PRUDENT FRAUD DETECTION IN INTERNET BANKING USING MACHINE LEARNING

[1] Mrs.Ch.Sindhu Priyanka, M.Tech, Assistant Professor, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.
[2] N.Lavanya, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.
[3] P.Lakshmi Manisha, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.
[4] B.Priyanka, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.
[5] B.Suhitha, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

Abstract: Most commercial Fraud Detection components of Internet banking systems use some kind of hybrid setup usually comprising a Rule-Base and an Artificial Neural Network. Such rule bases have been criticized for a lack of innovation in their approach to Knowledge Acquisition and maintenance. Furthermore, the systems are brittle; they have no way of knowing when a previously unseen set of fraud patterns is beyond their current knowledge. This limitation may have far reaching consequences in an online banking system. This project presents a viable alternative to brittleness in Knowledge Based Systems; a potential milestone in the rapid detection of unique and novel fraud patterns in Internet banking using machine learning. The experiments conducted with real online banking transaction log files suggest that Prudent based fraud detection may be a worthy alternative in online banking.

1. INTRODUCTION

According to The American Heritage dictionary, second college edition, fraud is defined as a deception deliberately practiced to secure unfair unlawful gain. Fraud detection is the recognition of symptoms of fraud where no prior suspicion or tendency to fraud exists. Examples include insurance fraud, credit card fraud and accounting fraud. Data from the Nigeria Inter-Bank Settlement System (NIBSS) has revealed that fraudulent transactions in the banking sector at its peak. Fraud has evolved from being committed by casual fraudsters to being committed by organized crime and fraud rings that use sophisticated methods to take over control of accounts and commit fraud. Some 6.8 million Americans were victimized by card fraud in 2007, according to Javelin research. Such fraud on existing accounts accounted for more than $3 billion in losses in 2007. The Nilson Report estimates the cost to the industry to be $4.84 billion. Javelin estimates the losses at more than six times that amount – some $30.6 billion in 2007. Of course, fraud is not a domestic product as it‘s everywhere. For instance, card fraud losses cost UK economy GBP 423 million in 2006. Credit card fraud accounts for the biggest cut of the $600 million that airlines lose each year globally. The motivation behind the project "PRUDENT FRAUD DETECTION IN INTERNET BANKING USING MACHINE LEARNING "is to prevent fraud by using Machine learning algorithms and provide the user a fraud free transaction experience. It can detect the fraud before it happens by using classification techniques. The prudence part of the system is engaged each time there are inconsistencies between the two components, and a warning is issued. Machine learning (ML) algorithms can detect and prevent financial fraud operations by looking at historical data and identifying trends and patterns connected to fraud. Predictive analytics enhances fraud detection by analyzing historical data to identify patterns and trends linked to fraudulent transaction.

2. LITERATURE SURVEY

R.Rajkumar,N.Kogila“Intelligent System for Fraud Detection in Online Banking using Improved Particle Swarm Optimization and Support Vector Machine” August2023[1]. The purpose of this research is to suggest a strategy for employing IPSO-SVM to detect and prevent financial fraud in the digital sphere. This investigation introduces an improved particle swarm optimization of support vector machine (IPSO-SVM) technique model for fraud detection by combining optimized particle swarm optimization (IPSO) and support vector machine (SVM). The proposed approach outperforms other two models such as CNN and SVM.2023 8th International Conference on Communication and Electronics Systems (ICCES)

S. Hossain, A. Abtahee, I. Kashem, M. M. Hoque, and I. H. Sarker, "Crime Prediction Using Spatio-Temporal Data," in Computing Science, Communication and Security, Singapore, 2020, pp. 277-289: Springer Singapore.[2] A crime is an action which constitutes a punishable offence by law. It is harmful for society so as to prevent the criminal activity, it is important to understand crime. Data driven researches are useful to prevent and solve crime. Recent research shows that 50% of the crimes are committed by only handful of offenders. The law enforcement officers need early information about the criminal activity to response and solve the spatio-temporal criminal activity.

M. Zamini and S. M. H. Hasheminejad, "A comprehensive survey of anomaly detection in banking, wireless sensor networks, social networks, and healthcare," (in English), Intelligent Decision Technologies-Netherlands, Article vol. 13, no. 2, pp. 229-270, 2019.[3] . This survey aims to create a structured and comprehensive overview of the research on anomaly detection. First, we tried to introduce the concept of anomalies and types of anomaly detection. We have tried to classify anomaly detection according to their application and then categorized their techniques. For each application and technique, we have described key assumptions, For each application, a basic anomaly detection technique has been provided, in the end; the differences among existing techniques in each specific category are discussed better concept of the various directions, which has been researched on that specific topic.Transaction Detection and Analysis Model to Prevent Illegal Money Transfer Through E-Banking Channels.

Gonzalez-Carrasco, J. L. Jimenez-Marquez, J. L. Lopez-Cuadrado, and B. Ruiz Mezcua, "Automatic detection of relationships between banking operations using machine learning," (in English), Information Sciences, Article vol. 485, pp. 319-346, Jun 2019[4]. This paper proposed a system to detect bank fraud using a community detection algorithm that identifies the patterns that can lead to fraud occurrences. An agile method was used to design the web-based application to detect the fraud. The application functioned as a central hub between the banks and customers. Neo4j, a graph database, was used for creating and representing the database, and the Cypher query was used as a graph query language. The proposed system successfully detected all frauds presented during the test experiment. This paper will assist bankers to combat fraud by detecting and preventing similar occurrences.

M. Pohoretskyi, D. Serhieieva, and Z. Toporetska, "The proof of the event of a financial resources fraud in the banking sector: problematic issues," (in English), Financial and Credit Activity-Problems of Theory and Practice, Article vol. 1, no. 28, pp. 36-45, 2019[5]. The object of the study is the process of assessing the risks and economic benefits of involving state-owned banks to the organization and conduct of lotteries. Based on the research conducted by the authors, the inefficiency and high riskiness of the combination of banks and lottery activities, as well as its economic disadvantage for the state, is proved by the authors. The main risk from the combination of these types of activities is that the bank can use the money attracted from the population not for their direct purpose, in particular: to use funds from the banking activity and the needs of the organization of lotteries or payouts, or to use funds of the prize money to goals other than payouts.

Gopinath Muruti; Fiza Abdul Rahim; Zul-Azri bin Ibrahim "A Survey on Anomalies Detection Techniques and Measurement Methods"November 2018[6]. Dynamic research area has been applied and researched on anomaly detection in various domains. And various techniques have been proposed to identify unexpected items or events in datasets which differ from the norm. This review tries to provide a basic and structured overview of the anomaly detection techniques. Also, this review discusses major anomaly detection techniques using statistical based and machine learning based techniques. The outcome of this review may facilitate a better understanding of the different techniques in which research has been done on this topic by comparing the pros and cons of the identified techniques. In addition, this review also discusses on the measurement methods used by other researchers in validating their anomalies detection techniques.

K. Noor et al., "Performance analysis of a surveillance system to detect and track vehicles using Haar cascaded classifiers and optical flow method," 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), Siem Reap, 2017, pp. 258 263.[7]. In this study, image processing and machine learning techniques are used to detect anomalies in vehicle movements. These anomalies include standing and traveling in reverse direction. Images are captured using CCTV cameras from front and rear side of the vehicle. This capability makes the results robust to the variations in operational and environmental conditions. Multiple consecutive frames are acquired for motion detection. Features such as edges and license plate corner locations are extracted for tracking purposes. Direction of the traffic flow is obtained from the trained classifier. K-nearest neighbor is chosen as the classifier model.

3. EXISTING SYSTEM

In case of bank fraud detection, the existing system is detecting the fraud after fraud has been happen. Existing system maintain the large amount of data when customer comes to know about inconsistency in transaction, he/she made complaint and then fraud detection system start it working. It first tries to detect that fraud has actually occur after that it transactions that was used to fraud detection mechanism developed by master and visa cardsA machine learning paradigm classification, with Bank Fraud Detection being the base. Intrusion detections to track fraud location and so on. In case of existing system there is no confirmation of recovery of fraud and Customer satisfaction. Secure electronic system used to analyze the behavior of legitimate users. Data Mining mechanisms to classify and preprocess the user‘s data Genetic algorithms.

DISADVANTAGES

- Each payment system has its limits regarding the maximum amount in the account the number of transactions per day and the amount of output.
- If Internet connection fails, you cannot get to your online account.
- If you follow the security rules the threat is minimal. The worse situation when the system of processing company has been broken because it leads to the leak of personal data on cards and its owners.
- The information about all the transactions, including the amount, time and recipient are stored in the database of the payment system. And it means the intelligence agency has

access to this information. Sometimes this is the path for fraudulent activities.

4. PROPOSED SYSTEM

In proposed methodology, Detection of fraudulent activity is thus critical to control these costs. This paper hereby addresses bank fraud detection via the use of machine learning techniques; association, clustering, forecasting, and classification to analyze the customer data to identify the patterns that can lead to frauds. Upon identification of the patterns, adding a higher level of verification / authentication to banking processes can be added. These kinds of frauds can be credit card fraud, insurance fraud, accounting fraud, etc. which may lead to the financial loss to the bank or the customers. Thus, detection of these kinds of frauds are very important. Fraud detection in banking sector is based on the machine learning techniques and their collective analysis from the past experiences and the probability of how the fraudsters can steal from customers and banks. Therefore, this paper addresses the analysis of data mining techniques of how to detect frauds and overcoming it in banking sector.

ADVANTAGES

- To eliminate real time fraud to the lowest level.
- To increase the confidence of customers in the banking system especially for online transactions.
- To discourage fraudsters (both present and intending ones)

SYSTEM ARCHITECTURE

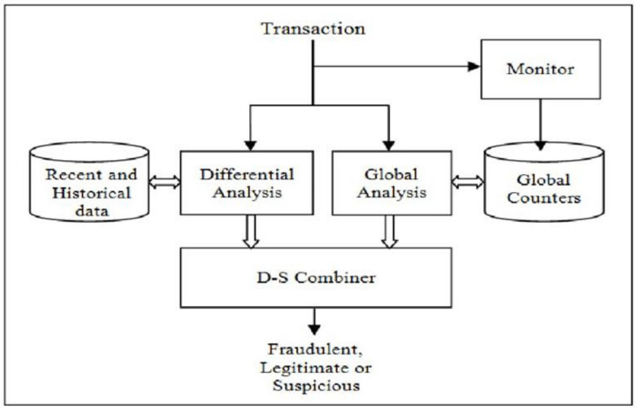


Fig1: System Architecture

5. UML DIAGRAMS

1. USECASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted

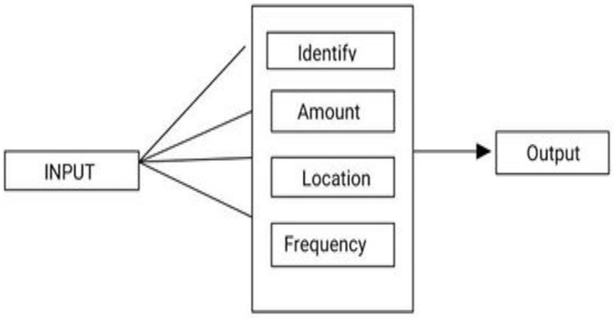


Fig 5.1 shows the Use case Diagram

SEQUENCE DIAGRAM

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. A sequence diagram shows, as parallel vertical lines ("lifelines"), different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchanged between them, in the order in which they occur. This allows the specification of simple runtime scenarios in a graphical manner.

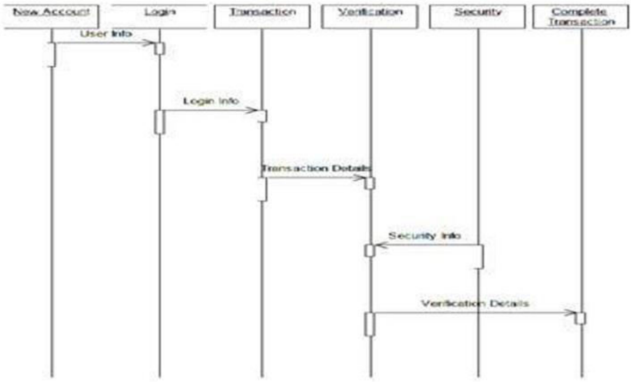
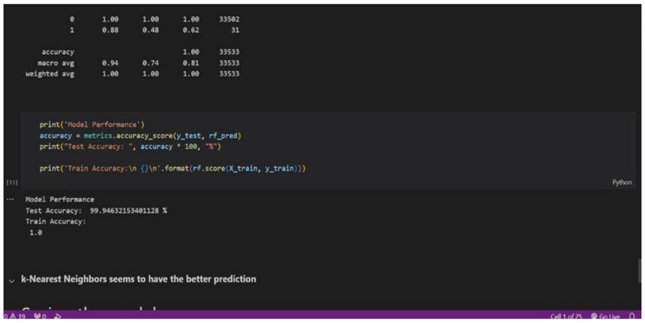


Fig 5.2 shows the Sequence Diagram

6. RESULTS

6.1 Output Screens

Hence, we use Machine Learning for detecting fraud. Here, a machine tries to learn by itself and becomes better by experience. Also, it is an efficient way of detecting fraud because of its fast computing. It does not even require the guidance of a fraud analyst. It helps in reducing false positives for transactions as the patterns are detected by an automated system for streaming transactions that are in huge volume.



```
           0       1.00      1.00      1.00     33502
           1       0.88      0.48      0.62        31

    accuracy                           1.00     33533
   macro avg       0.94      0.74      0.81     33533
weighted avg       1.00      1.00      1.00     33533
```

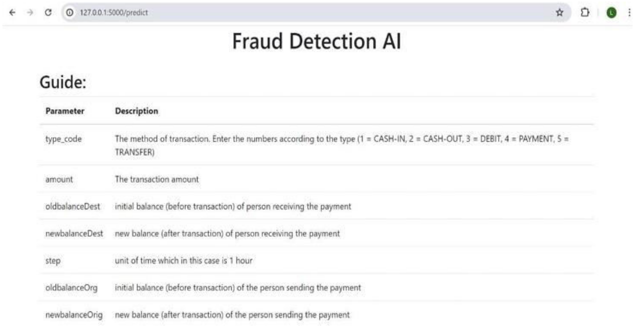
```
print('Model Performance')
accuracy = metrics.accuracy_score(y_test, rf_pred)
print("Test Accuracy: ", accuracy * 100, "%")

print('Train Accuracy:\n {}\n'.format(rf.score(X_train, y_train)))
```

```
Model Performance
Test Accuracy:  99.94632153401128 %
Train Accuracy:
 1.0
```

k-Nearest Neighbors seems to have the better prediction

Fig 6.1 Trained Accuracy



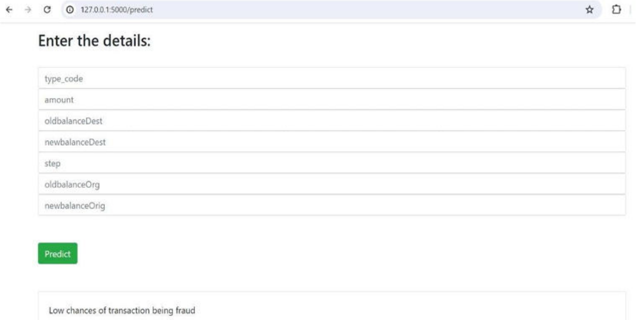
127.0.0.1:5000/predict

Fraud Detection AI

Guide:

Parameter	Description
type_code	The method of transaction. Enter the numbers according to the type (1 = CASH-IN, 2 = CASH-OUT, 3 = DEBIT, 4 = PAYMENT, 5 = TRANSFER)
amount	The transaction amount
oldbalanceDest	initial balance (before transaction) of person receiving the payment
newbalanceDest	new balance (after transaction) of person receiving the payment
step	unit of time which in this case is 1 hour
oldbalanceOrg	initial balance (before transaction) of the person sending the payment
newbalanceOrig	new balance (after transaction) of the person sending the payment

Fig 6.2 Fraud Detection of AI



127.0.0.1:5000/predict

Enter the details:

type_code
amount
oldbalanceDest
newbalanceDest
step
oldbalanceOrg
newbalanceOrig

Low chances of transaction being fraud

Fig 6.3 Fraud Detection Analysis

7. CONCLUSION

In conclusion, our study represents a significant step forward in addressing the fraud detection in internet banking through the utilization of advanced machine learning techniques .Machine learning is a technique used to extract vital information from existing huge amount of data and enable better decision-making for the banking and retail industries. They use data warehousing to combine various data from databases into an acceptable format so that the data can be mined. The data is then analyzed and the information that is captured is used throughout the organization to support decision-making. Data Mining techniques are very useful to the banking sector for better targeting and acquiring new customers, most valuable customer retention, automatic credit approval which is used for fraud prevention, fraud detection in real time, providing segment based products, analysis of the customers, transaction patterns over time for better retention and relationship, risk management and marketing.

FUTURE SCOPE

Machine learning has immense potential in improving fraud detection in internet banking. In the future, we can focus on developing more sophisticated algorithms that can learn from vast amounts of data to identify patterns and anomalies associated with fraudulent activities. Future work includes testing the IPA with more transaction from an online banking system. The research work and experiments completed so far suggest that RDR Prudence presents a potentially useful fraud detection method in commercial Internet banking systems. By continuously training these models with new data, we can enhance their accuracy and adaptability. Additionally, exploring the utilization of deep learning techniques, such as neural networks, can help uncover hidden patterns and detect complex fraud schemes. The key is to keep refining and optimizing these algorithms to stay one step ahead of fraudsters. It's an exciting field with endless possibilities.

8. REFERENCES

[1] R.Rajkumar,N.Kogila“Intelligent System for Fraud Detection in Online Banking using Improved Particle Swarm Optimization and Support Vector Machine” August2023.

[2] S. Hossain, A. Abtahee, I. Kashem, M. M. Hoque, and I. H. Sarker, "Crime Prediction Using Spatio-Temporal Data," in Computing Science, Communication and Security, Singapore, 2020, pp. 277-289: Springer Singapore.

[3] M. Zamini and S. M. H. Hasheminejad, "A comprehensive survey of anomaly detection in banking, wireless sensor networks, social networks, and healthcare," (in English), Intelligent Decision Technologies-Netherlands, Article vol. 13, no. 2, pp. 229-270, 2019.

[4] I. Gonzalez-Carrasco, J. L. Jimenez-Marquez, J. L. Lopez-Cuadrado, and B. Ruiz-Mezcua, "Automatic detection of relationships between banking operations using machine learning," (in English), Information Sciences, Article vol. 485, pp. 319-346, Jun 2019.

[5] M. Pohoretskyi, D. Serhieieva, and Z. Toporetska, "The proof of the event of a financial resources fraud in the banking sector: problematic issues," (in English), Financial and Credit Activity-Problems of Theory and Practice, Article vol. 1, no. 28, pp. 36-45, 2019.

[6] Gopinath Muruti; Fiza Abdul Rahim; Zul-Azri bin Ibrahim “A Survey on Anomalies Detection Techniques and Measurement Methods”November 2018.

[7] K. Noor et al., "Performance analysis of a surveillance system to detect and track vehicles using Haar cascaded classifiers and optical flow method," 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), Siem Reap, 2017, pp. 258-263.

[8] Si-Wan Yoo Journal of the korea institute of information security and cryptology “study on real time-based suspicious transaction detection and analysis model to prevent illegal money transfer through E-banking channels” december 2016

[9] Michele Carminati. Roberto Caron. Federico Maggi. Illenia Epifani “A Decision support system for online banking fraud analysis and investigation” Apirl 2015.

[10] Rohulla Kosari Langari.Nasrolla Moghaddam.Davood Vahdat “Introducing a model for suspicious behaviors detection in electronic banking by using decision tree.march 2013.