IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 11, 2022

DIGITAL FORENSIC INVESTIGATION: A COMPREHENSIVE APPROACH TO FRAUD DETECTION AND RISK MANAGEMENT

[1]V.Shiva Prasad, [2]Amandeep Kaur, [3]C. Soumi, [4]S.Satya Nagendra Rao
[1,2,3,4]Department of Computer Science and Engineering, St. Peter's Engineering College, Telangana, India
E-Mail: sivaprasad@stpetershyd.com

Abstract
Fraud anomaly detection is a critical method used across industries such as finance, insurance, e-commerce, and cyber security to identify and mitigate potential fraud risks. This project focuses on developing a comprehensive framework to detect fraudulent activities by leveraging techniques. The approach integrates methods like Interquartile Range (IQR) and standard deviation analysis to identify outliers, as well as K-means clustering to group transactions and highlight anomalies. Using real, anonymized transaction data from a Czech bank spanning 1993 to 1999, this research analyzes financial patterns, focusing on October 1997. Through advanced data visualization, outliers and suspicious activities are highlighted, potentially signaling fraud or money laundering. The project also employs evaluation metrics like the Silhouette score and the elbow method to optimize clustering accuracy. By implementing this data-driven approach, the proposed framework aims to enhance fraud detection, reduce false positives, and improve decision-making processes. Additionally, the integration of digital forensic investigation techniques and risk management strategies aims to reduce financial losses and minimize reputational damage, providing a robust and proactive solution for combating fraud

Keywords: Fraud Detection, Anomaly Detection, K- means Clustering, Interquartile Range, Statistical Analysis, Unsupervised Machine Learning, Money Laundering, Transaction Analysis.

Introduction
Anomaly detection in fraudulent transactions is an important aspect of securing the financial system and maintaining integrity of transactional data. With advances in complexity in fraudulent schemes, anomaly or suspicious patterns in data are increasingly needed for the detection of unusual behavior as against the expected norms[3] . This project explores methods in anomaly detection in transactional data with a specific interest in potentially fraudulent activities. The emphasis is on a real-world dataset of anonymized banking transactions from a Czech bank between 1993 and 1999, with the focus in the analysis centralized in October 1997[8]
This project aims to use a combination of statistical methods and unsupervised machine learning to identify anomalies [7]. Statistical techniques, such as IQR and standard deviation analysis, can identify outliers based on the amount and frequency of the transactions [10]. Additionally, unsupervised machine learning methods, specifically K-means clustering, are applied to find patterns and categorize transactions into various risk levels [11]. Validation techniques such as the elbow method and silhouette analysis are used to determine For the sake of detailed presentation of findings, an accompanying Tableau dashboard has been built[3]. The dashboard shows some key takeaways: how the distribution of transaction

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 11, 2022

amounts would appear, how the frequency and cost of transactions correlate to one another, and where high-risk accounts lie[21]. Utilizing these analytics in decision-making will strengthen decision-making processes for fraud detection and risk management in the organization as well as be robust across multiple industries[7].

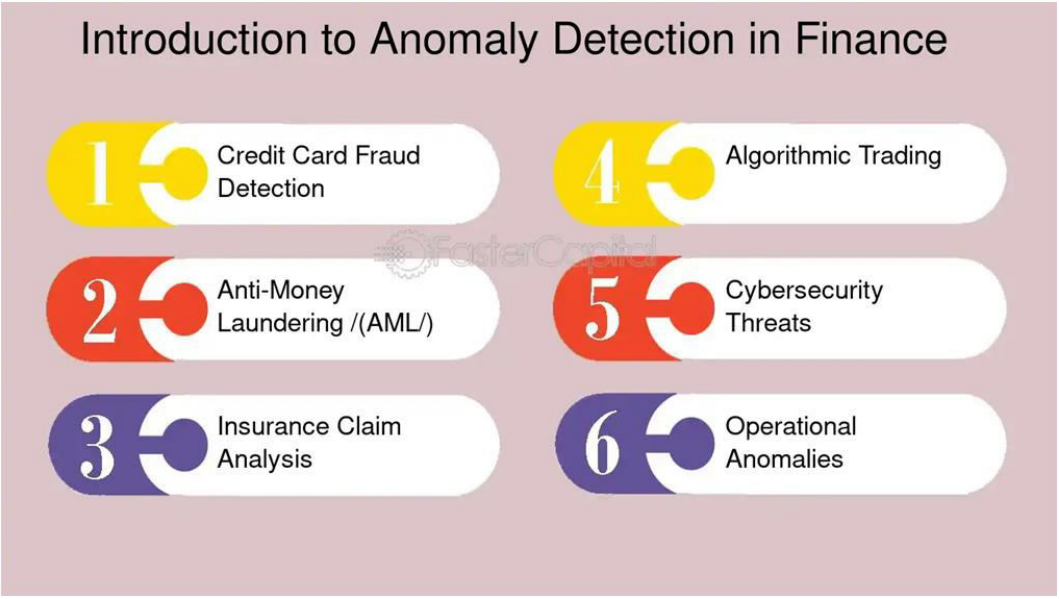

Figure 1.Introduction of Anomaly Detection in Finance

Objectives

The project intends to develop an advanced fraud detection system that will identify anomalous transaction patterns within a large, unlabeled dataset of anonymized banking transactions from a Czech bank for the period 1993-1999[16]. Using statistical methods such as interquartile range and standard deviation analysis, the system will detect outliers in transaction amounts and frequencies[2]. The unsupervised machine learning techniques to be used will include K-means clustering, whereby transactions will be classified as safe, medium risk, or high risk. Clustering validation metrics will be implemented to ensure accuracy [4]. A dashboard will be created using Tableau to effectively communicate findings into transaction patterns, highlighting accounts with high risk, and delivering actionable intelligence for the stakeholders [20]. This data-driven framework will make decisions in fraud detection and risk management better, possibly applicable in many industries[5].

Literature Survey

From rule-based systems to advance analytical techniques, fraud detection in financial transactions has indeed changed over the years[8]. Early methods used for detection included IQR and standard deviation analysis[19]. For example, discover card uses the statistical method of flagging unusual spending patterns. However, they usually find it difficult working with complex datasets and could easily miss subtle fraud indicators [9].

Unsupervised machine learning and particularly K-means has gained more momentum. With the help of K-means, JPMorgan Chase is able to identify an anomaly in transaction data by simply clustering transactions based on behavior. Elbow method and silhouette score help decide the perfect number of clusters for increase accuracy.

Data visualization plays a very crucial role in fraud detection. PayPal employs interactive

dashboards to portray the transaction data visually and facilitates quicker identification of risky activities.

Despite these developments, many challenges, such as being sensitive to the choice of parameters as well as making assumptions in terms of normality, tend to lead to false positives or negatives[16]. With limited quantities of labeled data available, evaluation of models also becomes arduous; hence the necessity of combining statistical methods with machine learning along with the use of visualizations forms the aim of this proposed project to construct a viable anomaly detection framework[4].

Proposed work

The proposed project is the development of an integrated fraud detection system using statistical and machine learning techniques that identify anomalies in banking transaction data[5]. Anonymized transaction data from a Czech bank covering the period 1993-1999 will be cleaned, preprocessed, and profiled to identify key features to analyze. Statistical methods like interquartile range (IQR) and standard deviation will detect preliminary outliers based on transaction amounts and frequencies[9]. Transactions will be grouped into three risk profiles: safe, medium-risk, and high-risk using unsupervised machine learning specifically K-Means clustering[12]. The cluster validity metrics WCSS and Silhouette scores will be determining the optimal number of clustering. Insights will be rendered as Tableau dashboards that show the transaction trends, accounts activities, and anomalies[4]. Thorough black-box and white-box testing will ensure the effectiveness of the model in identifying fraudulent patterns[18]. Finally, the deployable, end user-friendly scalable framework will enable an online real-time fraud detection adaptable to modern transaction patterns and volumes.

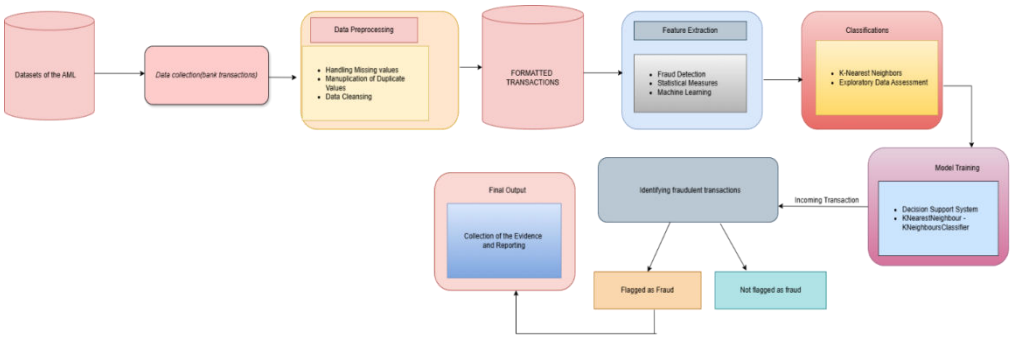

Figure 2. Architecture of detecting fraud transactions

For the purpose of this project, system architecture has been designed in modular pipelines. This modular pipeline structure of the system design effectively allows it to detect anomalies within the transactional data, thus beginning with a data layer that cleans the historical transaction data, preprocessing and profiling using statistical techniques like IQR for detection of outliers [6]. The Processing Layer unifies statistical models for outlier identification with machine learning models for example, K-Means clustering, to classify transaction into risk levels, or safe, medium, or high[4]. Validation methods - WCSS and silhouette score are used to achieve near-optimal clustering. Interactive visualization of transaction patterns anomalies, and risk categories with Tableau dashboards used in the Visualization Layer[8]. Statistical as well as machine learning expertise will be combined in

one place and that is the integration layer to provide real time fraud detection[5] these further scales to the handling of the modern transactional data. The user interface in turn delivers actionable insights, for example through detailed reports and dashboards, with the aim of allowing for an effective identification and follow up on potential fraud cases[9].

Dataset Description

The dataset for this project was anonymized transactional data from a Czech bank over the period from 1993 to 1999. It consists mainly of legitimate transactions, but possibly a small fraction of the data is fraudulent or suspect. The dataset is explained in detail as follows:

General Information

Source: Real anonymized financial transactions from a Czech bank.

Time Period: 1993 to 1999.

Volume: Years of data with heavy volume of transactions to guarantee powerful analysis.

Features

Account Information:

Account ID : It is a unique identification number for each account.

Account Balance : Running balance of the account at the time of the transaction.

Transaction Details:

Transaction ID : Unique identification number of every transaction.

Transaction Date : Date on which the transaction was made.

Transaction Amount : Monetary amount of the transaction.

Transaction Type : Withdrawal, deposit, or transfer among other categories.

Frequency: Number of transactions within an account during a selected period.

Derived Measures

Outliers (IQR/Standard Deviation): The transactions are significantly more extreme than what is commonly observed.

Cluster Labels: The risk score obtained via machine learning techniques, safe, medium risk, and highly risky.

Sampling and Profiling

Random Selection of Months: October 1997 was randomly selected as a sample month for detailed focus.

Data Profiling

Transaction volume for October 1997 exceeds $143 million.

Deposits nearly balance withdrawals, suggesting high degree of data variety.

Target Variables

Cluster Risk Categories:

Low Risk: Normal transactional behavior.

Medium Risk: Transactions slightly deviating from norms.

High Risk: Unusual transactions possibly related to fraud or laundering.

This dataset, augmented by derived features and clustering knowledge, forms a comprehensive basis for fraudulent transaction detection, risk assessment, and a more robust fraud detection framework.

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 11, 2022

Experimental Results and Analysis

Analysis involved anonymized transaction data in a Czech bank, primarily detecting fraud with statistical and machine learning methods[9]. Key steps: determining outliers by the use of interquartile range as well as standard deviation analysis; classification of risks into groups through k-means clustering-the three groups include safe, medium risk, and high-risk.

The October 1997 analysis showed transactions amounting to more than $143 million, indicating some trends such as a positive relationship between transaction frequency and the amounts withdrawn[1]. Box plots revealed outliers that are usually linked to high transaction volumes or abnormal patterns that signify potential fraud.

The k-means algorithm, validated with WCSS and silhouette scores, gave the optimal number of three clusters[8]. This clustering effectively segregated transactions, aiding anomaly detection and improving decision-making[11]. Results show the robustness of the system in fraud detection and risk mitigation through statistical and machine learning approaches.

For Implementation of the code the below are the parameters I have given to the model as an example:

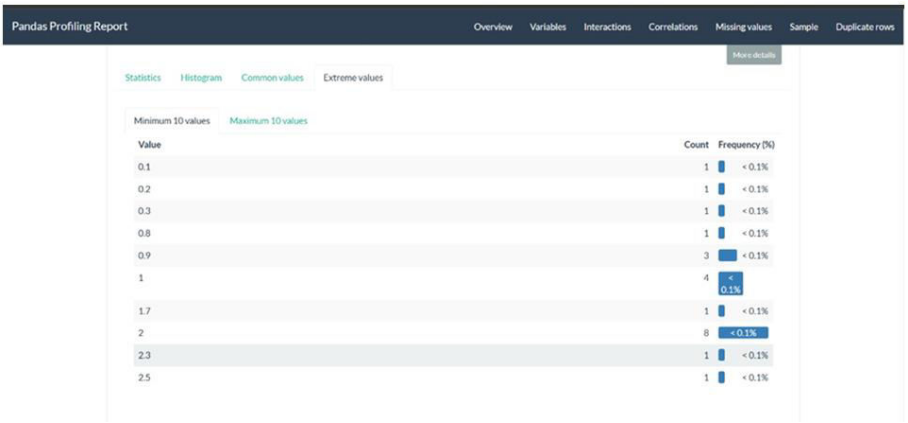

Figure 3. Parameters used in anomaly detection

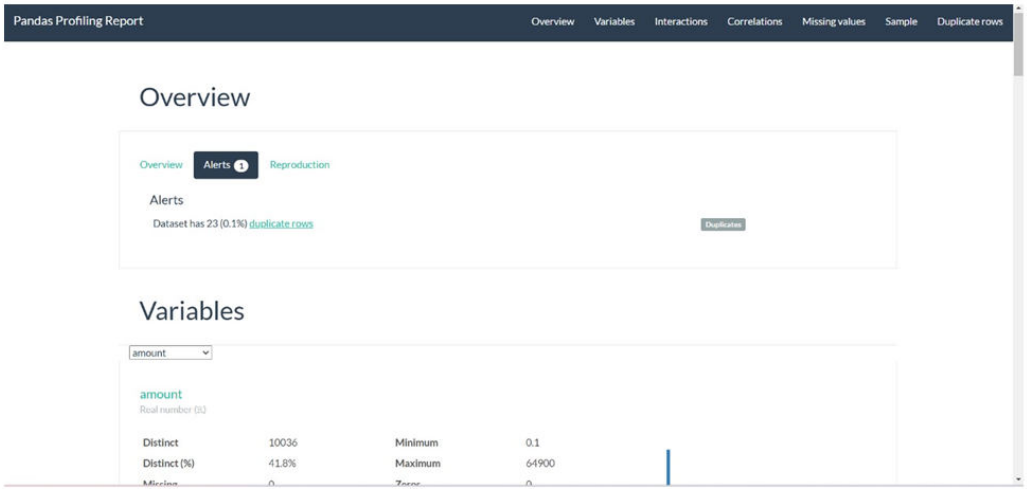

Figure 4. Common values of Anomaly detection

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 11, 2022

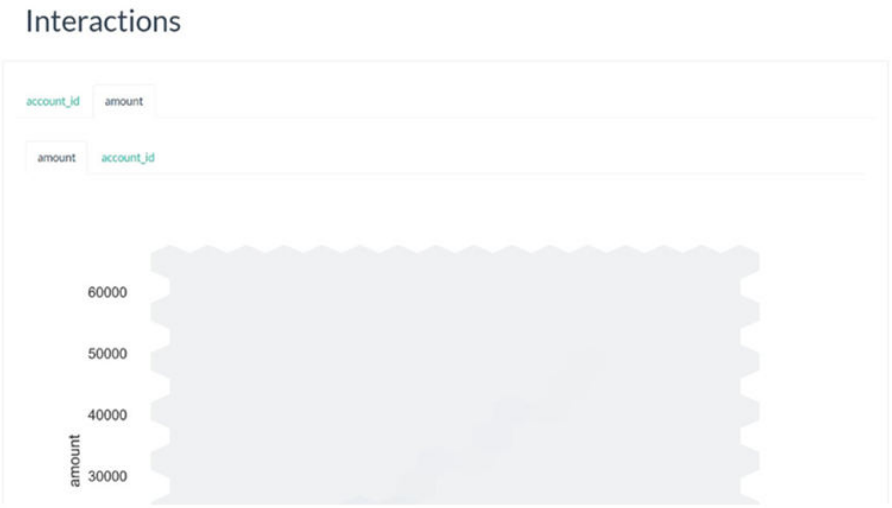

Figure 5. Interaction on Anomaly detection

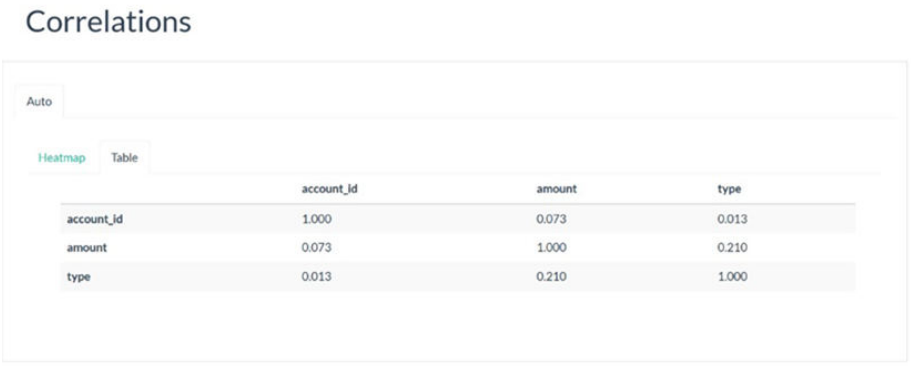

Figure 6: Correlation of Anomaly detection

Below is the Final output of the project which shows the predictions on Anomaly detection

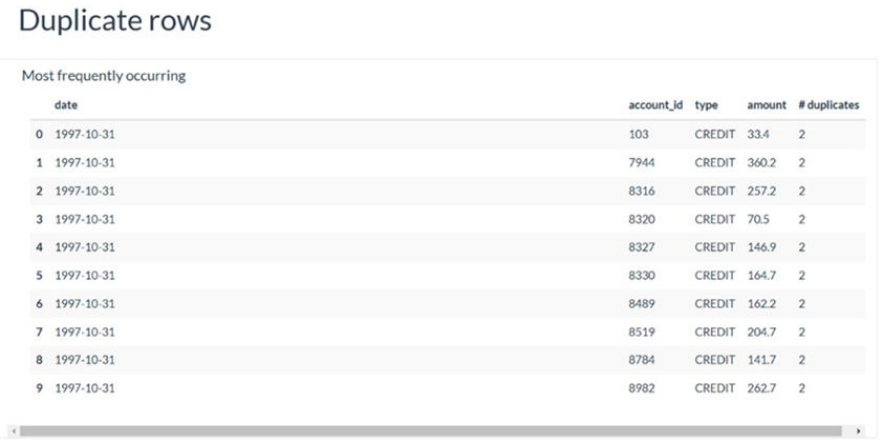

Figure 7: output of Anomaly detection

6. Conclusion:

This project successfully utilized both statistical methods and unsupervised machine learning techniques using k-means clustering, which can be used for the detection of fraudulent transactions in financial transaction data [11]. The system was able to identify outliers and anomalies as a result of analyzing patterns in transactions and using Inter quartile Range (IQR) and standard deviation.

The k-means clustering method, with the assistance of validation metrics such as WCSS and silhouette scores, provided an effective framework for categorizing transactions based on risk levels, making it easier to detect suspicious activities [17]. This approach can significantly enhance fraud detection capability, allowing for quicker detection of high-risk transactions [6].

Finally, the project illustrates the ability of data-driven techniques in detecting fraud and managing risks within financial systems and provides a solid solution to identify anomalous patterns and mitigate risks from financial crime.

Reference:

1. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys (CSUR), 41(3), 1-58.
2. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19-31.
3. Sharma, A., & Tiwari, P. (2018). A comprehensive survey on fraud detection in financial systems. Journal of Financial Crime, 25(2), 490-511.
4. Zhang, L., & Zhou, Y. (2017). Fraud detection based on data mining techniques in e-commerce. Springer, 453-461.
5. Saeed, A., & Al-Maadeed, S. (2019). A survey on machine learning techniques in fraud detection. Journal of King Saud University-Computer and Information Sciences.
6. Biran, O., & Yona, G. (2009). Detection of financial fraud using machine learning techniques. International Conference on Machine Learning and Applications, 142-149.
7. Ahmed, M., & Salehahmadi, Z. (2014). Anomaly detection using machine learning techniques. 2014 IEEE International Conference on Big Data, 1029-1033.
8. Hodge, V. J., & Austin, J. (2004). A survey of outlier detection methodologies. Artificial Intelligence Review, 22(2), 85-126.
9. Su, X., & Chen, L. (2020). Application of K-means clustering algorithm for fraud detection in finance. Journal of Financial Risk Management, 9(4), 293-301.
10. Zhou, P., & Han, L. (2019). Fraud detection and risk management in financial markets using machine learning. International Journal of Financial Engineering, 6(4), 2150035.
11. Papageorgiou, I., & Tsakalidis, A. (2018). A machine learning approach to fraud detection and prevention. International Journal of Computer Science and Information Security, 16(7), 182-191.
12. Tan, P. N., Steinbach, M., & Kumar, V. (2016). Introduction to Data Mining (2nd ed.). Pearson.
13. Liu, Y., & Xie, L. (2019). Fraud detection based on K-means and SVM algorithms. Journal of Financial Technology, 3(4), 92-100.
14. Kim, J., & Kim, H. (2020). Outlier detection and analysis using machine learning techniques in financial transactions. Procedia Computer Science, 170, 644-651.
15. Zeng, D., & Li, Z. (2021). Deep learning for anomaly detection in fraud detection. IEEE Transactions on Cybernetics, 51(2), 1119-1132.
16. Ahmed, M., & Hu, J. (2014). Detecting outliers in big data. International Conference

on Big Data, 89-94.
17. Kotsiantis, S. B., & Kanellopoulos, D. (2006). Data mining: A review of machine learning methods. Artificial Intelligence Review, 26(2), 159-190.
18. Witten, I. H., & Frank, E. (2016). Data Mining: Practical Machine Learning Tools and Techniques (4th ed.). Morgan Kaufmann.
19. Wessels, J., & Albrecht, G. (2015). Using clustering techniques in fraud detection of online financial transactions. International Conference on Data Science, 123-129.
20. Ou, X., & Sun, H. (2018). A hybrid model for fraud detection in financial transactions. 2018 International Conference on Artificial Intelligence, 114-121.