learning algorithms. Before chatting with the experts looking into these claims, cardholders need to determine whether a transaction was legitimate or fraudulent. The automated system that creates and refines the algorithm incorporates suggestions from the detectives to gradually increase the accuracy of fraud detection.

To progressively improve the accuracy of fraud detection, the automated system that develops and improves the algorithm considers recommendations made by the detectives. With the help of several technologies, we are trying to develop a Web application for machine learning- based fraud detection in this project. An online transaction fraud detection system is necessary to safeguard digital financial transactions and thwart unlawful or fraudulent activities. As more financial transactions are being done online, robust fraud detection techniques are becoming more and more important. The primary goal of these systems is to immediately identify and thwart fraudulent conduct in order to safeguard the integrity and security of online transactions. An online transaction fraud detection system is a set of instruments, algorithms, and processes designed to identify and prevent.

1.1 OBJECTIVE

To achieve high accuracy in fraud detection while minimizing false positives. Analyze user behavior and transaction patterns. Continuously adapt to evolving fraud tactics and patterns. Seamlessly integrate with various online payment gateways and financial institutions. Enhance user authentication methods. Ensure scalability to handle increasing transaction volumes Identify and mitigate potential risks associated with online transactions, to industry regulations and compliance standards which Build and maintain trust among users.

1.2 CHALLENGES

First off, even while transaction data contains extremely few fraudulent events, the data's imbalance may skew model performance in favour of the majority class, leading to imprecise fraud detection. Second, because fraudsters use dynamic techniques, idea drift occurs, and model adaption is

ongoing to identify changing fraud trends. Moreover, challenges exist in precisely identifying fraud indicators related to the selection and engineering of pertinent features from large and diverse datasets. Training and performance of the model are further complicated by problems with data quality, such as noisy or incomplete data. Furthermore, fraud detection systems' resilience is compromised by ML models' vulnerability to adversarial attacks, in which criminals benefit from flaws to avoid detection. Scalable infrastructures and effective algorithms are necessary for real-time processing requirements to quickly examine transactions.

1.3 RANDOM FOREST

To distinguish between authentic and fraudulent transactions, the Random Forest algorithm is utilized as a reliable and efficient instrument. To function, Random Forest builds many decision trees, each of which is trained using a different subset of the characteristics and data that are available. Together, these decision trees create an ensemble, with each tree adding to the final classification. The algorithm can recognize suspicious activity depending on multiple transaction variables such transaction amount, frequency, location, and user behavior because it has learned patterns and relationships within the transaction data during training. Random Forest's capacity to handle the imbalanced nature of transaction datasets—where fraudulent cases are frequently greatly outnumbered by genuine ones—is one of its main advantages for fraud detection.

This issue of class disparity can be successfully addressed by Random Forest by using strategies like class weighting or modifying decision criteria, which will enhance its accuracy in detecting fraudulent transactions. Moreover, high- dimensional data that is frequently encountered in fraud detection applications is a good fit for Random Forest models. Their real-time processing ability of vast amounts of transaction data makes them perfect for online fraud detection systems that need to quickly identify any suspicious activity. Furthermore, Random Forests guarantee dependable performance even in dynamic transaction contexts since they are resistant to overfitting and noise in the data. Random Forest also has the benefit of being interpretable. Through the examination of feature important scores produced during model training, fraud analysts can acquire valuable knowledge regarding the transaction attributes that most significantly affect that have the greatest impact on the identification of fraudulent activity.

1.4 GRADIENT BOOSTING

One effective and popular method for spotting fraud is the Gradient Boosting algorithm. Gradient Boosting, in contrast to conventional machine learning algorithms, creates an ensemble of decision trees successively, with every new tree of decision trees in a stepwise manner, with each new tree trying to improve on the mistakes of the one before it. By concentrating on the cases that were incorrectly classified in earlier iterations, this iterative procedure enables the algorithm to progressively increase its forecast accuracy. The methods are particularly effective in addressing imbalanced datasets, which have a large proportion of genuine transactions compared to fraudulent transactions, in the context of online transaction fraud detection. Also, it prioritizes the identification of fraudulent transactions, improving overall performance by assigning greater weight to misclassified occurrences.

These models are skilled in identifying intricate patterns and relationships in transaction information, allowing them to distinguish minute distinctions between authentic and fraudulent activity. Gradient Boosting can efficiently leverage features like transaction amount, frequency, location, and user behavior to spot suspicious activities. Furthermore, the flexibility that Gradient Boosting algorithms provide with regard to model complexity and parameter adjustment enables fraud detection systems to adjust to shifting fraud patterns and dynamic threats. This flexibility is essential in dynamic online contexts because fraud strategies are ever-changing. Moreover, Gradient Boosting offers feature important insights that help fraud analysts determine which transaction The most crucial elements for spotting fraudulent activity.

2. LITERATURE SURVEY

There is one good deal of published research on financial fraud detection; for an overview, see West and Bhattacharya, and for a thorough analysis of financial fraud detection techniques, read Hajek and Henriques. Investigations into the risk factors of financial fraud revealed that the most significant risk factor is pressure or incentive to commit fraud. The following general categories can be used to group related studies based on the type of financial fraud: The four primary fraud channels—physical, web, telephonic, and mobile—were also identified by account takeover fraud, payment fraud, and application fraud. [1]

The explanation in a study on the data and technique- oriented aspects of master card fraud detection, master card took a very important rule out of today's economic landscape. It is a pleasant neighborhood for families, businesses, and international endeavors. When used sensibly and properly, credit cards can be very beneficial; nevertheless, fraudulent activity can also seriously damage credit and cause financial loss. Several strategies are put forth to stop the growth of master card fraud.[2]

Virjanand, Rajkishan Bharti, Shubham Chauhan, and Suraj Pratap proposed, the several methods for identifying fraud in online transactions are enumerated in this review paper. To effectively handle the issues raised by fraud detection and prevention, it provides knowledge of numerous research articles in the field of online transaction fraud detection. To detect fraudulent online transactions, future research may

combine machine learning algorithms with various input and output consideration configurations.[3]

The most popular delivery strategies are balanced ones. Three categories comprise the most popular suggested solutions: synthesis, algorithm, and data layer solutions. Class imbalance may suffer if preprocessing is shortened as a resample to apply data-level solutions. The algorithmic level solution is to modify the learning biases of the current algorithms or develop new ones for the minority class.[4]

One of the key issues in the application domain has been identified as the lack of real-world datasets for the data utilized in prior studies. As a result, most past studies sought to create synthetic data simulations using features taken from actual fraud and valid transactions. Riekeet al. did this by taking payment laundering patterns from actual occurrences.
Still, since

early investigations (Coppolino et al., Rieke et al.) showed relatively low false negative rates, the number of cases was insufficient for effective fraud detection. The introduction of the PaySim financial simulator, which mimics typical mobile transactions and introduces fraudulent behavior to increase the frequency of financial frauds, is a significant advancement.[5]

3. EXISTING SYSTEM

Many fraud detection systems incorporate predefined rules and thresholds based on known patterns of fraudulent behavior. These rules can be set to flag transactions that deviate from normal parameters, such as unusually large transactions, multiple transactions within a short time frame, or transactions from high-risk geographic locations. Statistical analysis is often used to identify patterns and trends within transaction data. Deviations from established patterns may indicate potential fraud. For example, the system may look for anomalies in transaction amounts, frequency, or location.

One of the most important aspects of fraud detection is analyzing user behavior. For every user, the system creates a baseline of typical behavior based on variables including transaction history, spending habits, and the typical locations of transactions. When this baseline is departed from, notifications for additional research may be triggered. Systems for detecting fraud run in real time to keep an eye on transactions as they take place. Fraudulent acts must be stopped as soon as possible to avoid financial damage. Financial institutions such as banks and credit card firms frequently work together with fraud detection technologies.

4. PROPOSED SYSTEM

The recommended method for identifying online transaction fraud simply uses two powerful machine learning algorithms: random forest and gradient boosting.

These algorithms were chosen because they are good at handling the complexity of transaction data and spotting patterns that indicate fraud. The system's meticulous approach starts with the collection and preparation of transaction data, which prepares the ground for model training. After that, the dataset is split into training, validation, and testing sets to help in the creation and evaluation of the model. The random forest and gradient boosting models are trained using the training set. Two techniques are utilized to optimize the hyperparameters: grid search and randomized search along with cross- validation.

AUC, F1-score, recall, accuracy, precision, and ROC-AUC are among the metrics used to assess the models' performance on the validation set. To make sure the model can generalize, the top-performing model is then chosen for additional testing on the testing set that hasn't been seen yet. Following validation, the model is implemented in the online transaction system, where it is immediately integrated into the pipeline for real-time transaction classification and processing. In reaction to emerging fraud tendencies and distribution adjustments, the model's performance needs to be regularly assessed, extended, and retrained using new data. Throughout the development process, ethics and data protection rules are strictly adhered to, and interpretability techniques are employed to inspire trust and facilitate stakeholders' comprehension of the model.

5. SYSTEM ARCHITECTURE

Several interconnected components that cooperate to identify and stop fraudulent activity usually make up the system architecture utilized in online transaction fraud detection. The data ingestion and processing components, which are important to the architecture, oversee gathering transaction data from an assortment of a variety of sources, including online platforms, payment gateways, and financial institutions. To guarantee the quality and consistency of the data, these components preprocess and clean it by carrying out operations such data normalization, deduplication, and enrichment. After the data is ready, it is put into the fraud using machine learning methods as a detecting engine to evaluate transaction patterns and spot suspect activity. These algorithms include Random Forests, Gradient Boosting, and Neural Networks. Various algorithms are employed to differentiate between authentic and fraudulent transactions based on factors including transaction value, frequency, location, and user behavior. To apply particular fraud detection techniques or compliance needs, the fraud detection engine may additionally integrate rule-based systems.

6. DATASETS

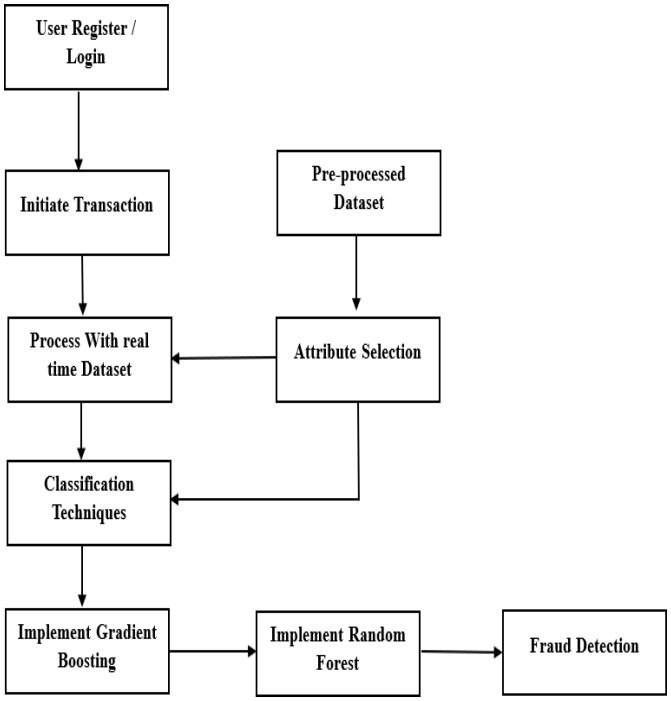
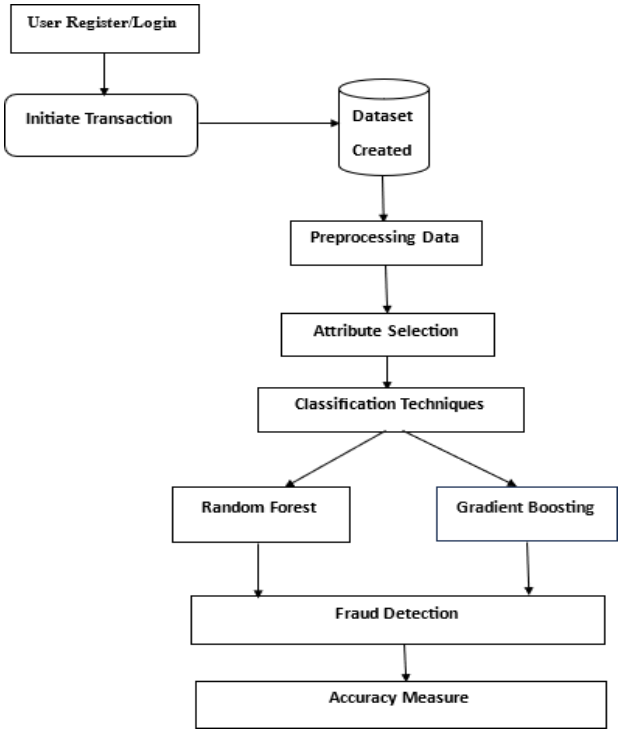


Fig-1: System Architecture

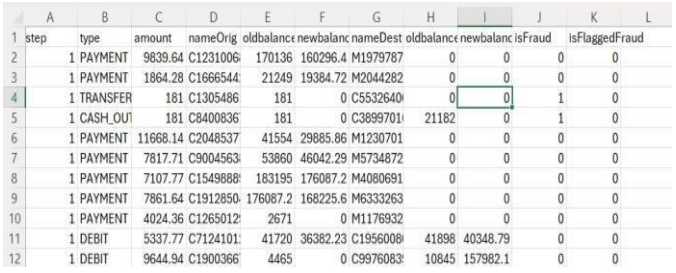
Initially, transaction data is collected from various sources such as online platforms, this raw transaction data then undergoes preprocessing, where it is cleansed, normalized, and enriched to ensure consistency and quality. Once the data is prepared, it is fed into the fraud detection system, where sophisticated algorithms analyze transaction patterns and behaviors to identify potentially fraudulent activity. These algorithms leverage features such as transaction amount, frequency, location, and user behavior to distinguish between legitimate and fraudulent transactions.

The real-time monitoring components continuously monitor transaction streams for any anomalies or suspicious patterns. If a transaction is flagged as potentially fraudulent, it is routed to an alerting system, which notifies relevant stakeholders for further investigation or intervention.

Fig-2: Data Flow



Our dataset obtained from the Kaggle website. Ten thousand transaction details were used as training data among them we divide as type of payment, amount, original name, old balance and new balance. One thousand datasets were used for testing and real-time datasets were generated as transactions occurred. To efficiently detect fraud transactions the real-time datasets are processed and compared with the acquired dataset.



	A	B	C	D	E	F	G	H	I	J	K	L
1	step	type	amount	nameOrig	oldbalance	newbalanc	nameDest	oldbalance	newbalanc	isFraud	isFlaggedFraud	
2	1	PAYMENT	9839.64	C1231006	170136	160296.4	M1979787	0	0	0	0	
3	1	PAYMENT	1864.28	C1666544	21249	19384.72	M2044282	0	0	0	0	
4	1	TRANSFER	181	C1305486	181	0	C5532640	0	0	1	0	
5	1	CASH_OUT	181	C8400836	181	0	C3899701	21182	0	1	0	
6	1	PAYMENT	11668.14	C2048537	41554	29885.86	M1230701	0	0	0	0	
7	1	PAYMENT	7817.71	C9004563	53860	46042.29	M5734872	0	0	0	0	
8	1	PAYMENT	7107.77	C1549888	183195	176087.2	M4080691	0	0	0	0	
9	1	PAYMENT	7861.64	C1912850	176087.2	168225.6	M6333263	0	0	0	0	
10	1	PAYMENT	4024.36	C1265012	2671	0	M1176932	0	0	0	0	
11	1	DEBIT	5337.77	C7124101	41720	36382.23	C1956008	41898	40348.79	0	0	
12	1	DEBIT	9644.94	C1900366	4465	0	C9976083	10845	157982.1	0	0	

Fig-3: Kaggle Dataset Image

7. IMPLEMENTATION AND RESULT

1. Login/Register: This step, the user registers or logs in to the website to make transactions. user login user their credentials like name or email-Id and password. If the user is a valid user he/she is logged in successfully other it will fail then the user needs to register to the website.

2. Make Transaction: After the successful login, the user needs to make Transactions on this website. Once the

transaction is completed the real-time dataset is created using the currently occurring transaction.

3. Pre-processing Dataset: This step will clean the data by handling missing values and noise. Perform features engineering to extract relevant features and transform the data into a suitable format for the Random Forest and Gradient Boosting algorithm.

4. Random Forest algorithm for efficient detection of fraudulent activities:

- The Random Forest algorithm offers several benefits in the detection of fraudulent activities. By leveraging the ability to handle large amounts of Dataset.

- Random Forest can analyze various features associated with the Transaction this will help in detecting the patterns that are indicative of fraudulent activities.

- This algorithm can able to identify anomalies in the transaction such as unusual purchasing patterns, huge amounts, inconsistent IP addresses, or typical user behaviors.

5. Gradient Boosting algorithm:

- This algorithm works by combining multiple weak learners sequentially, with each new learner focusing on the mistake made by the previous one.

- This process is very helpful to improve the model's predictive power

- Gradient boosting can effectively handle imbalanced datasets and capture complex relationships between various features.

- During the fraud detection phase, each decision tree votes on the class label of input data. The final prediction is determined by combining the individual predictions.

6. Fraud Detection: Following the use of algorithms the transaction undergoes feature extraction and attribute selection based on the training dataset it will detect whether the transaction is Fraud or not. Once the user completes the Transaction process it will detect the risk of the ongoing transaction and detect the chances of fraud.

7. Notify the user and the bank branch manager at the same time: If the transaction is detected as fraud, it will block the transaction and also send the notification to the user as well as the branch manager. Suppose the transaction is done by the user he receives the notification message and approves the transaction further or if it was done by the intruder the account holder receives the message, he/she can block the transaction.

From this project we can safeguard our money and personal information this will be very helpful in detecting the anomalies happening in our online transactions. This can be performed in real-time scenarios so, we can detect fraudulent activities at any time.

8. CONCLUSION

To sum up, using machine learning techniques to the identification of fraudulent online transactions provides a strong defence against fraudulent activity in financial institutions. Organizations may efficiently identify and minimize potential threats in real time by utilizing advanced algorithms and comprehensive data analysis. Machine learning models can optimize fraud protection efforts by improving detection accuracy while reducing false positives and negatives through constant refinement and adaption to changing fraud patterns. By preserving a smooth transaction experience, this not only results in significant cost savings and risk mitigation but also maintains client trust and happiness. In addition, the facilitation of regulatory standards compliance and the assurance of transparency in the detection process allow enterprises to achieve strict criteria without falling behind developing risks. In the end, machine learning-based online transaction fraud detection is a proactive strategy that fights financial fraud, strengthens security, and preserves the integrity of digital transactions.

ACKNOWLEDGEMENT

We would like to thank Dr. Ranjit KN, our department head, for her crucial advice and ideas that have helped our project go above and beyond what we had anticipated. We extend our heartfelt gratitude to our Project Coordinator, Dr. HK Chethan, for his unwavering support and dedication in helping us complete this project within a tight timeframe. Our profound gratitude is finally sent to our principal, Dr. Y T Krishne Gowda, for giving us the fantastic chance to work on a project titled "Online Transaction Fraud Detection using Machine Learning" and for his constant support of our study and research efforts.

REFERENCES

[1] Mettildha Mary, Priyadarshini, Dr. Karuppasamy.K, Ms. Margret Sharmila. F. 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) | 978-1-7281-7741-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICACITE51222.2021.9404750

[2] Ms. Kishori Dhanaji Kadam, Ms. Mrunal Rajesh Omanna, Ms. Sakshi Sunil Neje, Ms. Shraddha Suresh Nandai. 1,2,3,4 Student, Sharad Institute of Technology College of Engineering, Ichalkaranji.

[3] U.Siddaiah, P. Anjaneyulu, Y. Haritha, M Ramesh. 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) | 979-8-3503-9725- 3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICICCS56967.2023.10142404

[4] Yeming Chen, Xinyuan Han Clarity AI Beijing, China. 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE) | 978- 1-7281-8319-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCECE51280.2021.9342475.

[5] Petr Hajek, Mohammad Zoynul Abedin, Uthayasankar Sivarajah.

[6] Darshan Aladakatti, Gagana P, Ashwini Kodipalli, Shoaib Kamal. Published in: 2022 International Conference on Smart and Sustainable Technologies in Energy and Power Sectors (SSTEPS)| DOI: 10.1109/SSTEPS57475.2022.00063.

[7] Hongwei Chen, Lun Chen. Published in: 2022 4th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)| DOI: 10.1109/MLBDBI58171.2022.00064