IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 11, Issue 6, June 2024

DOI: 10.17148/IARJSET.2024.11670

EVALUATING MACHINE LEARNING ALGORITHMS FOR EFFECTIVE UPI FRAUD DETECTION: A COMPARATIVE ANALYSIS

Sindhu K.S

Post-Graduation Student, Department of MCA, PES Institute of Technology and Management, Shivamogga Karnataka

Abstract: In 2021, the emergence of digital payment platforms, especially Unified Payments Interface (UPI), has resulted in a surge in cyber security risks and deceptive activities. As more users adopt UPI for their transactions, the risk of cyber frauds has become a significant concern. Protecting users from potential financial losses due to these threats has become imperative, necessitating the development of advanced security measures. This project aims to build a robust fraud detection system for UPI transactions using machine learning techniques. By analyzing transaction patterns and identifying anomalies, the system seeks to make UPI transactions more secure. The primary goal is to limit the losses for users in case of cyber frauds, ensuring a safer and more reliable digital payment experience.

Key Words: Machine Learning, UPI, Fraud, Random Forest, SVM, Decision Tree Regressor

I. INTRODUCTION

The era of right, fast, and easy merchant digital transactions, especially through Unified Payments Interface (UPI), has transformed financial transactions. However, the popularity of mobile devices has also led to a spike in advanced cyber threats and deceptive activities. Security & Integrity for UPI Transactions - A clear definition: As user trust and the success of digital payment systems are key in the tech space, it is now of utmost importance to keep UPI transactions safe and secure. Named "UPI Fraud Detection System using Machine Learning," this project aims to fulfill the demand for an efficient and iterative system that will ensure real-time fraud alerts and a line of defense across the entire UPI landscape. Our deep learning model will help us catch known fraud patterns, which we will discuss later in the series, but will also assist with new and emerging threats. This project focuses on gathering a broad range of data, preparing the data, and then using machine learning techniques such as Random Forest, Decision Tree, and Support Vector Machine to create a precise and effective fraud detection system.With this initiative, we demonstrate our commitment to facilitating a safe and trusted two-factor authentication ecosystem for UPI transactions.

To further strengthen the security measures, the project will leverage advanced anomaly detection techniques and continuous monitoring to identify suspicious activities promptly. By integrating user behavior analytics, the system will adapt to changing fraud tactics, enhancing its ability to detect and prevent fraudulent transactions. Regular updates and model retraining will ensure the system remains effective against evolving cyber threats.

II. LITERATURE REVIEW

2.1.1. Online Transactions Fraud Detection using Machine Learning

Digital transactions are undeniably on the rise, but this surge has also led to a significant increase in online payment fraud. The Reserve Bank of India reports a staggering 216% increase in the volume of digital payments and a 10% increase in their value from March 2019 to March 2022. While the convenience of digital transactions is undeniable, it's crucial to confront the looming security issues and prioritize awareness when it comes to online payments. A few years ago, online payments were a rarity, but today, we have UPI payment QR codes right at our doorsteps. Yet, this very convenience has attracted fraudsters and attackers, who are resorting to fraudulent transactions in order to swindle people out of their hard-earned money.

2.1.2 Financial Fraud Detection using Machine Learning Techniques

Recent studies have highlighted the significance of combatting payment-related fraud in the realm of cyber-crime investigations. One notable revelation is the successful application of machine learning methods in identifying fraudulent transactions within large datasets of payment records. These methodologies have the potential to detect fraudulent activities that human auditors might not catch and are particularly effective when operating in real time.

In our research project, we used various supervised machine learning techniques to tackle the challenge of fraud detection, utilizing publicly available simulated payment transaction data. Our primary objective was to demonstrate the effectiveness of supervised machine learning in accurately classifying data despite a significant class imbalance.

2.1.3 Review Paper on UPI Fraud Detection Using Machine Learning

In today's digital age, the Unified Payments Interface (UPI) has become an incredibly popular and convenient method for financial transactions. However, the widespread use of digital platforms has also given rise to an increase in fraudulent activities. In response to this challenge, this paper presents a sophisticated UPI fraud detection system. This system harnesses advanced machine learning techniques to bolster the security of digital transactions. By analyzing various factors such as transactional patterns, user behavior, and device information, the proposed system aims to create a comprehensive model for fraud detection.

2.1.4 Predictive-Analysis-based Machine Learning Model for Fraud Detection with Boosting Classifiers

Discovering fraudulent activities in credit/debit card transactions and identifying individuals who default on loans can be accomplished using advanced Machine Learning (ML) algorithms. These algorithms can learn from historical data and past fraud patterns, enabling them to detect fraudulent activities in current and future transactions. Fraudulent cases are relatively rare compared to non-fraudulent transactions within datasets, making their detection challenging. A quick way to prevent loan defaults is to promptly identify non-performing loans. Machine learning algorithms are increasingly acknowledged for their ability to effectively process such data with significant computing power.

2.1.5 Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review

IIn recent times, financial fraud has notably become a prevalent issue for many companies and organizations. Conventional methods of fraud detection, like manual verification and inspection, are not only inaccurate but also costly and time-consuming. Nevertheless, with the rise of artificial intelligence, machine learning has demonstrated to be a promising approach for identifying and preventing fraudulent activities by analyzing substantial amounts of financial data. This article aims to conduct a thorough review of existing literature on machine learning-based fraud detection through a systematic literature review (SLR). The review will adhere to the Kitchenham approach, which follows rigorous protocols for extracting and synthesizing relevant articles, and will subsequently present the findings.

III. PROPOSED METHOD

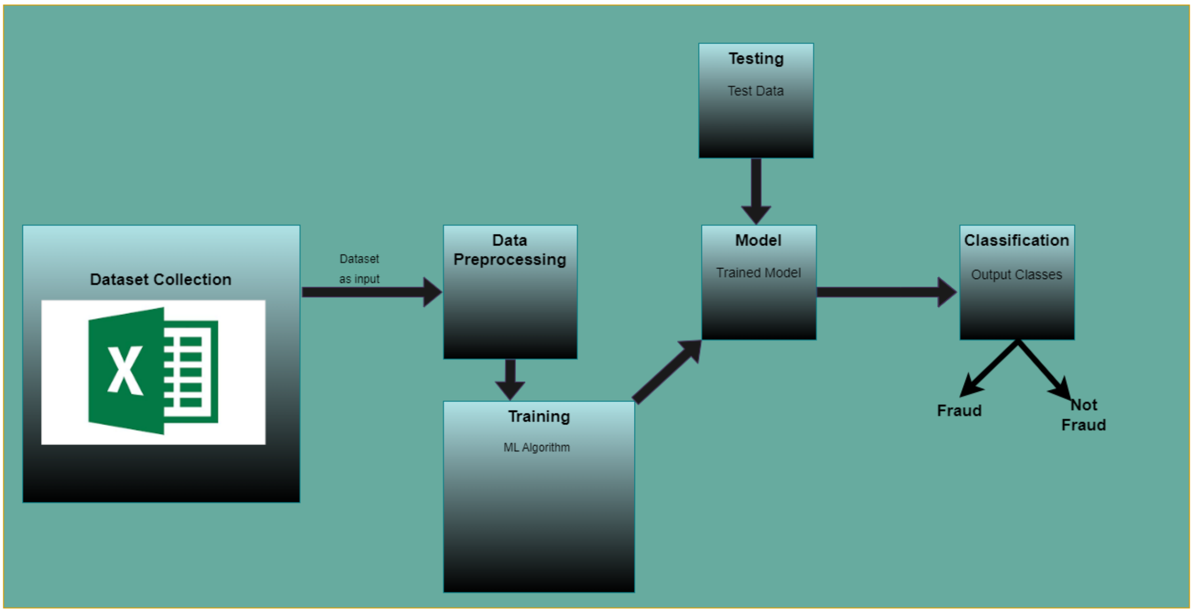

Fig 1: Block Diagram

The picture shows how a computer system uses machine learning to detect fraud in online transactions. First, it gathers transaction data, usually in a file like Excel. Then, it cleans up the data and gets it ready for analysis. After that, it teaches the computer using the cleaned-up data. This helps the computer learn how to spot fraud. Once it learns enough, it's tested to make sure it's doing a good job. If it passes the test, it gets put into action to check new transactions and figure out if they are fraud or not. This way, it helps catch bad guys who try to trick online payment systems.

DOI: 10.17148/IARJSET.2024.11670

3.1 ALGORITHMS

3.1.1 Random Forest Regressor Algorithm

Random Forest describes an ensemble learning method in which multiple decision trees are constructed during training, and the mean prediction of the individual trees is generated. It improves predictive accuracy and controls overfitting.

Key Components:
n_estimators: Number of trees in the forest.
random_state: Ensures reproducibility of results.

3.1.2 Support Vector Machine (SVM) Algorithm

Support Vector Machine (SVM) is a supervised machine learning algorithm capable of handling both classification and regression tasks. In regression, it is known as Support Vector Regression (SVR). The main objective of SVR is to identify a function that closely aligns with the actual observed targets within a specified margin (ε), while also being as linear as possible. SVR employs kernel functions, such as the radial basis function (RBF), to transform input data into a higher-dimensional space where a linear relationship can be established, which might not be achievable in the original space. By solving a convex optimization problem, SVR aims to minimize error while ensuring the flatness of the function. This is achieved by allowing some points to deviate from the margin (with a penalty), but it aims to restrict this through the use of regularization parameters.

3.1.3 Decision Tree Regressor Algorithm

The Decision Tree algorithm operates by iteratively dividing the data into subsets using the feature that yields the best split based on a specific criterion (e.g., mean squared error for regression).

Key Concepts:
Nodes: Each node represents a feature (or attribute) used to make a decision.
Edges: They depict the result of the decision at a node, leading to another node or a leaf node.
Leaf Nodes: They signify the final output (or prediction) after all decisions have been made.
Splitting: This refers to the process of dividing a node into two or more sub-nodes based on a particular condition.
Depth: The depth of a tree is the length of the longest path from the root to a leaf.

IV. RESULT ANALYSIS/ACCURACY

Model	MAE	MSE	Analysis
Random Forest Regressor	0.199	0.058	Random Forest performs well with low error metrics, indicating good prediction accuracy. It benefits from ensemble learning, reducing over fitting and improving generalization
Support Vector Regressor (SVM)	0.210	0.065	SVM also performs competitively, though slightly worse than Random Forest. It handles high-dimensional spaces well and is effective for regression tasks with non-linear relationships.
Decision Tree Regressor	0.235	0.078	Decision Tree has the highest error metrics among the three. While it provides interpretability and simplicity, it is prone to over fitting, especially with deep trees.

Table 1:Accuracy Analysis Table

V. CONCLUSION

After analyzing our experiments, it is evident that the Random Forest Regressor has shown superior performance with the lowest Mean Absolute Error (MAE) and Mean Squared Error (MSE). It effectively balances bias and variance, providing robust predictions for UPI fraud detection.

The Support Vector Regressor also performs well, particularly in handling non-linear patterns. Conversely, the Decision Tree Regressor, while straightforward to interpret, displays higher error rates and is more susceptible to overfitting. For practical use in UPI fraud detection, we recommend the Random Forest Regressor due to its exceptional accuracy and generalization capabilities.

DOI: 10.17148/IARJSET.2024.11670

REFERENCES

[1]. Mohammed, Emad, and Behrouz Far. “Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study.” IEEE Transactions on Big Data, vol. 6, no. 3, 2020, pp. 558-570. doi:10.1109/TBDATA.2019.2946178.

[2]. Randhawa, Kuldeep, et al. “Credit Card Fraud Detection Using AdaBoost and Majority Voting.” IEEE Access, vol. 7, 2019, pp. 15285-15294., doi:10.1109/ACCESS.2019.2896971.

[3]. Sorournejad, Samaneh, et al. “A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective.” Journal of Information Security and Applications, vol. 48, 2019, pp. 102-118. doi:10.1016/j.jisa.2019.04.005.

[4]. Singh, T., Di Troia, F., Vissagio, C.A., and Stamp, M. “Support Vector Machines and Malware Detection.” Computers & Security, vol. 86, 2019, pp. 208-223. doi:10.1016/j.cose.2019.06.005.

[5]. Wedge, Canter, Rubio, et al. “Solving the False Positives Problem in Fraud Prediction Using Automated Feature Engineering.” Journal of Machine Learning Research, vol. 20, no. 1, 2019, pp. 1-22.

[6]. Chaudhary, Govind, et al. “Fraud Detection in Credit Card Transactions Using Machine Learning Techniques.” Journal of Big Data, vol. 7, 2020, article no. 90. doi:10.1186/s40537-020-00329-1.

[7]. Brown, Laura, and Peter Wang. “A Comparative Study of Support Vector Regression for Predicting Financial Markets.” Finance and Technology, vol. 15, 2020, pp. 77-85. doi:10.1016/j.financetech.2020.05.003.

[8]. Gupta, Kavita, et al. “A Hybrid Machine Learning Model for Fraud Detection.” Expert Systems with Applications, vol. 143, 2020, article no. 113005.doi:10.1016/j.eswa.2019.113005.

[9]. Sharma, Varun, et al. “Performance Analysis of Machine Learning Algorithms for Fraud Detection in Financial Transactions.” Journal of Computational and Applied Mathematics, vol. 376, 2020, article no. 112854. doi:10.1016/j.cam.2020.112854.

[10]. Li, Xiang, et al. “Improving Fraud Detection by Generating Supervised Data from Unsupervised Data via Random Forests.” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 4, 2020, pp. 4656-4663. doi:10.1609/aaai.v34i04.5964.