International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 10, Issue 5, May-2025, http://ijmec.com/, ISSN: 2456-4265

Machine Learning Based UPI Fraud Detection

Nagaraju Daniyal Raju
PG scholar, Department of MCA, CDNR collage, Bhimavaram, Andhra Pradesh.
K.Venkatesh
(Assistant Professor), Master of Computer Applications, DNR collage, Bhimavaram, Andhra Pradesh.

Abstract: ***Fraud is a large scale problem which affects the various entities from public sector to private sectors including government, profit and non-profit organizations. It is hard to predict the exact scale of the fraud because most of the time it remains undetected. It is very important to detect UPI frauds and save the company's or the tax payer's money. The data mining model developed in this research will help organization to analyse their UPI transaction and will blow the early whistle against the fraudsters. In our process, the system is developed to detect the fraud in UPI by using the machine learning algorithm. We predict the values as fraud/non fraud for more accuracy and predict the future fraud. First, we select and view the dataset for future purpose. And we split the data as training data and test data for getting to predict the values. It is essential to train the models on data which includes fraud and relevant non fraud. By using the ML algorithm the system is, to classify the fraud and non-fraud and results shows that the accuracy, precision, recall and f1-score and also prediction. This shows that method used in this project can predict the possibility of fraud accurately in most of the cases. This module is the simple and effective way to avoid such frauds and save those expenditures.***

I. INTRODUCTION

UPI fraud refers to the use of fraudulent and illegal methods or deceptive tactics to gain UPI benefits. Fraud can be committed in different areas of finance, including banking, insurance, taxation, and corporates, and more. Fiscal fraud and evasion, including credit card fraud, tax evasion, UPI statement fraud, money laundry, and other types of UPI fraud, has become a growing problem. Despite efforts to eliminate UPI fraud, its occurrence adversely affects business and society as hundreds of millions of dollars are lost to fraud each year. This significant UPI loss has dramatically affected individuals, merchants, and banks.

Nowadays, fraud attempts have increased drastically, which makes fraud detection more important than ever. The Association of Certified Fraud Examiners (ACFE) has announced that 10% of incidents concerning white-collar crime involves falsification of UPI statements. They classified occupational fraud into three types: asset misappropriation, corruption, and UPI statement fraud. UPI statement fraud resulted in the most significant losses among them.

Although the occurrence frequency of asset misappropriation and corruption is much higher than UPI statement fraud, the UPI implications of these latter crimes are still far less severe. In particular, as reported in a survey from Eisner Amper, which is among the prominent accounting firms in the U.S., ''the average median loss of UPI statement fraud ($800,000 in 2018) accounts for over three times the monetary loss of corruption ($250,000) and seven times as much as asset misappropriation ($114,000)''.

The focus of this study is on UPI statement fraud. UPI statements are documents that describe details about a company, specifically their business activities and UPI performance, including income, expenses, profits, loans, presumable concerns that may emerge later, and managerial comments on the business performance.

All firms are obligated to announce their UPI statements in a quarterly and annual manner. UPI statements can be used to indicate the performance of a company. Investors, market analysts, and creditors exploit UPI reports to investigate and assess the UPI health and earnings potentials of a business. UPI statements consist of four sections; income statement, balance sheet, cash flow statement, and explanatory notes. The income statement places a great emphasis on a company's expenses and revenues during a specific period.

The company's profit or net income is provided in this section, which subtracts expenses from revenues. The balance sheet provides a timely snapshot of liabilities, assets, and stockholders' equity. The cash flow statement measures the extent to which a company is

successful in making cash to fund its operating expenses, fund investments, and pay its debt obligations. Explanatory notes are supplemental data that provide clarification and further information about particular items published UPI statements of a company.

These notes cover areas including disclosure of subsequent events, asset depreciation, and significant accounting policies, which are necessary disclosures that demonstrate the amounts reported on the UPI statements. UPI statement fraud involves falsifying UPI statements to pretend the company more profitable than it is, increase the stock prices, avoid payment of the taxes, or get a bank loan.

Fraud triangle in auditing is a framework to demonstrate the motivation behind an individual's decision to commit fraud. The fraud triangle has three elements that increase the risk of fraud: incentive, rationalization, and opportunity, which, together, lead to fraudulent behaviour. Auditing professionals have extensively used this theory to explain the motivation behind an individual's decision to commit fraud.

It is indispensable to understand the fraud triangle to evaluate UPI fraud. Gupta and Singh suggested that when there are incentives such as the obligation to achieve an outcome or cover losses, the potential for fraud increases. The company will encounter temptations or pressures to adopt fraudulent practices.

II. LITEARTURE SURVEY

2.3.1 UPI fraud detection research: A multidisciplinary analysis,

Author*:* **A. Albizri, D. Appelbaum, and N. Rizzotto**

Methodology

Prior research in the fields of accounting and information systems has shed some light on the significant effects of UPI reporting fraud on multiple levels of the economy. In this paper, we compile prior multi-disciplinary literature on UPI statement fraud detection. UPI reporting fraud detection efforts and research may be more impactful when the findings of these different domains are combined. We anticipate that this research will be valuable for academics, analysts, regulators, practitioners, and investors.

Advantages:

- Reduced Manual power
- Low cost

Disadvantages:

- Too Many False Negatives.
- Run to failure prediction is low.

2.3.2 Interpretable fuzzy rule-based systems for detecting UPI fraud, 2019

Author*:* **P. Hajek**

Methodology

Systems for detecting UPI frauds have attracted considerable interest in computational intelligence research. Diverse classification methods have been employed to perform automatic detection of fraudulent companies. However, previous research has aimed to develop highly accurate detection systems, while neglecting the interpretability of those systems. Here we propose a novel fuzzy rule-based detection system that integrates a feature selection component and rule extraction to achieve a highly interpretable system in terms of rule complexity and granularity. Specifically, we use a genetic feature selection to remove irrelevant attributes and then we perform a comparative analysis of state-of-the-art fuzzy rule-based systems, including FURIA and evolutionary fuzzy rule-based systems. Here, we show that using such systems leads not only to competitive accuracy but also to desirable interpretability. This finding has important implications for auditors and other users of the detection systems of UPI fraud.

Advantages:

- Avoid the over fitting from the dataset.

Disadvantages:

- It can be intimidating.

2.3.3 An application of ensemble random forest classifier for detecting UPI manipulation of Indian listed companies, 2019

Author*:* **H. Patel, S. Parikh, A. Patel, and A. Parikh**

Methodology

A rising incidents of UPI frauds in recent time has increased the risk of investor and other stakeholders. Hiding of UPI losses through fraud or manipulation in reporting and hence resulted into erosion of considerable wealth of their stakeholders. In fact, a number of global companies like WorldCom, Xerox, Enron and number Indian companies such as Satyam, Kingfisher and Deccan Chronicle had committed fraud in UPI by manipulation. Hence, it is imperative to create an efficient and effective framework for detection of UPI fraud. This can be helpful to regulators, investors, governments and auditors as preventive steps in avoiding any possible UPI fraud cases. In this context, increasing number of researchers these days have started focusing on developing systems, models and practices to detect fraud in early stage to avoid the any attrition of investor's wealth and to reduces the risk of financing.

In Current study, the researcher has attempted to explore the various 42 modeling techniques to detect fraud in UPI. To perform the experiment, researcher has chosen 86 FFS and 92 non-fraudulent UPI of manufacturing firms. The data were taken from Bombay Stock Exchange for the dimension of 2008-2011. Auditor's report is considered for classification of FFS and Non-FFS companies. T-test was applied on 31 important UPI ratios and 10 significant variables were taken in to consideration for data mining techniques. 86 FFS and 92 non-FFS during 2008-2017 were taken for testing data set. Researcher has trained the model using data sets. Then, the trained model was applied to the testing data set for the accuracy check. Random forest gives best accuracy. Here, modified random forest model was developed with improved accuracy.

Advantages:

- Change of detecting unknown prediction.
- Fraud Detection more efficient than fraud detection, if fraud detection file is large.

Disadvantages:

- Run to failure prediction is low.
- Reliability is unclear.

2.3.4 Detecting fraudulent UPI for the sustainable development of the socio-economy in China: A multi-analytic approach, 2019

Author*:* **J. Yao, Y. Pan, S. Yang, Y. Chen, and Y. Li**

Methodology

Identifying UPI fraud activities is very important for the sustainable development of a socio-economy, especially in China's emerging capital market. Although many scholars have paid attention to fraud detection in recent years, they have rarely focused on both UPI and non-UPI predictors by using a multi-analytic approach. The present study detected UPI statement fraud activities based on 17 UPI and 7 non-UPI variables by using six data mining techniques including support vector machine (SVM), classification and regression tree (CART), back propagation neural network (BP-NN), logistic regression (LR), Bayes classifier (Bayes) and K-nearest neighbor (KNN). Specifically, the research period was from 2008 to 2017 and the sample is companies listed on the Shanghai stock exchange and Shenzhen stock exchange, with a total of 536 companies of which 134 companies were allegedly involved in fraud. The stepwise regression and principal component analysis (PCA) were also adopted for reducing variable dimensionality. The experimental results show that the SVM data mining technique has the highest accuracy across all conditions, and after using stepwise regression, 13 significant variables were screened and the classification accuracy of almost all data mining techniques was improved. However, the first 16 principal components transformed by PCA did not yield better classification results. Therefore, the combination of SVM and the stepwise regression dimensionality reduction method was found to be a good model for detecting fraudulent UPI .

III. PROPOSED METHOD

In our proposed system, we detect the fraud in UPI statements by using the machine learning algorithm. First, we select and view the imported dataset for future purpose. And we get missing values and fill the default values to the dataset. We

encoding the label in the dataset. And we split the dataset to the Train and Test data for predict the fraud or non-fraud. Then we use three algorithms for more accuracy, prediction and which is more accurate value. There are Random forest algorithm, KNN classifiers and Ada-Boost Algorithm. Now, we fit the training data from the dataset. Then we predict the test dataset using training dataset. Then the test values get the results of actual and predicted. And we get the performance of the dataset. It is essential to train the models on data which includes fraud and relevant non fraud. By using the ML algorithm the system is, to classify the fraud and non-fraud and results shows that the accuracy, precision, recall and f1-score and also prediction. This shows that method used in this project can predict the possibility of fraud accurately in most of the cases. This module is the simple and effective way to avoid such frauds and save those expenditures.

2.2.1 ADVANTAGES

- It is efficient for large number of datasets.
- The experimental result is high when compared with existing system.
- Time consumption is low.
- Provide accurate prediction results.

IV. RESULTS

DATA SELECTION:

```
#--------------------Data Selection--------------#
********************
```

	step	type	amount	...	newbalanceDest	isFraud	isFlaggedFraud
0	1	PAYMENT	NaN	...	0.00	0	0
1	1	PAYMENT	1864.28	...	0.00	0	0
2	1	TRANSFER	NaN	...	0.00	1	0
3	1	CASH_OUT	181.00	...	0.00	1	0
4	1	PAYMENT	11668.14	...	0.00	0	0
5	1	PAYMENT	7817.71	...	0.00	0	0
6	1	PAYMENT	7107.77	...	0.00	0	0
7	1	PAYMENT	7861.64	...	0.00	0	0
8	1	PAYMENT	4024.36	...	0.00	0	0
9	1	DEBIT	5337.77	...	40348.79	0	0
10	1	DEBIT	9644.94	...	157982.12	0	0
11	1	PAYMENT	3099.97	...	0.00	0	0
12	1	PAYMENT	2560.74	...	0.00	0	0
13	1	PAYMENT	11633.76	...	0.00	0	0
14	1	PAYMENT	4098.78	...	0.00	0	0
15	1	CASH_OUT	229133.94	...	51513.44	0	0
16	1	PAYMENT	1563.82	...	0.00	0	0
17	1	PAYMENT	1157.86	...	0.00	0	0
18	1	PAYMENT	671.64	...	0.00	0	0
19	1	TRANSFER	215310.30	...	0.00	0	0

DATA PREPROCESSING

Find Missing Values

```
#--------------------Find missing values--------------------#
********************
step              0
type              0
amount            0
nameOrig          0
oldbalanceOrg     0
newbalanceOrig    0
nameDest          0
oldbalanceDest    0
newbalanceDest    0
isFraud           0
isFlaggedFraud    0
dtype: int64
```

Handling Missing values:

```
#--------------------Fill 0 from missing Values---------------#
********************
step              0
type              0
amount            0
nameOrig          0
oldbalanceOrg     0
newbalanceOrig    0
nameDest          0
oldbalanceDest    0
newbalanceDest    0
isFraud           0
isFlaggedFraud    0
dtype: int64
```

Label Encoding:

```
#--------------------Before Label Encoding---------------#
********************
```

	step	type	amount	...	newbalanceDest	isFraud	isFlaggedFraud
0	1	PAYMENT	0.00	...	0.00	0	0
1	1	PAYMENT	1864.28	...	0.00	0	0
2	1	TRANSFER	0.00	...	0.00	1	0
3	1	CASH_OUT	181.00	...	0.00	1	0
4	1	PAYMENT	11668.14	...	0.00	0	0
5	1	PAYMENT	7817.71	...	0.00	0	0
6	1	PAYMENT	7107.77	...	0.00	0	0
7	1	PAYMENT	7861.64	...	0.00	0	0
8	1	PAYMENT	4024.36	...	0.00	0	0
9	1	DEBIT	5337.77	...	40348.79	0	0
10	1	DEBIT	9644.94	...	157982.12	0	0
11	1	PAYMENT	3099.97	...	0.00	0	0
12	1	PAYMENT	2560.74	...	0.00	0	0
13	1	PAYMENT	11633.76	...	0.00	0	0
14	1	PAYMENT	4098.78	...	0.00	0	0
15	1	CASH_OUT	229133.94	...	51513.44	0	0
16	1	PAYMENT	1563.82	...	0.00	0	0
17	1	PAYMENT	1157.86	...	0.00	0	0
18	1	PAYMENT	671.64	...	0.00	0	0
19	1	TRANSFER	215310.30	...	0.00	0	0

```
#--------------------After Label Encoding---------------#
********************
```

	step	type	amount	...	newbalanceDest	isFraud	isFlaggedFraud
0	1	3	0.00	...	0.00	0	0
1	1	3	1864.28	...	0.00	0	0
2	1	4	0.00	...	0.00	1	0
3	1	1	181.00	...	0.00	1	0
4	1	3	11668.14	...	0.00	0	0
5	1	3	7817.71	...	0.00	0	0
6	1	3	7107.77	...	0.00	0	0
7	1	3	7861.64	...	0.00	0	0
8	1	3	4024.36	...	0.00	0	0
9	1	2	5337.77	...	40348.79	0	0
10	1	2	9644.94	...	157982.12	0	0
11	1	3	3099.97	...	0.00	0	0
12	1	3	2560.74	...	0.00	0	0
13	1	3	11633.76	...	0.00	0	0
14	1	3	4098.78	...	0.00	0	0
15	1	1	229133.94	...	51513.44	0	0
16	1	3	1563.82	...	0.00	0	0
17	1	3	1157.86	...	0.00	0	0
18	1	3	671.64	...	0.00	0	0
19	1	4	215310.30	...	0.00	0	0

DATA SPLITTING:

```
#--------------------Data Splitting------
********************
Total no of dataset : (80000, 11)
Training set Without Target (64000, 10)
Training set only Target (64000,)
Testing set Without Target (16000, 10)
Testing set only Target (16000,)
```


ISSN: 2456-4265
IJMEC 2025

CLASSIFICATION:

```
#--------------------Random Forest Algorithm---------------#
********************
Matrix:
[[15976     0]
 [   12    12]]
classfication:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     15976
           1       1.00      0.50      0.67        24

    micro avg       1.00      1.00      1.00     16000
    macro avg       1.00      0.75      0.83     16000
weighted avg       1.00      1.00      1.00     16000

Accuracy:  99.925
```

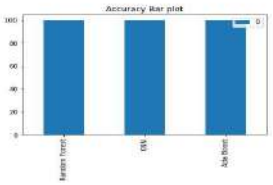
```
#--------------------KNN Algorithm---------------#
********************

Matrix:
[[15975     1]
 [   24     0]]
classfication:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     15976
           1       0.00      0.00      0.00        24

    micro avg       1.00      1.00      1.00     16000
    macro avg       0.50      0.50      0.50     16000
weighted avg       1.00      1.00      1.00     16000

Accuracy:  99.84375
```

```
#--------------------Ada Boost---------------#
********************
0.999

Matrix:
[[15973     3]
 [   13    11]]
classfication:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     15976
           1       0.79      0.46      0.58        24

    micro avg       1.00      1.00      1.00     16000
    macro avg       0.89      0.73      0.79     16000
weighted avg       1.00      1.00      1.00     16000

Accuracy:  99.9
```

PREDICTION:

GRAPH:

```
#--------------------Camparision between 3 Algorithm Accuracy---------------#
********************
```



V. CONCLUSION

In this project, we propose an approach to utilise the Random Forest algorithm, KNN and Adaboost algorithm for fraud detection in UPI statements. We call the approach the three algorithms on datasets with significantly reduced dimensionality. The Classifications classifier gives high accuracy results that are comparable or superior to other fraud detection techniques in spite of working with reduced data and also compared with graph.

REFERENCES

1. Albizri, D. Appelbaum, and N. Rizzotto, ‘‘Evaluation of UPI statements fraud detection research: A multi-disciplinary analysis,’’ Int. J. Discl. Governance, vol. 16, no. 4, pp. 206–241, Dec. 2019.
2. R.Albright, ‘‘Taming text with the SVD.SAS institute white paper, ’’SAS Inst., Cary, NC, USA, White Paper 10.1.1.395.4666, 2004.
3. M. S. Beasley, ‘‘An empirical analysis of the relation between the board of director composition and UPI statement fraud,’’ Accounting Rev., vol. 71, pp. 443–465, Oct. 1996.
4. T. B. Bell and J. V. Carcello, ‘‘A decision aid for assessing the likelihood of fraudulent UPI reporting,’’ Auditing A, J. Pract. Theory, vol. 19, no. 1, pp. 169–184, Mar. 2000.
5. M.D.BeneishandC.Nichols,‘‘The predictable cost of earnings manipulation,’’Dept.Accounting,KelleySchoolBus.,IndianaUniv.,Bloomington, IN, USA, Tech. Rep. 1006840, 2007.
6. R. J. Bolton and D. J. Hand, ‘‘Statistical fraud detection: A review,’’ Stat. Sci., vol. 17, no. 3, pp. 235–249, Aug. 2002.
7. M.Cecchini, H.Aytug, G.J.Koehler, and P.Pathak,‘‘Making words work: Using UPI text as a predictor of UPI events,’’ Decis. Support Syst., vol. 50, no. 1, pp. 164–175, 2010.
8. Q. Deng, ‘‘Detection of fraudulent UPI statements based on naïve Bayes classifier,’’ in Proc. 5th Int. Conf. Comput. Sci. Educ., 2010, pp. 1032–1035.
9. S. Chen, Y.-J.-J. Goo, and Z.-D. Shen, ‘‘A hybrid approach of stepwise regression, logistic regression, support vector machine, and decision tree for forecasting fraudulent UPI statements,’’ Sci. World J., vol. 2014, pp. 1–9, Aug. 2014.
10. X. Chen and R. Ye, ‘‘Identification model of logistic regression analysis on listed Firms’ frauds in China,’’ in Proc. 2nd Int. Workshop Knowl. Discovery Data Mining, Jan. 2009, pp. 385–388.
11. Chimonaki, S. Papadakis, K. Vergos, and A. Shahgholian, ‘‘Identification of UPI statement fraud in greece by using computational intelligencetechniques,’’inProc.Int.WorkshopEnterpriseAppl.,Markets Services Finance Ind. Cham, Switzerland: Springer, 2018, pp. 39–51.
12. R. Cressey, ‘‘Other people’s money; a study of the social psychology of embezzlement,’’ Amer. J. Sociol., vol. 59, no. 6, May 1954, doi: 10.1086/221475.
13. B. Dbouk and I. Zaarour, ‘‘Towards a machine learning approach for earningsmanipulationdetection,’’AsianJ.Bus.Accounting,vol.10,no.2, pp. 215–251, 2017.
14. Q. Deng, ‘‘Application of support vector machine in the detection of fraudulent UPI statements,

ISSN: 2456-4265
IJMEC 2025

''inProc.4thInt.Conf.Comput.Sci.Educ., Jul. 2009, pp. 1056–1059.

15. S. Chen, ''Detection of fraudulent UPI statements using the hybrid data mining approach,'' SpringerPlus, vol. 5, no. 1, p. 89, Dec. 2016.