e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:09/September-2024 Impact Factor- 8.187 www.irjmets.com

UPI FRAUD DETECTION USING MACHINE LEARNING

Yash Patil[*1], Amar Shinde[*2], Yash Parthe[*3], Sameer Sayyad[*4]

[*1,2,3,4]Computer Department Savitribai Phule Pune University Pune, India.

ABSTRACT

The Unified Payments Interface (UPI) has revolutionized digital payments with its ease of use and instantaneous transactions. However, this rapid adoption has also led to an increase in fraudulent activities. It presents a novel approach to detecting UPI fraud using advanced machine learning techniques and behavioural analytics. By leveraging a diverse dataset encompassing transaction patterns, user behaviours, and historical fraud cases, we develop a robust detection system that identifies suspicious activities with high accuracy. Our approach integrates features such as anomaly detection, pattern recognition to enhance the security of UPI transactions. Experimental results demonstrate a significant improvement in fraud detection rates compared to traditional methods, highlighting the effectiveness of our model in minimizing financial losses and safeguarding user trust in digital payment systems.

Keywords: SVM, Dataset, Preprocessing, Feature Extraction, Classification.

I. INTRODUCTION

Unified Payments Interface (UPI) has significantly transformed the landscape of digital financial transactions, offering a convenient, real-time method for transferring money between accounts. Since its inception, UPI has gained widespread acceptance due to its ease of use, low transaction costs, and integration with various financial institutions and service providers. Its ability to facilitate instant payments with minimal friction has made it a preferred choice for millions of users globally. However, the rapid expansion of UPI usage has also brought to light the vulnerabilities associated with digital payments.

The simplicity and openness of the UPI system, while advantageous for users, have also made it a target for fraudulent activities. In response to these challenges, it introduces a novel fraud detection framework designed specifically for UPI transactions. By integrating advanced machine learning algorithms and behavioral analytics, our approach aims to enhance the detection of fraudulent activities. We explore the use data analysis to identify anomalies, improve the accuracy of fraud detection, and safeguard the integrity of UPI transactions.

II. RELATED WORK

FRAUD DETECTION IN UPI TRANSACTIONS USING ML, et.al, J. Kavitha, G. Indira, A. Anil kumar, A. Shrinita, D. Bappan Significant obstacles to financial security have arisen as a result of the quick uptake of Unified Payments Interface (UPI) for online transactions and a commensurate rise in fraudulent activity. This paper suggests a novel fraud detection method that makes use of cutting-edge machine learning (ML) algorithms to address this urgent issue. It focuses on integrating a Hidden Markov Model (HMM) into the UPI transaction process. In order to enable the system to identify departures from these learnt behaviour's as possibly fraudulent, the HMM is trained to predict the typical transaction patterns for particular cardholders.

UPI Based Mobile Banking Applications – Security Analysis and Enhancements et.al K. Krithiga Lakshmi, Himanshu Gupta, Jayanthi Ranjan Technology advancements have reduced the cost of both a mobile device and data connection making it affordable to all. In parallel, mobile applications are also rising providing the quick, easy door-step solution(s) to one's professional and personal requirements. In the current trend of the digital and cashless economy, mobile-based app solutions are easy to use and ubiquitous, facilitating a wide range of banking financial services (pay/collect money etc.) and non-financial services (cheque request, account balance, view transaction history etc.).

Online Fraud Detection System et.al Prof. D.C. Dhanwani, Aniruddh Tonpewar, Devashish Ikhar, Komal Ladole, Suyog Mahant Financial services are used everywhere and function with high complexity. With the increase in online transacting, frauds too are increasing alarmingly. An automated Fraud Detection System is thus required. With millions of transactions taking place, it is practically impossible to detect frauds manually with good speed and accuracy. We propose a system is that provides a robust, cost effective, efficient yet accurate solution to detect frauds in both online payment transactions and credit card payments. The proposed solution is a

Machine Learning model that will serve the purpose of detecting “fraudulent” and all the “genuine” transactions in real time.

III. METHODOLOGY

UPI fraud detection refers to the process of identifying and preventing fraudulent transactions conducted through the Unified Payments Interface (UPI). UPI allows users to transfer money instantly between bank accounts using a mobile phone. While it has become highly popular due to its ease of use, this widespread adoption has also made it a target for various types of fraud.

Machine Learning in UPI Fraud Detection:

Machine learning (ML) plays a critical role in detecting UPI fraud by building models that can:

- Classify Transactions: ML models are trained to distinguish between fraudulent and legitimate transactions based on historical data.
- Predict Fraud Patterns: Using supervised learning, these models learn from labeled data (fraudulent vs. non-fraudulent transactions) and predict which future transactions are likely fraudulent.
- Handle Imbalanced Data: Fraud detection datasets are usually imbalanced (i.e., there are far fewer fraudulent transactions than legitimate ones).

What is SVM:

Support Vector Machine (SVM**)** is a powerful supervised machine learning algorithm used for classification, regression, and outlier detection tasks. It is widely used for binary classification tasks, but it can also handle multi-class problems. SVM aims to find the best boundary (decision boundary) that separates data points belonging to different classes in such a way that the margin between the boundary and the closest points from either class is maximized.

Support Vector Machine (SVM) is a powerful supervised machine learning algorithm used for classification, regression, and outlier detection tasks. It is widely used for binary classification tasks, but it can also handle multi-class problems. SVM aims to find the best boundary (decision boundary) that separates data points belonging to different classes in such a way that the margin between the boundary and the closest points from either class is maximized.

Key Concepts of SVM:

1. Hyperplane: In SVM, a hyperplane is a decision boundary that separates different classes. For a 2D space, it's a line; for 3D, it's a plane; and for higher dimensions, it's a hyperplane. SVM tries to find the hyperplane that best separates the classes.
2. Support Vectors: The data points closest to the hyperplane are called support vectors. These points are crucial because they define the optimal hyperplane. The algorithm only relies on these support vectors to determine the decision boundary, hence the name Support Vector Machine.
3. Margin: The margin is the distance between the hyperplane and the closest data points (support vectors) from each class. SVM tries to maximize this margin. A larger margin reduces the risk of misclassification on new data points, leading to a better generalization of the model.
4. Linear vs. Nonlinear SVM:

- ○ Linear SVM: When the data is linearly separable, SVM tries to find a straight hyperplane that separates the data.
- ○ Nonlinear SVM: In many real-world problems, data is not linearly separable. In such cases, SVM uses a technique called the kernel trick to map data into a higher-dimensional space where a linear hyperplane can separate the classes.

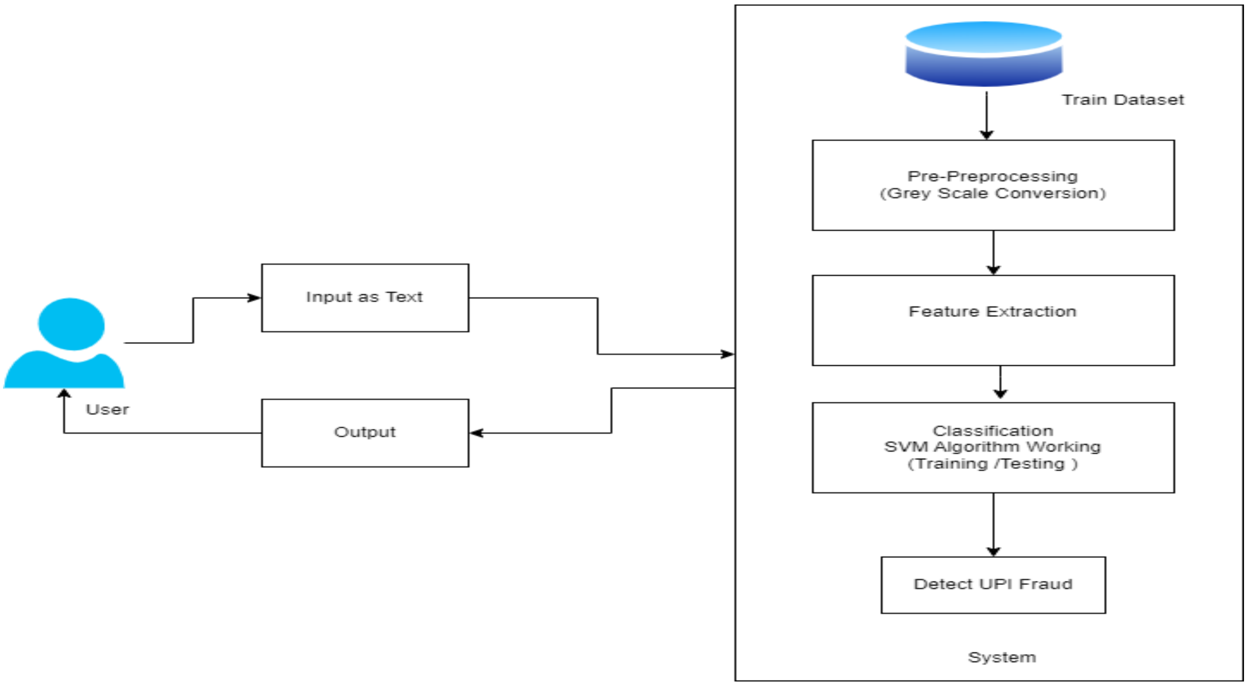

Figure 1: Architecture

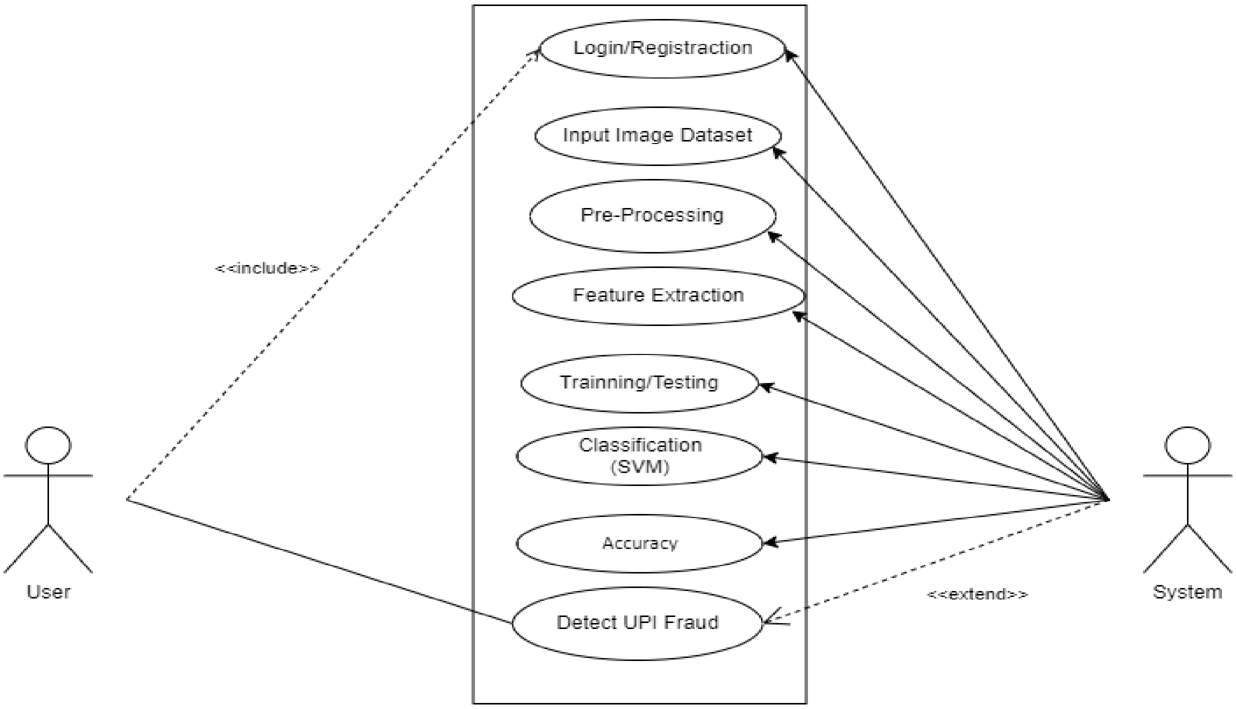

Figure 2: Use Case Diagram

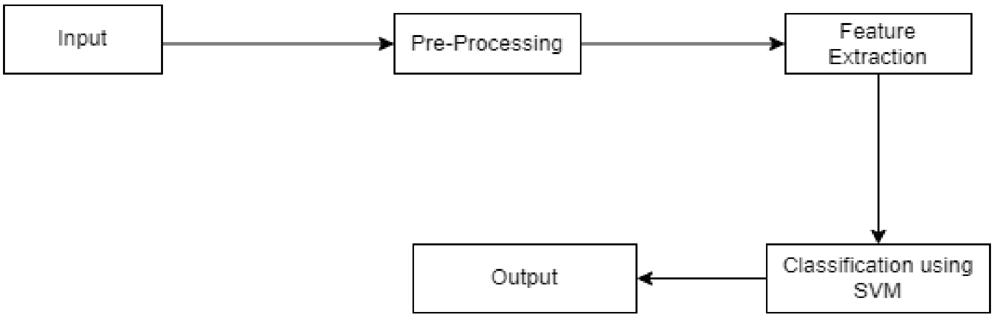

Figure 3: Data Flow Diagram

IV. COMPARISION AND ANALYSIS

We compare the performance of the Support Vector Machine (SVM) model against other machine learning algorithm and analyze key findings. This analysis helps assess how well SVM performs in detecting fraud, particularly in comparison to other methods.

- **Accuracy**: The proportion of correctly classified transactions (fraudulent and legitimate) out of all transactions.
- **Precision**: The proportion of true fraud cases among the transactions flagged as fraudulent.
- **Recall (Sensitivity)**: The proportion of actual fraud cases correctly identified by the model.
- **F1-Score**: The harmonic mean of precision and recall, especially useful in imbalanced datasets like fraud detection.

V. DISCUSSION

SVM has proven to be effective in detecting fraud in UPI transactions, especially in terms of its precision and accuracy. By defining a clear margin between fraudulent and non-fraudulent transactions, SVM is capable of drawing an optimal hyperplane in the feature space, particularly useful in non-linear scenarios. UPI fraud detection inherently deals with imbalanced data, where fraudulent transactions form only a small percentage of the overall dataset. This imbalance causes the model to favor legitimate transactions unless handled carefully. Identifying relevant features from transaction data, such as time-based patterns, transaction velocity, and geographical mismatches, was crucial. However, the selection process was challenging due to the sheer volume of features available and the dynamic nature of fraud strategies.

One of the challenges with using SVM in fraud detection is its interpretability**.** While SVM is a powerful classifier, its decisions—especially with non-linear kernels—arenot always easily understood. In a fraud detection system, interpretability is crucial as banks and financial institutions need to understand why certain transactions were flagged as fraudulent.

VI. CONCLUSION

The SVM-based UPI fraud detection model performs well in identifying fraudulent transactions, particularly with high precision. However, it faces challenges with data imbalance, scalability, and interpretability. Comparatively, Random Forest may be better suited for maximizing fraud detection, but SVM holds strong potential, especially when optimized for the user. As fraud strategies evolve, continuous updates, feature engineering, and potentially hybrid models will be necessary to maintain the model's effectiveness. Financial institutions should consider integrating these advanced models into their systems while balancing customer experience and security.

By leveraging an SVM classifier with appropriate feature selection, data balancing, and hyperparameter tuning, the system can classify transactions as fraudulent or legitimate with a high degree of accuracy. The continuous updating of the model with new data ensures that the fraud detection system remains effective over time, adapting to new fraud techniques and patterns.

ACKNOWLEDGEMENTS

We would like to express our deepest gratitude to all those who contributed to the success of this UPI fraud detection project using SVM.

Firstly, we extend our sincere thanks to our project supervisor for their invaluable guidance, continuous support, and constructive feedback throughout the project's development.

We are also grateful to our institution for providing the necessary infrastructure, resources, and learning environment, without which this project would not have been possible.

Lastly, we would like to acknowledge the open-source community for making data sets, machine learning frameworks, and resources available, which played a key role in shaping our model and its implementation.

Without the collective efforts and support from all these individuals and organizations, this project would not have come to fruition.

We would like to extend our gratitude to Project staff at the Sinhgad Academy of Engineering, Department of Computer Engineering for their support and help without which this work could not have been accomplished. Additionally, we would like to give Prof. P. N. Wadibhasme, our project manager, our profound gratitude for leading us through the process

VII. REFERENCES

[1] ALESKEROV E, FREISLEBEN, B., and, RAO B CARDWATCH: A neural network-based database mining system for credit card fraud detection. In Conference (pp. 220–226). IEEE, Piscataway, NJ

[2] Sahin M Understanding Telephony Fraud as an Essential Step to Better Fight it [Thesis]. École Doctorale Informatique, Télécommunication et Électronique, Paris

[3] Abdallah A, Maarof MA, Zainal A Fraud detection system: A survey. J Netw Comput Appl 68:90–113

[4] ANDREWS PP, PETERSON MB (eds) Criminal Intelligence Analysis. Palmer Enterprises, Loomis, CA

[5] ARTÍS M, AyUSO M, GUILLÉN M Modeling different types of automobile insurance fraud behavior in the Spanish market. Insurance Math Econ 24:67–81.

[6] 6.BARAO MI, TAWN JA \ Extremal analysis of short series with outliers: Sea-levels and athletics records. Appl Stat 48:469–487

[7] BLUNT G, HAND DJ The UK credit card market. Technical report, Department of Mathematics, Imperial College, London.

[8] BOLTON RJ, HAND DJ Unsupervised profiling methods for fraud detection. In Conference on Credit Scoring and Credit Control 7, Edinburgh, UK, 5–7 Sept

[9] Phua C, Lee V, Smith K, Gayler R A comprehensive survey of data mining-based fraud detection research.

[10] Summers SL, Sweeney JT Fraudulently misstated financial statements and insider trading: An empirical analysis.