

CREDIT CARD TRANSACTION ANALYSIS AND FRAUD DETECTION USING MACHINE LEARNING

¹ P Srinu, ² Suraj Kumar, ³ Mohd saud Shareef, ⁴ M Manish, ⁵ Mesishetty Amarnath

¹AssistantProfessor, ²³⁴⁵Students

Department of CSE(Data Science)

Siddhartha Institute of Technology & Sciences, Narapally

vijaykoraveni.cse@siddhartha.co.in, 24TQ1A6714@siddhartha.co.in, 24TQ1A6730@siddhartha.co.in,
24TQ1A6735@siddhartha.co.in, 24TQ1A6761@siddhartha.co.in

Abstract

The analysis of credit card transactions has become increasingly important in today's digital financial ecosystem, where online payments and cashless transactions are rapidly growing. Financial institutions generate large volumes of transaction data daily, containing valuable insights into customer spending behavior and potential fraudulent activities. Identifying patterns and detecting anomalies in such data is essential for improving financial security, enhancing user awareness, and supporting better decision-making. This project presents a Credit Card Transaction Analysis and Fraud Detection using Machine Learning system developed using Python programming language. The system analyzes transaction data to understand spending patterns, classify expenses, detect fraudulent activities, and segment customers based on their behavior. The dataset includes transaction details such as time, amount, and anonymized features, which are processed and transformed into meaningful information for analysis.

The project follows a complete data analytics pipeline, including data preprocessing, exploratory data analysis, feature engineering, behavioral segmentation, and anomaly detection. Machine learning techniques such as clustering (K-Means) are used to group customers based on spending behavior, while statistical methods like Z-score and anomaly detection algorithms are applied to identify suspicious transactions. Additionally, the system incorporates interactive data visualization using tools like Streamlit, enabling users to explore insights such as spending trends, fraud patterns, and customer segmentation through dashboards and charts. These visualizations make complex financial data easy to understand and interpret.

I. Introduction

The analysis of credit card transactions has become increasingly important in today's digital economy, where cashless payments and online purchases are rapidly growing. Financial institutions and individuals generate large volumes of transaction data through daily activities such as shopping, bill payments, subscriptions, travel, and online services. This transactional data contains valuable insights into spending behavior, financial habits, and lifestyle patterns. However, extracting meaningful information from such large and complex datasets is difficult when handled manually. Understanding customer spending behavior is essential for both financial institutions and individuals. For banks and fintech companies, transaction analysis helps in detecting unusual patterns, identifying potential fraud, and offering personalized financial services. For individuals, it supports better financial planning, budgeting,

and awareness of spending habits. Despite the availability of data, many existing systems fail to categorize and analyze transactions effectively, limiting their usefulness.

With advancements in data analytics and machine learning, it has become possible to process and analyze transaction data more efficiently. Using Python programming language, techniques such as data preprocessing, feature engineering, and classification can be applied to group transactions into meaningful categories like food, travel, entertainment, bills, and shopping. These approaches provide deeper insights into spending trends, frequent transactions, and monthly expenditure patterns. This project focuses on developing a Credit Card Transaction Analysis and Fraud Detection using Machine Learning system. The system analyzes transaction data to understand customer behavior, classify spending patterns, and detect fraudulent activities. It follows a complete data analysis pipeline including data collection, preprocessing, exploratory data analysis, feature engineering, and anomaly detection.

Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn are used for efficient data processing and visualization. The system generates visual reports such as spending distribution and trend analysis, making the results easy to interpret.

II. Literature Survey

Credit card transaction analysis has become a significant research area in data mining and financial analytics due to its importance in understanding user spending behavior and improving financial decision-making. Early studies primarily relied on statistical and rule-based methods to analyze transaction data and identify basic spending patterns. However, with the growth of large-scale financial datasets, advanced data mining techniques have been introduced to extract meaningful insights. These techniques analyze attributes such as transaction amount, time, frequency, and category to understand financial behavior. Feature selection methods further enhance analysis by identifying the most relevant attributes, improving model accuracy and interpretability. Overall, data mining provides a strong foundation for transforming raw transaction data into actionable insights .

In recent years, machine learning techniques have gained significant importance in transaction analysis due to their ability to handle large and complex datasets. Algorithms such as Decision Trees, Random Forest, Support Vector Machines, and clustering methods like K-Means are widely used for spending behavior analysis and categorization. Decision tree-based models are particularly useful for classification tasks, while ensemble methods like Random Forest improve accuracy and reduce overfitting. Clustering techniques help in identifying hidden patterns by grouping similar transactions without predefined labels. Advanced approaches such as Gradient Boosting further enhance predictive performance by capturing complex relationships within data. These models also support adaptability, allowing continuous updates as new transaction data becomes available, making machine learning a powerful tool for analyzing financial behavior.

Furthermore, transaction analysis plays a crucial role in modern financial systems and fintech applications. Financial institutions use analytics to categorize transactions, track spending habits, and detect unusual patterns. Studies highlight that factors such as transaction frequency, category-wise expenditure, and time-based trends are essential for understanding financial behavior. Many applications provide features like expense tracking, budgeting, and personalized financial recommendations by analyzing transaction data. Additionally, organizations use these insights to improve customer experience and design better financial products. Despite these advancements, there remains a need for integrated systems that combine data mining, machine learning, and visualization techniques. This project addresses this gap by developing a comprehensive system that analyzes transaction data, identifies patterns, and supports fraud detection, thereby enhancing financial security and decision-making.

III. System Analysis

Credit card transactions generate massive volumes of data in today's digital financial ecosystem. Managing and analyzing this data is essential for detecting fraud and understanding spending behavior. Traditional systems struggle to process large datasets efficiently. There is a need for intelligent systems that can analyze transaction patterns in real time. Machine learning techniques help in identifying anomalies and suspicious activities. Understanding customer spending behavior improves financial decision-making. Data preprocessing and feature engineering are crucial for accurate analysis. Visualization tools help interpret complex financial data. The system should support real-time monitoring and fraud detection. It must be scalable and secure to handle sensitive data. Integration of analytics and visualization is necessary. This project addresses these needs using modern data science techniques.

Existing System

In the existing system, fraud detection is often based on rule-based methods and manual monitoring. These systems rely on predefined rules such as transaction limits and location checks. Data is stored in separate systems without proper integration. Analysis is limited to basic statistical methods. There is no advanced machine learning implementation. Fraud detection is often reactive rather than proactive. Systems generate static reports with limited insights. Visualization capabilities are minimal. Handling large datasets is difficult. False positives and false negatives are common. Decision-making is slow and inefficient. Overall, the system lacks accuracy and scalability.

Disadvantages of Existing System

- Dependence on rule-based detection
- High false positive and false negative rates
- Lack of real-time fraud detection
- Limited data analysis capabilities
- No advanced machine learning models
- Poor scalability for large datasets
- Fragmented data systems

- Lack of interactive visualization
- Slow response to fraud activities
- Inefficient decision-making
- Manual and time-consuming processes
- Reduced accuracy in fraud detection

Proposed System

The proposed system is a machine learning-based solution for credit card transaction analysis developed using Python programming language. It processes transaction data to analyze spending behavior and detect fraudulent activities. The system uses data preprocessing techniques to clean and structure the dataset. Feature engineering is applied to enhance model performance. Machine learning algorithms such as clustering and classification are used for fraud detection. Statistical methods are also applied to identify anomalies. The system supports real-time or near real-time analysis. Visualization tools are used to present insights clearly. A dashboard is developed for interactive data exploration. The system identifies high-risk transactions effectively. It is scalable and secure. Overall, it improves fraud detection accuracy and financial insights.

Advantages of Proposed System

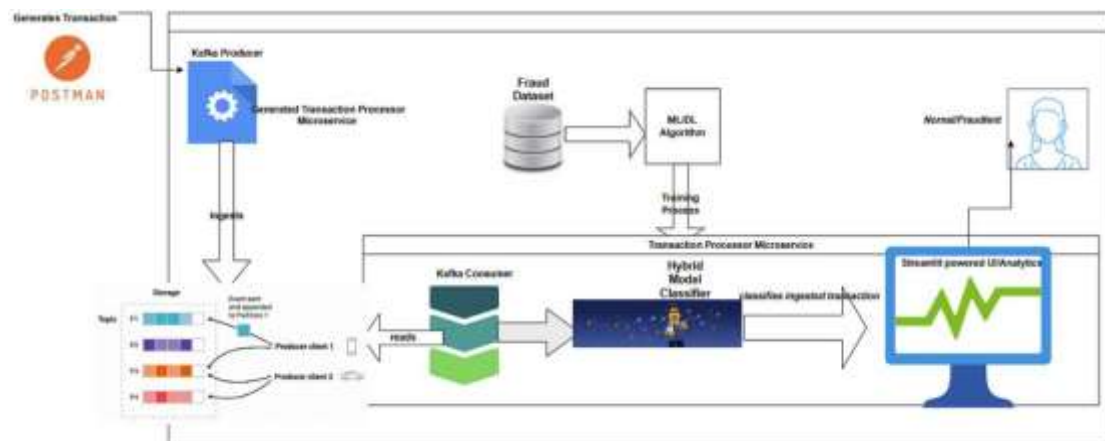
- Accurate fraud detection using machine learning
- Real-time monitoring of transactions
- Reduced false positives and negatives
- Efficient handling of large datasets
- Advanced data analysis capabilities
- Interactive dashboards and visualization
- Improved decision-making support
- Scalable and flexible system
- Automated detection of anomalies
- Enhanced financial security

IV. Methodology

The project begins with collecting credit card transaction data. Data preprocessing is performed to clean and remove inconsistencies. Missing values are handled and data is normalized. Feature engineering is applied to create meaningful attributes. Exploratory Data Analysis (EDA) is conducted to understand patterns. Machine learning models are trained for fraud detection. Clustering techniques are used for behavioral segmentation. Statistical methods such as Z-score are applied for anomaly detection. The model is evaluated using performance metrics. Visualization techniques are used to present results. A dashboard is developed for user interaction. The system is tested for accuracy and performance. Finally, it is deployed for real-world use.

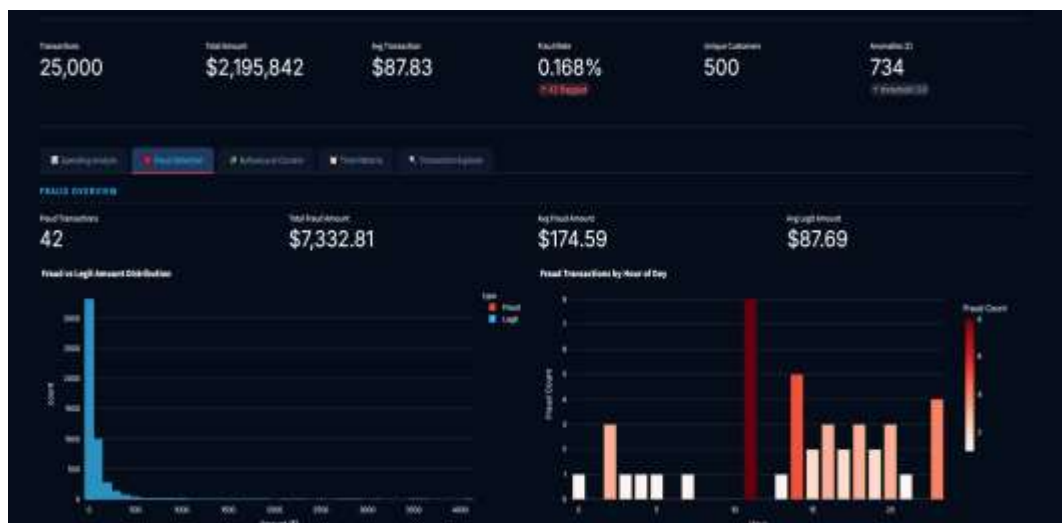
System Architecture

The system architecture consists of multiple layers for efficient data processing. The first layer is the Data Source Layer, which includes transaction datasets. The second layer is the Data Preprocessing Layer, where data cleaning and transformation are performed. The third layer is the Data Storage Layer, where processed data is stored securely. The fourth layer is the Data Analysis Layer, where machine learning models are applied. The fifth layer is the Fraud Detection Layer, where anomalies are identified. The sixth layer is the Visualization Layer, where dashboards and reports are created. The seventh layer is the User Interaction Layer, where users access insights. The system supports real-time updates. It ensures data security and accuracy. The architecture is scalable and flexible. Overall, it provides a complete fraud detection solution.



V. Result and Output





VI. Conclusion

The Credit Card Transaction Analysis and Fraud Detection using Machine Learning system developed in this project demonstrates how data analytics and machine

learning techniques can be effectively applied to analyze financial transactions and improve security. In today's digital economy, raw transaction data alone does not provide meaningful insights. By processing and analyzing this data, users and financial institutions can gain a clear understanding of spending behavior and detect potential fraud.

In this project, credit card transaction data was analyzed to identify patterns in user behavior and detect anomalies. Various features such as transaction amount, time, and transaction characteristics were utilized to perform the analysis. The dataset was preprocessed using Python programming language to handle missing values, remove inconsistencies, and structure the data for effective analysis.

The system focuses on both behavioral profiling and fraud detection. Transactions are analyzed to identify spending patterns and categorized into different groups such as food, travel, shopping, and bills. At the same time, anomaly detection techniques and machine learning models are applied to identify suspicious or fraudulent transactions. This dual approach enhances both financial understanding and security.

Exploratory data analysis and visualization techniques are used to present insights clearly. Graphs such as bar charts, pie charts, and line charts help in understanding spending distribution and trends over time. These visualizations simplify complex data and make it easier to interpret key findings.

References

1. Kumar, R. D., Prudhviraaj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In *Handbook of Artificial Intelligence and Wearables* (pp. 145-158). CRC Press.
2. Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In *The International Conference on Artificial Intelligence and Smart Environment* (pp. 557-564). Cham: Springer Nature Switzerland.
3. Sv satyakraishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
4. Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
5. V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in *Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT)*, Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
6. B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in *Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart*

-
- Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.
7. R. D. Kumar, V. N. S. Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in *Blockchain for Smart Cities*, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
 8. Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
 9. Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," *Journal of Machine and Computing*, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
 10. Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in *Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE)*, Apr. 2025.
 11. Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," *SN Computer Science*, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.
 12. Gaddam, S. Integrating Analytics into the Development Process: Bridging the Gap between Data Insights and Design Execution.
 13. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
 14. Poojari, R. Frameworks for Data Management and Lineage in Large-Scale Healthcare Data Systems.
 15. Purmani, S. S. R. (2025). Streamlining IT operations and service management with agile frameworks. *European Journal of Advances in Engineering and Technology*, 12(4), 76–81.
 16. Viswanathan, V. (2024). Embedding Ethical Principles into Generative AI Workflows for Project Teams.
 17. Mudusu, S. K. (2026, April 15). The secure intelligence framework: Architecting AI systems for a data-driven world. CIO (Foundry Expert Contributor Network).
 18. Viswanathan, V. (2024). Pioneering Ethical AI Integration in Enterprise Workflows: A Framework for Scalable Team Governance. Available at SSRN 5375619.
 19. Mudusu, S. K. (2025, June 3). Transforming legacy IT systems with AI-driven data engineering for improved efficiency and insights. *Hampton Global Business Review (HGBR)*.
 20. Gajula, S. (2026, March). Two Pillars of Banking Intelligence: A Comparative Analysis of AI Techniques for Fraud Prevention and Churn Mitigation. In *2026 14th International Symposium on Digital Forensics and Security (ISDFS)* (pp. 1-6). IEEE.
 21. Maturi, S. Y. (2023). Crowdsourced frontier: Unveiling autonomous adversarial cybercapabilities via open AI competition. *International Journal of Intelligent Systems and Applications in Engineering*, 11(1s), 275–284.
-

22. Chowdhury, A. K., Muhit, M. M. I., & Islam, M. M. (2023). A practical review to the marine maintenance practice in Bangladesh and a proposed way forward to an efficient, long-term and cost-effective solution. In Proceedings of the 13th International Conference on Marine Technology (MARTEC 2022). <https://doi.org/10.2139/ssrn.4445071>
23. Manoharan, D. (2025). Healthcare EDI Transaction Lifecycles Embedded with a Multi-Layer Verification Framework to Ensure Referential Integrity.
24. Ravishankara, M. (2026, February). CircuChain: Disentangling Competence and Compliance in LLM Circuit Analysis. In SoutheastCon 2026 (pp. 1-7). IEEE.
25. Doragacharla, V. R. (2023). Comprehensive Benchmarking Analysis of Auto Scaling Approaches in Cloud Native Streaming Pipelines During Flash Sales and Holiday Traffic Peaks. Available at SSRN 6566479.
26. P. Venkata Ramana. (2024). AI-driven predictive analytics in ERP systems for proactive supply chain optimization. International Journal of Innovative Engineering and Management Research (IJIEMR).
27. Kumar Adabala, P. (2021). Optimizing ERP Modernization: A Smart Data Migration Framework Approach. International Journal of Enhanced Research in Science, Technology & Engineering, 10(07), 61–72. <https://doi.org/10.55948/ijerste.2021.0708>
28. Kavuri, S. (2025). Critical Review of Software Testing Problems in the Current Decade. International Journal on Science and Technology, 16(2). <https://doi.org/10.71097/ijstat.v16.i2.9469>
29. Srikanth Kavuri. (2024). Probabilistic Generative Modeling for Synthesizing High-Coverage Test Data in Safety-Critical Software Applications. Computer Fraud and Security, 633–642. <https://doi.org/10.52710/cfs.838>
30. Venkata Pavan Kumar Gummadi. (2024). API Design and Implementation: RAML and OpenAPI Specification. Journal of Electrical Systems, 16(4), 76–85. <https://doi.org/10.52783/jes.9329>
31. Venkata Pavan Kumar Gummadi. (2025). MuleSoft's Role in Advancing Sustainable Digital Infrastructure: An Enterprise Integration Perspective. Journal of Information Systems Engineering and Management, 10(62s), 1313–1321. <https://doi.org/10.52783/jisem.v10i62s.13783>