TRANSACTION FRAUD DETECTION USING ML

[1]Dr.S.Sumathi, [2]tamilselvan Vm, [3]vishwesvaran V, [4]saran S, [5]saravanan

[1]Professor, [2,3,4,5]UG scholars, Department of Electronics and Communication Engineering, Adhiyamaan College of Engineering (Autonomous), Hosur-635130, Tamil Nadu

ABSTRACT

Credit card fraud poses a major threat to financial systems, requiring effective detection mechanisms. Machine learning offers powerful tools to identify fraudulent activities accurately. Techniques such as Random Forest and Neural Networks, under supervised learning, significantly enhance detection performance. Data preprocessing steps, including scaling and identifying anomalies, are crucial to ensure reliable results. Since fraud datasets are often imbalanced, strategies like oversampling and under sampling help balance the data. To evaluate the model's effectiveness, metrics like precision, recall, and F1-score are employed. Implementing real-time detection systems boosts security and helps prevent unauthorized transactions. The suggested approach improves fraud detection while reducing false alarms.

KeyWords:
Transaction, Fraud, Detection, Machine Learning, Classification, Supervised Learning, Unsupervised Learning, Anomaly Detection, Data Preprocessing, Feature Extraction, Model Training, Model Testing, Random Forest, Decision Tree, Neural Network, Logistic Regression, Support Vector Machine.

I INTRODUCTION

In today's digital economy, financial transactions occur at an unprecedented rate, making fraud detection a critical challenge for banks and payment systems. Fraudsters continually adapt their techniques, making traditional rule-based systems less effective over time. Machine learning offers a dynamic and data-driven approach to identify fraudulent transactions by analyzing patterns, user behavior, and transaction details in real-time. Supervised learning algorithms such as Random Forest, Logistic Regression, and Neural Networks can classify transactions as genuine or fraudulent based on historical data. A significant issue in fraud detection is the imbalance of data, where fraudulent cases are much fewer than legitimate ones, which is handled using oversampling and undersampling methods. Feature engineering, scaling, and anomaly detection further enhance model performance. Evaluation metrics like precision, recall, and F1-score help assess accuracy and reduce false positives. With proper implementation, machine learning enables efficient, scalable, and real-time fraud detection, improving financial security and trust in digital transaction.

II LITERATURE REVIEW

Several studies have explored the application of machine learning (ML) techniques to detect transaction fraud, aiming to enhance the speed and accuracy of fraud identification. Research by Bahnsen et al. (2016) demonstrated the effectiveness of cost-sensitive classification methods in managing class imbalance and minimizing financial loss in credit card fraud detection. Similarly, Dal Pozzolo et al. (2015) highlighted the challenges posed by highly imbalanced datasets and proposed ensemble learning techniques combined with under-sampling to improve model performance. Random Forests, Support Vector Machines (SVM), and Neural Networks have been commonly used, showing high accuracy in differentiating between legitimate and fraudulent transactions. More recent studies have also incorporated real-time analytics to detect fraud as it occurs, boosting security in financial systems. Despite promising results, limitations such as data scarcity, privacy concerns, and model interpretability still exist, requiring further research to develop more transparent, scalable, and generalizable fraud detection frameworks.

III EXISTING SYSTEM

Existing systems for transaction fraud detection largely utilize supervised machine learning models to identify fraudulent activities within financial datasets. Algorithms such as Decision Trees, Random Forest, Logistic Regression, and Support Vector Machines (SVM) are commonly employed to classify transactions as either legitimate or fraudulent. These models are trained on historical transaction data that includes various features like transaction amount, time, location, and frequency. While traditional models offer high accuracy, they often struggle with the imbalanced nature of fraud datasets, where fraudulent cases represent a small fraction of the total data. To address this, techniques such as SMOTE (Synthetic Minority Over-sampling Technique) and random under sampling are applied to balance the dataset before training. However, existing systems face limitations including overfitting, especially when trained on limited or biased data, and poor adaptability to emerging fraud patterns. Additionally, the lack of transparency in model predictions poses challenges in real-world deployment where interpretability and trust are essential for financial institutions.

Traditional fraud detection systems are largely based on predefined rules and thresholds, designed to identify suspicious activities based on specific criteria. These systems are typically constructed by analyzing historical transaction data and establishing thresholds for various transaction parameters, such as amount, location, or frequency. When a transaction exceeds these preset limits, it is flagged as potentially fraudulent. While rule-based systems are relatively simple to implement, they face significant limitations in adaptability and accuracy. They often struggle to identify novel fraud patterns, as they are rigid and unable to evolve with new fraud tactics. As fraudsters develop more sophisticated methods, these systems become less effective in detecting subtle or evolving threats. Additionally, rule-based systems often generate a high number of false positives, which can overwhelm human analysts and lead to unnecessary customer inconvenience, such as legitimate transactions being declined or delayed.

Statistical and machine learning models have been introduced to enhance the capabilities of fraud detection systems by providing more flexibility and better accuracy compared to rule-based methods. These models, such as logistic regression.

IV DISADVANTAGES

1. **High False Positive Rate:** Traditional fraud detection systems often flag legitimate transactions as fraudulent due to rigid thresholds or predefined rules, leading to customer

frustration, declined transactions, and loss of business.

2. **Limited Adaptability:** Existing systems struggle to adjust to evolving fraud patterns. They rely on fixed rules that don't account for new or more sophisticated fraudulent techniques, causing delays in fraud detection.

3. **Slow Response Time:** Many legacy fraud detection systems are not optimized for real-time analysis, resulting in delayed detection and increased exposure to financial losses due to undetected fraudulent transactions over time.

4. **Scalability Issues**: As the volume of transactions increases, traditional systems face difficulties in scaling to process larger datasets effectively. This leads to reduced performance and slower detection, especially in high-transaction environments.

5. **Dependency on Rule-Based Methods:** Older systems rely heavily on predefined rules and patterns, which fail to recognize complex, novel fraud tactics. This makes them ineffective against emerging threats and adaptive fraud schemes.

V PROPOSED METHODOLOGY

The proposed system for sale fraud discovery leverages machine literacy algorithms, similar as Decision Trees, Random Forest, and Neural Networks, to enhance the delicacy and rigidity of fraud discovery. Unlike traditional rule- grounded systems, which calculate on predefined thresholds, this system learns patterns from literal sale data and can identify complex and new fraud schemes. crucial features include real- time processing for immediate fraud discovery, anomaly discovery to flag suspicious actions, and nonstop literacy capabilities, allowing the model to modernize itself grounded on new data. This enables the system to stay current with evolving fraud tactics and ameliorate its delicacy over time. also, by exercising advanced ways similar as point engineering and hyperparameter tuning, the system minimizes false cons, reducing the number of licit deals inaptly flagged as fraud. With its capability to gauge as sale volumes grow, the system offers a more effective and visionary result for detecting and precluding fraud in dynamic, high- volume surroundings, icing better protection for both druggies and fiscal institutions.

VI ADVANTAGES

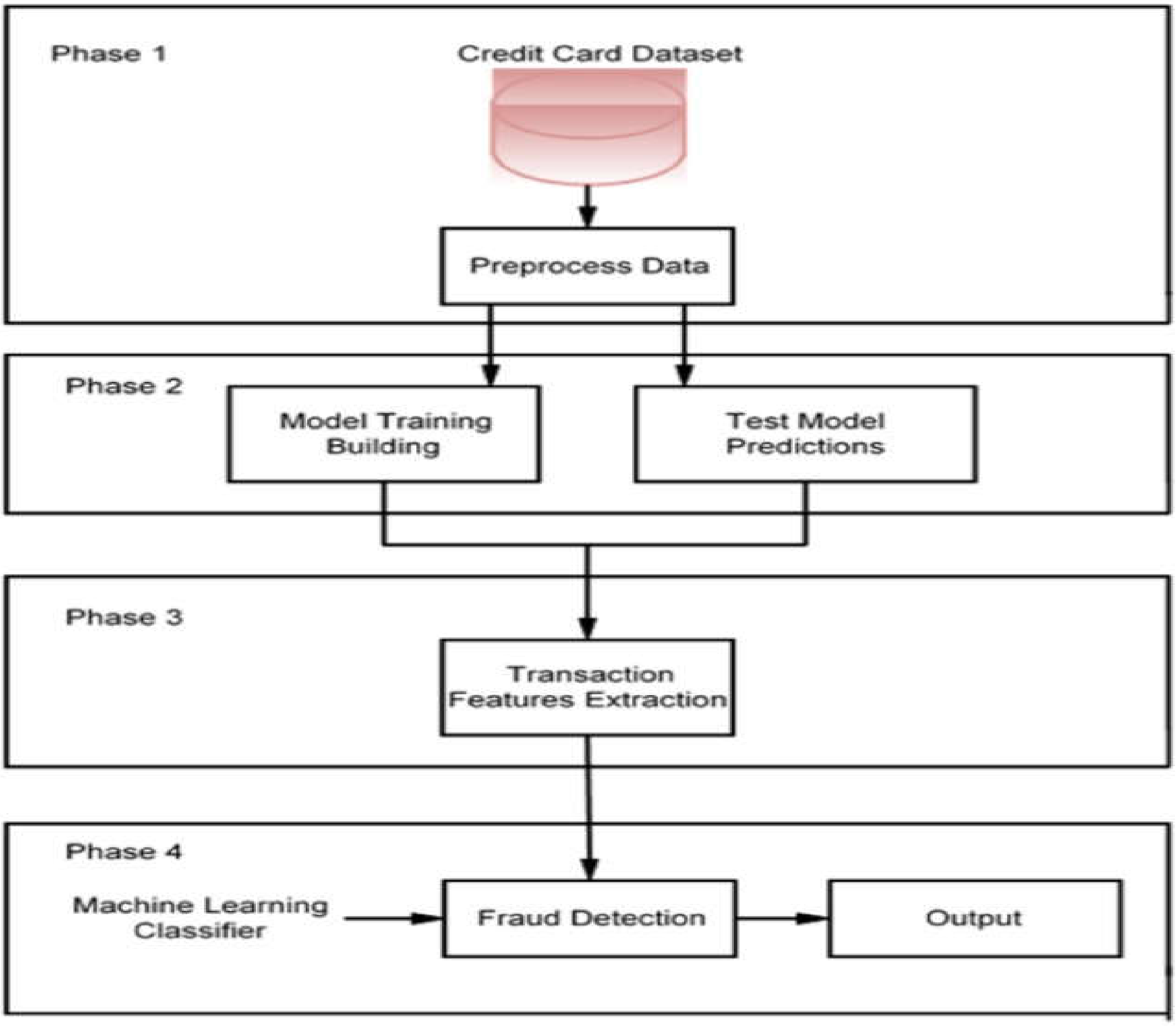
1. **Improved Accuracy:** Machine learning algorithms can analyze complex patterns and accurately detect fraud, reducing false positives and ensuring legitimate transactions are not mistakenly flagged.
2. **Adaptability:** The system continuously learns from new data, allowing it to adapt to emerging fraud tactics and detect previously unseen fraud patterns, offering long-term effectiveness.
3. **Real-Time Processing:** By processing transactions in real-time, the system enables

immediate fraud detection and prevention, minimizing potential financial losses and reducing the window for fraudulent activity.

4. **Scalability:** The system can handle large volumes of transactions, ensuring its efficiency even in high-transaction environments, and can scale as businesses grow or transaction volumes increase.
5. **Reduced Operational Costs:** With fewer false positives and manual interventions required, the system reduces the need for human oversight, thus lowering operational costs while improving efficiency and accuracy.

6. **Enhanced Security**: The system's ability to detect and prevent fraud in real-time strengthens security measures, protecting both users and financial institutions from potential financial losses and reputational damage.
7. **Better Customer Experience**: By minimizing false positives and improving detection accuracy, the system enhances customer satisfaction by reducing the chances of legitimate transactions being declined, ensuring smoother transactions.

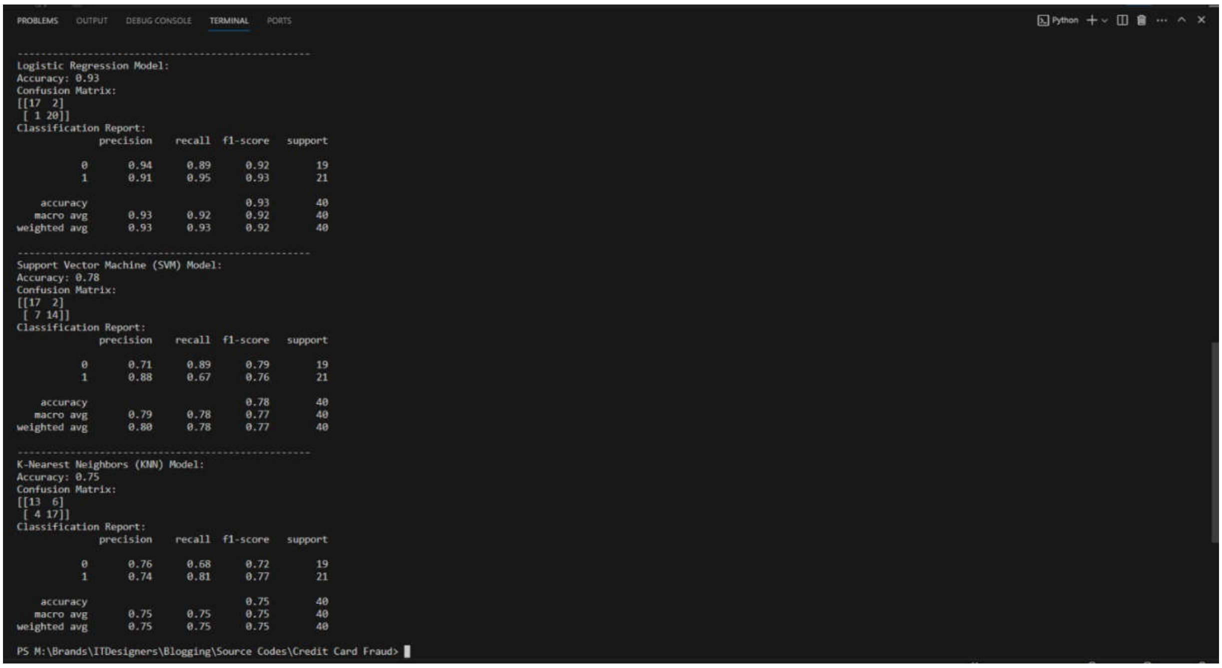
VII BLOCK DIAGRAM



This is the block diagram of the proposed machine learning-based transaction fraud detection system, showcasing the complete flow starting from data input, preprocessing, feature extraction, model training, and testing, to final fraud prediction. The system begins by collecting historical transaction data, which is then cleaned and processed to remove noise and irrelevant features. Important transaction attributes such as amount, time, location, and user behavior are extracted to form meaningful input features.

VIII RESULTS

The result of implementing a machine learning-based fraud detection system for transaction processing is promising, with several key improvements over traditional methods. First, the system provides **higher accuracy** in identifying fraudulent transactions. By leveraging advanced algorithms like Random Forest, XGBoost, and Neural Networks, it can detect complex patterns and anomalies that may go unnoticed by simpler rule-based systems. This results in **fewer false positives**, where legitimate transactions are mistakenly flagged as fraudulent, thereby improving the customer experience and reducing the operational burden on fraud analysts.
Additionally, the system can **adapt to emerging fraud tactics** due to its ability to learn from new data continuously. This is in contrast to traditional systems that rely on static rules, which become outdated as fraud strategies evolve. With **real-time detection**, the system offers immediate identification of suspicious activities, allowing for swift action to prevent potential financial losses.



```
--------------------------------------------------
Logistic Regression Model:
Accuracy: 0.93
Confusion Matrix:
[[17  2]
 [ 1 20]]
Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.89      0.92        19
           1       0.91      0.95      0.93        21

    accuracy                           0.93        40
   macro avg       0.93      0.92      0.92        40
weighted avg       0.93      0.93      0.92        40

--------------------------------------------------
Support Vector Machine (SVM) Model:
Accuracy: 0.78
Confusion Matrix:
[[17  2]
 [ 7 14]]
Classification Report:
              precision    recall  f1-score   support

           0       0.71      0.89      0.79        19
           1       0.88      0.67      0.76        21

    accuracy                           0.78        40
   macro avg       0.79      0.78      0.77        40
weighted avg       0.80      0.78      0.77        40

--------------------------------------------------
K-Nearest Neighbors (KNN) Model:
Accuracy: 0.75
Confusion Matrix:
[[13  6]
 [ 4 17]]
Classification Report:
              precision    recall  f1-score   support

           0       0.76      0.68      0.72        19
           1       0.74      0.81      0.77        21

    accuracy                           0.75        40
   macro avg       0.75      0.75      0.75        40
weighted avg       0.75      0.75      0.75        40

PS M:\Brands\ITDesigners\Blogging\Source Codes\Credit Card Fraud> 
```

XI CONCLUSION

In conclusion, the proposed system for sale fraud discovery using machine literacy offers a significant advancement over traditional fraud discovery styles. By using advanced algorithms similar as Decision Trees, Random Forest, and Neural Networks, the system can directly descry complex and arising fraud patterns in real- time. With nonstop literacy capabilities, it adapts to new fraud ways, perfecting its effectiveness over time. The integration of anomaly discovery further enhances the system's capability to identify suspicious conditioning, while real- time processing ensures quick action to help fiscal losses. also, the system's scalability makes it suitable for high- sale surroundings, icing effective fraud discovery indeed as sale volumes grow. By reducing false cons and perfecting the delicacy of fraud discovery, the system enhances client satisfaction and reduces functional costs. Eventually, this approach provides a robust, automated, and effective result to guard businesses and druggies from fiscal fraud. unborn advancements in model conception and integration with broader fiscal systems will further enhance its performance, icing its continued applicability in combating evolving fraud tactics.

IX REFERENCES

1. Dal Pozzolo, A., Caelen, O., Le Borgne, Y. A., Waterschoot, S., & Bontempi, G. (2014). *Learned lessons in credit card fraud detection from a practitioner perspective*. Expert Systems with Applications, 41(10), 4915–4928.

2. Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., & Caelen, O. (2018). *Sequence classification for credit-card fraud detection*. Expert Systems with Applications, 100, 234–245.

3. Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). *Transaction aggregation as a strategy for credit card fraud detection*. Data Mining and Knowledge Discovery, 18(1), 30–55.

4. Chen, C., & Liu, Z. (2020). *Machine learning for fraud detection: Challenges and opportunities*. Journal of Financial Crime, 27(2), 371–384.

5. Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). *Data mining for credit card fraud: A comparative study*. Decision Support Systems, 50(3), 602–613.

6. Van Vlasselaer, V., Eliassi-Rad, T., Akoglu, L., Snoeck, M., Baesens, B., & Snoeck, M. (2015). *APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions*. Decision Support Systems, 75, 38–48.

7. Carcillo, F., Le Borgne, Y. A., Caelen, O., & Bontempi, G. (2019). *Combining unsupervised and supervised learning in credit card fraud detection*. Information Sciences, 557, 317–331.