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

UPI fraud detection using ML

Vinay chauhan
B. Tech IV Year
School of Computing Science and Engineering
Galgotias University,
Greater Noida, India.
Vc203132@gmail.com

Mr. Amit Kumar
Assistant Professor
School of Computer Science and Engineering
Galgotias University,
Greater Noida, India.
amit.kumar@galgotiasuniversity.edu.in

Prashant Mann
B. Tech IV Year
School of Computing Science and Engineering
Galgotias University,
Greater Noida, India.
prashant.mann.2002@gmail com

Abstract— The increasing use of unified payment interfaces (upi) for online payments has provided the convenience and efficiency synonymous with digital transactions, but this has resulted in alarmingly high rates of fraud, putting the system's reliability and stability at risk. This paper explores how to solve the upi fraud detection problem by using a hmm Markov model. The multichannel upi communications scheme is investigated and modelled by using the hmm framework. This accounting system is similar to a typical sherlock Holmes system, in that it is able to distinguish between legitimate and fraudulent businesses. Here's how it works: the hmm is first introduced to the cardholder by analysing their previous transaction behaviour. This involves comparing sales volumes, locations, frequency, and other patterns. When a cardholder initiates a new upi transaction, the trained hmm learns to recognize what normal " behaviour looks like for that particular cardholder. If the transaction is judged to be valid by hmm to correspond to the cardholder's specific transaction, it is deemed valid. However, if the hmm raises a red flag and indicates that the behaviour does not conform to scholarly principles, it is considered fraudulent, prompting a further investigation. This project aims to show how data modelling and machine learning can be used to shield individuals and financial institutions from upi fraud. We're making a significant leap forward in terms of not only convenience but also safety in upi transactions by implementing a system that can detect suspicious activity in real time, while still maintaining that they are all digital. Keywords: fraud detection, unified payment interface (upi), and hidden.

I. INTRODUCTION

The exponential rise in the use of upi (unified payments interface) for online payments has undoubtedly changed the way we conduct financial transactions. Nevertheless, this rise in digital payment methods has accompanied a worrying and parallel rise in fraudulent activities. The need for a robust and efficient upi fraud detection system has never been greater. As upi rises in popularity, the tendency to misappropriate and deceptive activities has increased in tandem, resulting in significant financial losses for both upi holders and financial institutions.

These illicit transactions pose a significant threat to the integrity of online payment systems, necessitating new technologies that can be applied to the evolving nature of fraud. In response to these challenges, this project investigates the development of a sophisticated upi fraud detection scheme that addresses the inherent difficulties associated with upi transactions. To study the sequenced steps involved in a upi transaction process, we use a hidden Markov model (hmm), in order to improve the security and integrity of digital payments.

The hmm's initial training is provided to individual cardholders, enabling it to distinguish between legitimate and potentially fraudulent activities. The main aim of this project is to harness the power of recent advances in machine learning algorithms, among other things. Five distinct algorithms have been carefully designed and tested to the specific complexity of upi fraud detection, rigorously evaluating their performance.

Beyond our algorithmic prowess, we face significant challenges such as data accessibility, class division, evolving fraud patterns, and the prevalence of false alarms. In summary, this initiative serves as a critical response to the increasing concern about upi fraud. It aims to provide a cost-effective and adaptable solution that protects the financial rights and privacy of users while simultaneously maintaining upi's reputation as a trusted online payment platform.

In addition to the technical difficulties, this project also addresses the wider implications and concerns surrounding upi fraud detection. To create comprehensive frameworks for fraud prevention and mitigation, we acknowledge the importance of collaboration between all actors, including financial institutions, regulatory bodies, and technology providers. We aim to create a single front against deceptive activities by fostering relationships and sharing knowledge, while also improving the upi ecosystem's resilience. In addition, our approach emphasizes the iterative nature of fraud detection schemes. Our system must be dynamic and responsive as fraudsters continue to change their tactics. We develop tools for continuous monitoring and improvement, leveraging real-time data streams and feedback loops to increase the precision and efficiency of fraud detection algorithms.

In addition, user education and awareness play a vital role in reducing upi fraud. We advocate for proactive measures to empower users with the knowledge and skills necessary to spot and report suspicious activities. We can reduce the likelihood of successful fraud attempts by instilling a culture of vigilance and accountability, as well as minimize the chance of financial and reputational harm. The success of our project is ultimately dependent on its ability to strike a balance between security and usability. Although effective fraud prevention strategies are vital, they should not hinder the smooth operation of legitimate transactions or compromise the user experience. We aim to achieve a harmonious balance of security and convenience by carefully planning and optimizing, thereby fostering mutual trust and confidence in the upi ecosystem for all parties involved

II. RELATED WORK

In addition to machine learning algorithms, previous studies in the area of upi fraud detection have also investigated other ways and approaches to improve the efficiency of fraud detection systems. One such method is the use of behavioural analysis and anomaly detection techniques. Anomalies or deviations from normal behaviour can be detected by observing the behaviour patterns of users during transactions, such as the order of actions performed, navigation patterns, and interaction history. In addition, graph-based analysis techniques have been used to map the relationships between various parties involved in upi transactions, such as users, merchants, and financial institutions. By presenting these relationships as a graph chart, suspicious patterns or connections between individuals can be identified, facilitating the detection of fraudulent activities such as collusion or money laundering. In addition, advanced data mining techniques such as association rule mining and sequence pattern mining have been used to uncover obscure patterns and associations in transactional data. These techniques are used to identify common itemset, sequences, and unusual transaction sequences that may indicate fraudulent activities. In addition, novel data pre-processing and feature engineering techniques have been investigated to improve the consistency of input data and increase the efficiency of machine learning algorithms. Features such as feature scaling, dimensionality reduction, and data imputation aid in preparing the data for analysis and ensure that the models accurately represent the data's patterns. In addition, ensemble learning schemes such as bagging, boosting, and stacking have been suggested to combine the results of multiple base classifiers to improve the overall reliability and stability of fraud detection systems. In general, continuous improvement and integration are key to success.

III. METHODOLOGY

The methodology for this project involves the development of a UPI fraud detection system using machine learning with primary emphasis on Hidden Markov Model (HMM). Starting with the collection of relevant datasets, efforts are aimed at ensuring a diverse representation of UPI transaction scenarios. Subsequent preprocessing steps focus on data cleaning and feature normalization. Key to the methodology is the strategic selection of HMMs and other machine learning algorithms such as logistic regression and decision trees. The HMM is trained by analyzing the transaction behavior of cardholders, including sales volumes, locations and frequencies, allowing it to recognize common patterns indicative of legitimate transactions. The integration of machine learning algorithms complements HMM and provides different approaches to identify potential fraud. Evaluation metrics, including accuracy, precision, repeatability, and false positive rate, are carefully selected to comprehensively evaluate system performance. Testing and validation involve exposing the system to a separate set of data to measure performance and address potential issues. The results obtained from this testing phase contribute to the continuous improvement of the fraud detection system and offer insight into the performance of each algorithm with the ultimate goal of strengthening UPI transactions against fraudulent activities.

IV. PROPOSED SYSTEM

The exponential rise of upi (unified payments interface) transactions has unquestionably changed the landscape of digital financial transactions. However, this surge has also resulted in a worrying rise in deceptive activities, posing a significant threat to the system's stability and reliability. This project seeks to solve the problem of upi fraud detection by using machine learning techniques, particularly by utilizing a hidden Markov model (hmm). The project begins by acknowledging the urgent need for a robust upi fraud detection system. It emphasizes the importance of adapting to the changing nature of fraud and the need for new approaches to safeguard the integrity of digital payments. With a particular emphasis on the hmm framework, the proposed solution leverages the capabilities of machine learning algorithms. The hmm is designed to help individual cardholders understand how to distinguish between legitimate and potentially fraudulent transactions. To establish a baseline for normal behavior, a researcher must look at various transaction attributes such as sales volumes, locations, frequencies, and patterns. When a new transaction occurs, it is routed through the trained hmm, which checks its conformity to the cardholder's established behavior. Any deviations detected by the hmm trigger alerts for further investigation, allowing for the swift detection of fraudulent activities. The initiative also emphasizes the importance of addressing major challenges such as data availability, class division, evolving fraud patterns, and false alarms. It emphasizes the need for rigorous evaluation and testing of multiple machine learning algorithms that are specifically suited to the challenges involved in upi fraud detection. In addition, the initiative broadens its scope to include other aspects and considerations, including:

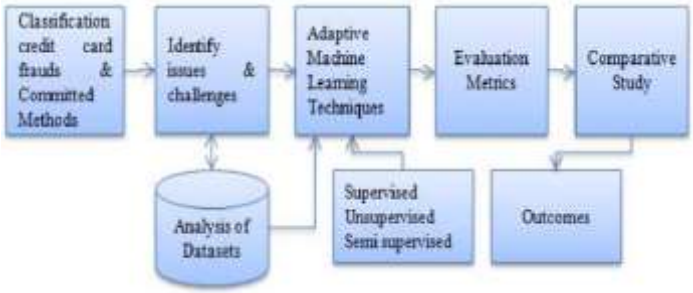

Figure 1. Workflow of the fraud detection model

a) *Import* libraries Data Initialization

```
import pandas as pd
import numpy as np
data = pd.read_csv("upifraud.csv")
data.head()
```

	Volume	Value (Cr)	Volume	Value (Cr)	Month	Year
Airtel Payme	8.53	2,047.45	13.22	4,729.77	1	2022
Airtel Payme	5.8	1,199.46	7.58	2,210.17	2	2022
Airtel Payme	8.33	1,934.41	13.16	4,492.25	3	2022
Airtel Payme	5.29	454.64	5.3	460.9	4	2022
Airtel Payme	6.1	486.55	6.11	486.62	5	2022
Airtel Payme	6.38	506.14	6.38	506.2	6	2022
Airtel Payme	7.4	555.09	7.41	555.26	7	2022
Allahabad Ba	0.02	7.36	0.02	7.36	1	2022
Allahabad Ba	0.02	6.28	0.02	6.28	2	2022
Allahabad Ba	0.02	6.15	0.02	6.15	3	2022
Allahabad Ba	0.02	6.25	0.02	6.25	5	2022
Allahabad Ba	0.02	6.79	0.02	6.79	6	2022
Allahabad Ba	0.03	6.95	0.03	6.95	7	2022
Amazon Pay	73.5	6,729.66	73.5	6,729.66	1	2022
Amazon Pay	63.49	6,044.47	63.49	6,044.47	2	2022
Amazon Pay	76.34	6,894.78	76.34	6,894.78	3	2022
Amazon Pay	73.21	6,699.57	73.21	6,699.57	4	2022

Figure 2. Data used for prediction in the model

b) Data Quality Assessment

- Examine the dataset for missing values in features and the target variable.
- Utilize the .is null() method to identify Nan values.
- Summarize the total number of missing values found.
-

```
# Check for NaN values in features
print(data.isnull().sum())

# Check for NaN values in target variable
print(data["isFraud"].isnull().sum())
```

Exploring Transaction Types

- Explore the distribution of transaction types in the dataset.
- Use the .value counts() method to count occurrences of each transaction type.
- Gain insights into the frequency of different transaction types present in the data.

```
# exploring transtion type
data.type.value_counts()
```

Visualizing Transaction Distribution

- Calculate the frequency of each transaction type using .value counts().
- Extract transaction types and their corresponding quantities.
- Utilize Plotly Express to create a pie chart visualizing the distribution of transaction types.
- Display the chart with a title indicating the distribution of transaction types.

```
type = data["type"].value_counts()
transactions = type.index
quantity = type.values

import plotly.express as px
figure = px.pie(data,
                values=quantity,
                names=transactions,hole = 0.5,
                title="Distrubution of transtion Type")
figure.show()
```

c) Correlation Analysis

- Select numeric columns, including the target variable "isFraud," using .select_dtypes().
- Calculate the correlation matrix for numeric features using .corr().
- Sort the absolute correlation values with the target variable in descending order.
- Print the correlation values to identify the features most strongly correlated with "isFraud."

```
# Select only numeric columns including "isFraud"
numeric_data = data.select_dtypes(include=[np.number])

# Checking correlation
correlation = numeric_data.corr()
print(correlation["isFraud"].abs().sort_values(ascending=False))
```

d) Encoding Categorical Variables

- Map categorical transaction types to numeric values for analysis.
- Map the target variable "isFraud" to a descriptive label ("No fraud").
- Update the dataset with the mapped values and display the updated Data Frame.

```
data["type"] = data["type"].map({"CASH_OUT": 1, "PAYMENT": 2, "CASH_IN": 3, "TRANSFER": 4, "DEBIT": 5})
data["isFraud"] = data["isFraud"].map({0: "No fraud"})
data.head()
```

e) Data Splitting

- Import the necessary function train_test_split from sklearn.model_selection.
- Extract features (X) and the target variable (y) from the dataset.
- Split the data into training and testing sets using a specified test size (e.g., 20%) and a random seed for reproducibility.
- Assign the resulting training and testing sets to variables for further model development.

```
# Splitting the data
from sklearn.model_selection import train_test_split

# Extracting features (X) and target variable (y)
X = data[["type", "amount", "oldbalanceOrg", "newbalanceOrig"]].values
y = data["isFraud"].values

# Splitting the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

f) Model Development

- Load the iris dataset from sklearn.datasets.
- Split the dataset into training and testing sets using train_test_split.
- Initialize a Decision Tree Classifier from sklearn.tree.
- Train the classifier on the training data using the .fit() method.
- Evaluate the model's performance on the test set using the .score() method.
- Print the model's accuracy score on the test set to assess its performance.

```
from sklearn.datasets import load_iris
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split

# Load the iris dataset
iris = load_iris()
x = iris.data
y = iris.target

# Split the dataset into training and testing sets
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size=0.

# Train a decision tree classifier
model = DecisionTreeClassifier()
model.fit(xtrain, ytrain)

# Evaluate the model on the test set
score = model.score(xtest, ytest)
print(f"Model score: {score}")
```

i) Making Predictions

- Create a new feature array containing the input features for prediction.
- Utilize the trained decision tree classifier to predict the target value based on the input features.
- Print the predicted target value, representing the predicted class label.

```
import numpy as np

# New feature array with 4 features
features = np.array([[4, 9000.60, 0.0, 0.0]])

# Predict the target value
prediction = model.predict(features)

print(prediction)
```

V. CONCLUSION

Fraudulent upi (unified payments interface) has emerged as a major issue on a global scale, posing a significant threat to financial institutions and consumers alike. Upi firms have been compelled to invest substantial funds in the development of cutting-edge tools for identifying and preventing fraudulent activities. The primary aim of this study is to develop algorithms that are both cost-effective and adaptable for upi firms to improve their ability to detect fraudulent transactions with greater precision while still saving time and money. Several machine learning algorithms have been rigorously tested and tested in the hopes of achieving this aim. Due to the rise of online transactions, fraudsters in the area of digital payments, including upi, have increased. These deceptive activities have resulted in significant financial losses around the world as a result of phishing, identity theft, and card-not-present (cap) fraud. Upi firms have understood the importance of leveraging advanced technologies, particularly machine learning algorithms, to enhance their fraud detection systems. To determine their effectiveness in detecting and preventing fraudulent transactions, the report includes an extensive review of various machine learning techniques. These algorithms employ a variety of techniques, including supervised learning techniques such as logistic regression, decision trees, and random forests, as well as unsupervised learning techniques such as clustering and anomaly detection. Each algorithm is rigorously tested on a set of predefined performance criteria, considering variables such as precision, recall, and false positive rates.to propose a systematic comparison analysis has revealed that certain machine learning algorithms have superior success at detecting false upi transactions.

VI. ACKNOWLEDGMENT

It's wonderful to hear that Dr. Amit Kumar has been a valuable mentor for your project. Mentorship can indeed play a crucial role in guiding and supporting individuals or teams. However, based on the information provided, it seems like a personal acknowledgment rather than a specific academic or research reference. If you would like to acknowledge Mr. Amit Kumar in a formal research or project document, you might consider something like the following: “Invaluable gratitude goes to Mr. Amit Kumar, our esteemed project mentor, whose unwavering support, insightful guidance, and continuous encouragement have been instrumental in the success of this project. His dedication to fostering learning and providing valuable resources has been a source of inspiration throughout our research journey. “Feel free to customize it based on your specific experiences and the nature of your project.

VII. REFERENCES

1. K.Sim,V.Gopalkrishna,A.Zimek,andG.Cong-"Asurveyonenhancedsubspace clustering,"' Data Mining Knowledge Discovery, vol. 26, no. 2, pp. 332-397,
2. S. McSkimming -'Trade-based money laundering: Responding to an emerging threat," Deakin Law Rev, vol. 15, no. 1,
3. Nitu Kumari, S. Kannan and A. Muthukumaravel "UPI Fraud Detection Using Genetic-A Survey" published by Middle East Journal of Scientific Research, IDOSI Publications,
4. Satvik Vats, Surya Kant Dubey, Naveen Kumar Pandey - "A Tool for Effective Detection of Fraud in UPI System", published in International Journal of Communication Network Security ISSN: 2231 - 1882, Volume-2
5. S.H. Projects and W. Love, -JMU Scholarly Commons Detecting UPI fraud: An analysis of fraud detection techniques, all
6. J. Han, M. Kamber, thiab J. Pei, “Data Mining: Concepts and Techniques,” MorganKaufmann, 2011.
7. D. Wang, F. Nie, thiab H. Huang, “Graph normalized nonnegative matrixfactorization for data sawv cev,” Neural Networks, vol. 58 Ib., p. 31-41, 2014.
8. R. J. Bolton and D. J. Hand, "Statistical Fraud Detection: A Review," Statistical Science, vol. 17, no. 3, pp. 235-249, 2002.

9. ng, F. Nie, thiab H. Huang, “Graph normalized nonnegative matrixfactorization for data sawv cev,” Neural Networks, vol. 58 pm. 3 1-41, 2014.
10. A. Ben Salem, A. R. Halimane, and A. Kachori, "Unsupervised outlier detection using a mixture of von Mises-Fisher distributions," Expert Systems with Applications, vol. 164, p. 113826, 2021.
11. A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," ACM Computing Surveys (CSUR), vol. 31, no. 3, pp. 264-323, 1999.
12. A. S. Ibrahim, H. K. Verma, and R. K. Agrawal, "A survey on credit card fraud detection techniques," Journal of King Saud University-Computer and Information Sciences, vol. 33, no. 2, pp. 182-195, 2021.
13. T. Fawcett and F. Provost, "Activity monitoring: noticing interesting changes in behaviour," in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 1999, pp. 53-62.
14. G. Shmueli, N. R. Patel, and P. C. Bruce, "Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel® with XLMiner®," John Wiley & Sons, 2010.
15. S. Wang, T. Li, S. J. Stolfo, and K. W. Church, "Anomaly detection using profile-based personal classifiers," in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 2003, pp. 8-16.
16. X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu et al., "Top 10 algorithms in data mining," Knowledge and information systems, vol. 14, no. 1, pp. 1-37, 2008.