COMPARATIVE EXAMINATION OF DEBIT CARD FRAUD DETECTION METHODS EMPLOYING MACHINE LEARNING INTELLIGENCE APPROACHES

R. Abdulyekeen [1]
F. O. Echobu [2]
A. Zakariyya [3]

Received: 21/01/2025
Revised: 26/02/2025
Accepted: 27/02/2025

[1] Department of Computer Science, Federal University, Dutsin-Ma, Nigeria
[2] Department of Information Technology, Federal University Dutsin-ma, Nigeria
[3] Department of Software Engineering, Federal University Dutsin-Ma, Nigeria
*Corresponding Author's Email: rabdulyekeen@fudutsinma.edu.ng

https://doi.org/10.20428/jst.v30i4.2756

Comparative Examination of Debit Card Fraud Detection Methods Employing Machine Learning Intelligence Approaches

R. Abdulyekeen
Department of Computer Science, Federal University, Dutsin-Ma, Nigeria,

F. O. Echobu
Department of Information Technology, Federal University, Dutsin-ma, Nigeria

A. Zakariyya
Department of Software Engineering, Federal University Dutsin-Ma, Nigeria

Abstract— For this study, we look at how to spot fake debit card transactions using supervised intelligent expert systems techniques like Extreme Gradient Boosting, Random Forest, KNN, Logistic Regression, SVM, and Decision Trees to solve classification problems. A debit card dataset with two parts, 30 variable fields, and a total of 248,807 records was used to do a thorough evaluation of data mining and machine learning algorithms. This dataset helped with the investigation of fraud detection. After finishing the basic structure, I can say with certainty that the combined technique model performs much better than a separate system that uses k-means, XGBoost, SVM, and logistic regression to find outliers. Given that the main objective is to detect fraudulent activities in a debit card dataset, its effectiveness is measured by how frequently it identifies outliers or atypical user behaviour patterns.

Keywords— Artificial Intelligence, Machine Learning, Dataset, SVM, And Logistic Regression.

I. INTRODUCTION

Debit card fraud involves the unauthorised use of a bank account by an individual apart from the rightful account holder. To mitigate this kind of abuse, appropriate strategies can be implemented, and understanding the patterns associated with fraudulent behavior can aid in decreasing the chances of future incidents and protecting against repeat occurrences [1]. In other terms, debit card fraud occurs when someone uses another person's debit card without the knowledge or consent of the owner or the card-issuing institution. Fraud discovery consists of observing user behaviour to identify, predict, or prevent negative activities, such as fraud, unauthorised access, and defaulting on payments. Domains such as machine learning and data science, which can create automated approaches, must focus on this vital matter. The intricacy of fraud detection arises from issues like class imbalance, where valid transactions significantly exceed fraudulent ones, and the shifting transaction trends that evolve over time.

Data mining refers to extract and identify useful information from large sets of data [2]. Fundamentally, data mining is the process of scrutinising large collections of facts to gain vital insights that guide future choices. Once we establish an effective model, we can use it for forecasting by assigning categories to incoming data. In the realm of fraud detection, the process entails categorising transaction information to pinpoint possible fraudulent actions. We generate a classification model by identifying trends from earlier fraudulent incidents, which we then apply to fresh data sets to highlight questionable behaviours. Even so, data mining and model creation can take a long time, which makes it hard to find fraud right away. This is especially true for situations like online debit card transactions, where fraud needs to be found right away, often while the fraudster is still in the bank.

Additionally, the detection of fraud in the real world encounters further hurdles. Automated systems rapidly handle payment requests, determining which transactions to validate. Algorithms powered by machine learning scrutinise approved transactions to pinpoint and highlight those that seem suspicious. These identified transactions undergo a detailed review by experts who contact cardholders for confirmation. Insights gained from these evaluations are integrated back into the system, allowing for the ongoing improvement and advancement of the fraud detection algorithms as time progresses.

Fraud detection techniques are constantly evolving to counteract criminals as they adapt their fraudulent tactics. We categorise these fraud types as follows:

- Debit Card Frauds: Online and Offline
- Card Theft
- •Account Takeover
- Device Breach
- Application Fraud
- Counterfeit Card Fraud
- Communication Network Fraud

Debit card fraud encompasses a broad spectrum of crimes related to the unlawful use or manipulation of this card during transactions. The intent may be to acquire items without payment or to illicitly move money from an individual's account. Additionally, debit card fraud is often tied to identity theft.

Throughout the years, advancements in technology have significantly impacted the evolution of Nigeria's financial sector. This includes the ability to transfer money from the convenience of our homes and pay utility bills effortlessly without needing to attend to the service providers, as well as the innovative use of biometrics for uniquely identifying bank customers. Numerous state-of-the-art services and products have emerged, transforming how I engage and conduct transactions. The era of lengthy waits at banks is now behind us. The advancement, clarity, and speed that technology has

introduced to Nigeria's financial landscape are remarkable. The criminals are perpetually devising new ways to fulfil their malicious goals and exploit the system. Nonetheless, the industry consistently engages in discussions and adopts measures and regulations to counter the activities of these fraudsters. Although it has been a challenging fight, we are undoubtedly making progress. An effort by the Central Bank of Nigeria (CBN) to set up fraud desks in the industry, the routing of all electronic interbank transactions through the Central Anti-Fraud Solution (HEIMDALL), the use of biometric systems, and, most importantly, our ability to work together have all helped reduce fraud in Nigeria's financial sector. Figure 2 illustrates that in 2016, there were 19,531 reported fraud cases in deposit money banks, a significant increase from 10,743 cases in 2015. Despite the 82% rise in reported cases, there was a marginal reduction in both the attempted fraud value and the actual loss, amounting to ₦4,368,437,371.64 and ₦2,196,509,038.78, respectively. Additionally, actual losses due to fraud decreased by 2.65% in 2016 compared to 2015. Table 1 provides a summary of the fraud reports, highlighting that the fraud volume in 2016 was notably higher than in 2015.

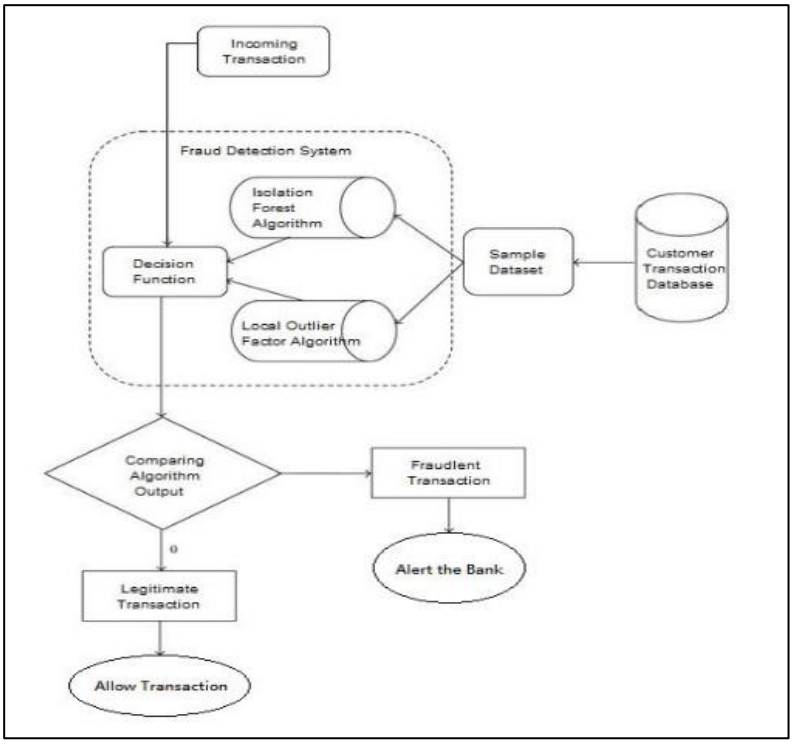

Fig 1. Detection and Response Algorithm

Table 1: Summary Fraud Report

YEAR	Fraud Volume	Attempted Fraud Volume	Actual Loss Value
2015	10,743	4,374,512,776.64	2,256,312,660.00
2016	19,531	4,368,437,371.64	2,196,509,038.78

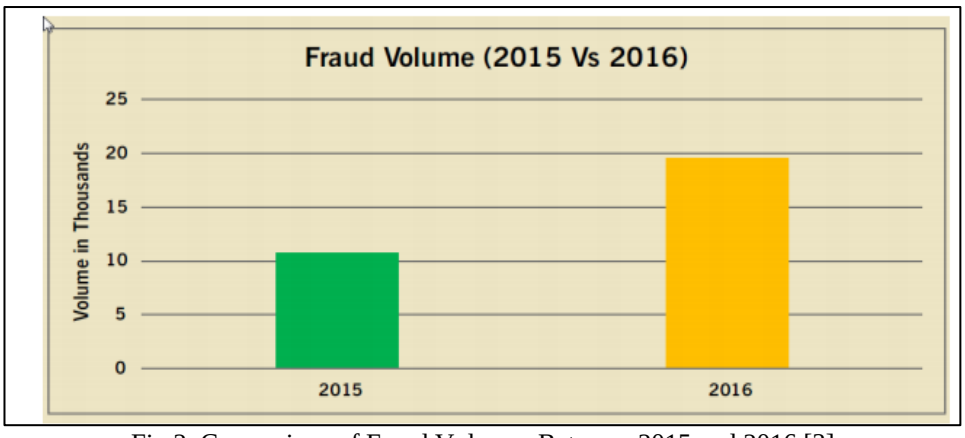

Fig 2. Comparison of Fraud Volumes Between 2015 and 2016 [3]

In the rapidly evolving milieu of the financial sector, Deposit Money Banks (DMBs) in Nigeria are faced with the considerable challenge of addressing fraud, an omnipresent concern that significantly undermines the integrity of financial systems. As the nation endeavors to cultivate a robust and secure banking framework, the function of technological innovation emerges as a crucial driver in strengthening measures aimed at fraud prevention [4].

A. What are the major techniques?

1. Cybercriminals utilise numerous methods to execute debit card fraud, including:
2. Cross-channel fraud: acquiring customer details on one platform, like a call centre, and utilising that information for fraudulent activities on another platform, such as an ATM.
3. Data theft: Cyber attackers breach both secure and unsecured websites, obtain personal information, and market it.
4. Spoofing of Email: altering the header details in an electronic mail to disguise the sender's identity and create the illusion that the email is from a legitimate source.
5. Phishing: the act of obtaining important data such as debit card details, user credentials, personal account numbers, and passwords through spoofing of email tactics.
6. Text Phishing: Perpetrators employ short message service (SMS) to trick individuals, usually including a contact number to call.
7. Social Engineering Call: Scammers make calls to extract sensitive private data from victims.
8. Shoulder Surfing: employing direct observation methods, like peering over someone's shoulder, to gather personal information such as PINs and passwords.
9. Underground Websites: Criminals acquire personal data, including PINs and PANs, from illicit websites.
10. Keylogger Software: the deployment of harmful software designed to capture sensitive details like passwords and card information.
11. Web Application Vulnerability: Attackers exploit weaknesses in web applications to gain unauthorised access to vital systems.
12. Eavesdropping: intercepting and viewing sensitive data as it travels across a network.
13. Google Hacking: leveraging Google search techniques to find confidential information about a target, intending to use it to commit fraud.
14. Session Hijacking: unlawfully taking charge of a communication session to extract information or compromise a system in some way.
15. Man-in-The-Middle Attack: a fundamental approach to capturing information and enabling more elaborate attacks.

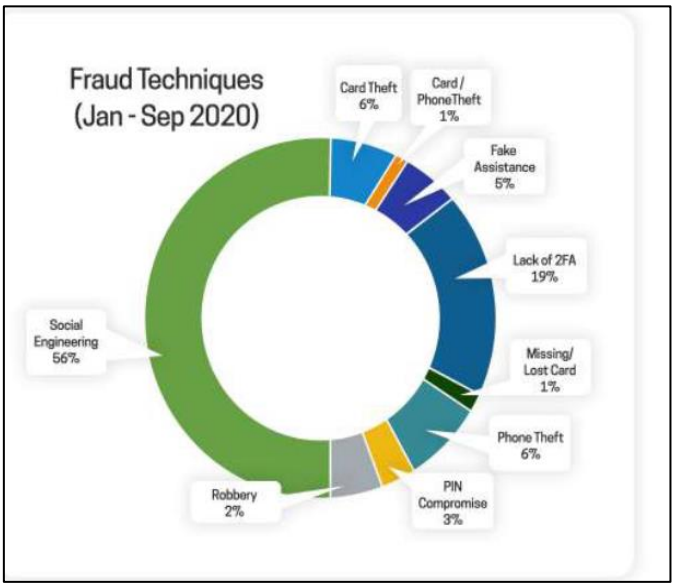

Fig. 3. Fraud Techniques [5]

B. Some of the currently used approaches to detection of such fraud are

- Artificial Neural Network
- Fuzzy Logic
- Genetic Algorithm
- Logistic Regression
- Decision tree
- Support Vector Machines
- Bayesian Networks
- Hidden Markov Model
- K-Nearest Neighbour

II. METHODOLOGY

This study employs supervised learning models (e.g., Logistic Regression, KNN, Random Forest) and unsupervised techniques (e.g., Peer Group Analysis, Break Point Analysis) to develop a comprehensive fraud detection framework. The methodology includes preprocessing the dataset, Exploratory data analysis (EDA) and the process of selecting relevant feature, model training and evaluation using Python libraries.

Table 2: Algorithms steps followed for this experiment

Algorithms Steps	Description
1. Import Packages	Importing the required packages into our python environment.
2. Import Data	Import the Data (Debit card.csv)
3. Data-preprocessing	Processing the data to our needs and Exploratory Data Analysis
4. Data Split	Feature Selection and Data Split
5. Modeling	Building six types of classification models
6. Evaluation	Evaluating the created classification models using the evaluation metrics

A. *Data Preprocessing & Handling Challenges*

1. *Data Cleaning:* We resolved missing values by using median imputation and outlier removal techniques.
2. *Feature Engineering:* We derived new features such as transaction frequency, user behaviour patterns, and deviations from typical spending behaviours.
3. *Feature Scaling:* We used StandardScaler to normalise transaction amounts.
4. *Handling Class Imbalance:* As fraudulent transactions constitute a small minority, the Synthetic Minority Over-sampling Technique (SMOTE) was utilised along with undersampling of the majority class to enhance model reliability. Additionally, cost-sensitive learning approaches were implemented to minimise bias towards nonfraudulent transactions.
5. *Exploratory Data Analysis (EDA):* We applied Principal Component Analysis (PCA) to diminish dimensionality while maintaining key features. Correlation heatmaps and variance analyses were used to eliminate redundant variables.

B. *Packages Used*

1. *Python:* A programming language ranked 1 in use for implementing machine learning and data science.
2. *XGBoost:* XGBoost, short for Extreme Gradient Boosting, is a highly regarded Python library designed for implementing gradient boosting algorithms, which enhance the performance and accuracy of machine learning models. It is widely recognised as one of the most effective tools for booster-based machine learning tasks. Similarly, libraries like LightGBM and CatBoost are also well-equipped with advanced functionalities for gradient boosting, though their methods may sometimes lack precise definition. These libraries are primarily focused on optimising model performance through gradient boosting techniques.
3. *Numerical Python:* Numerical NumPy is a foundational Python library extensively used for numerical computations. Its standout feature is its ability to handle multi-dimensional arrays, enabling efficient mathematical and logical operations. NumPy provides a wide range of functions for tasks such as indexing, sorting, reshaping, and even representing complex data like images and sound waves as multidimensional arrays of real numbers.
4. *Pandas*: Pandas is a significant statistical library that is utilised across various fields, including statistics, finance, economics, and data analysis. The NumPy array serves as the foundation for this library, which handles Pandas data objects. NumPy, Pandas, and SciPy are closely interrelated and rely on each other for scientific computing, data manipulation, and similar tasks.
5. *Modelling Tools:*
   - Logistic regressions
   - Decision trees
   - KNN
   - Random Forest
   - Support Vector Machine

III. RESULTS

The research examined six machine learning algorithms using a publicly accessible dataset for debit card fraud. Models such as KNN and logistic regression were evaluated for their accuracy and F1 scores. The findings revealed that KNN attained the highest accuracy at 99%, whereas logistic regression demonstrated the lowest accuracy. While unsupervised anomaly detection techniques were more resource-demanding, they successfully pinpointed fraudulent transactions.

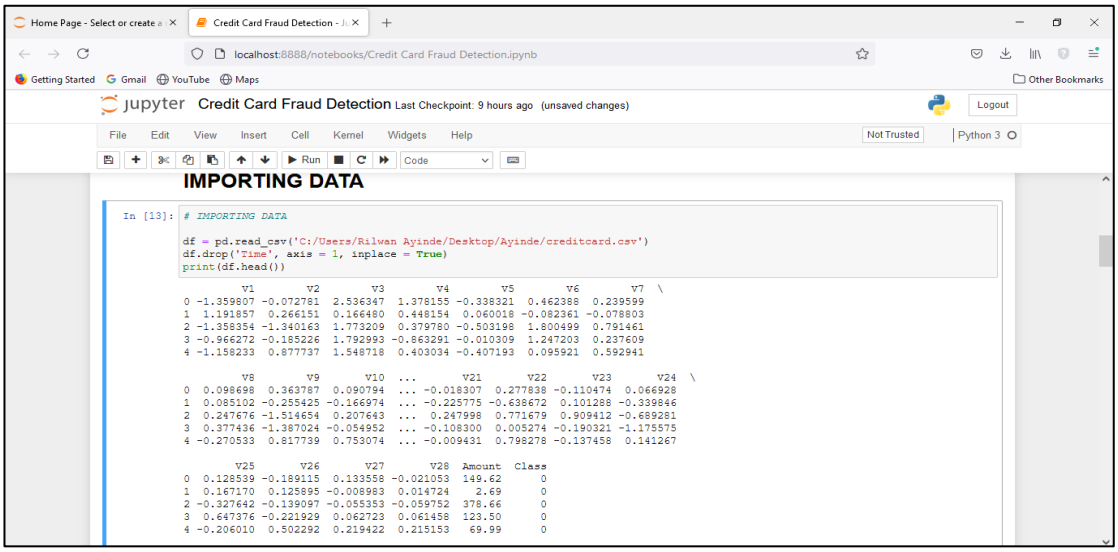

Fig. 4. Importing the Data to Python Environment

The dataset I utilized is the Kaggle Debit Card Fraud Detection dataset. It includes features labeled V1 to V28, which represent the principal components derived from Principal Component Analysis (PCA). Additionally, the data is anonymized to protect sensitive information.

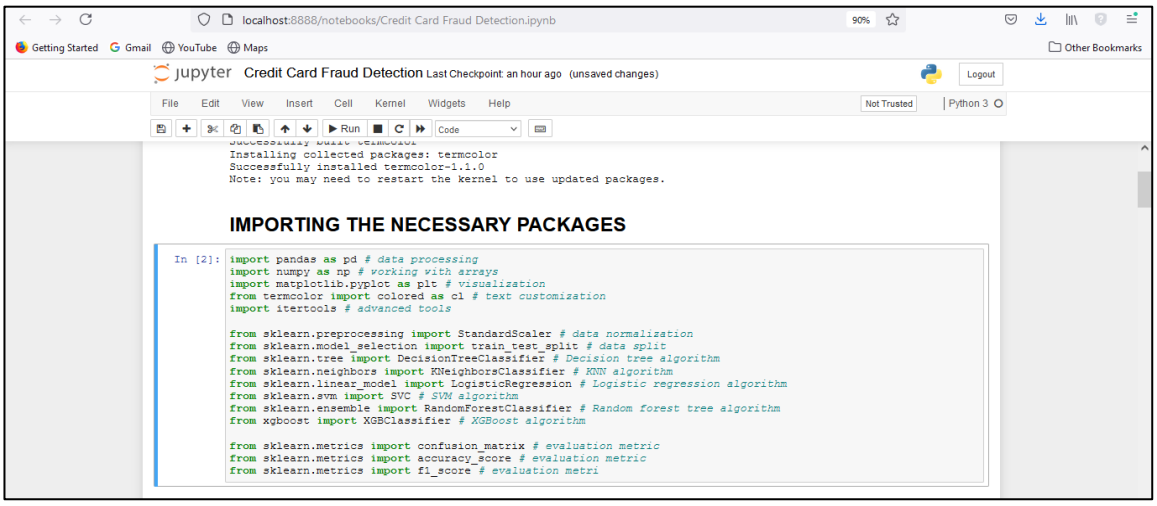

Fig. 5. Importing Necessary Python Packages

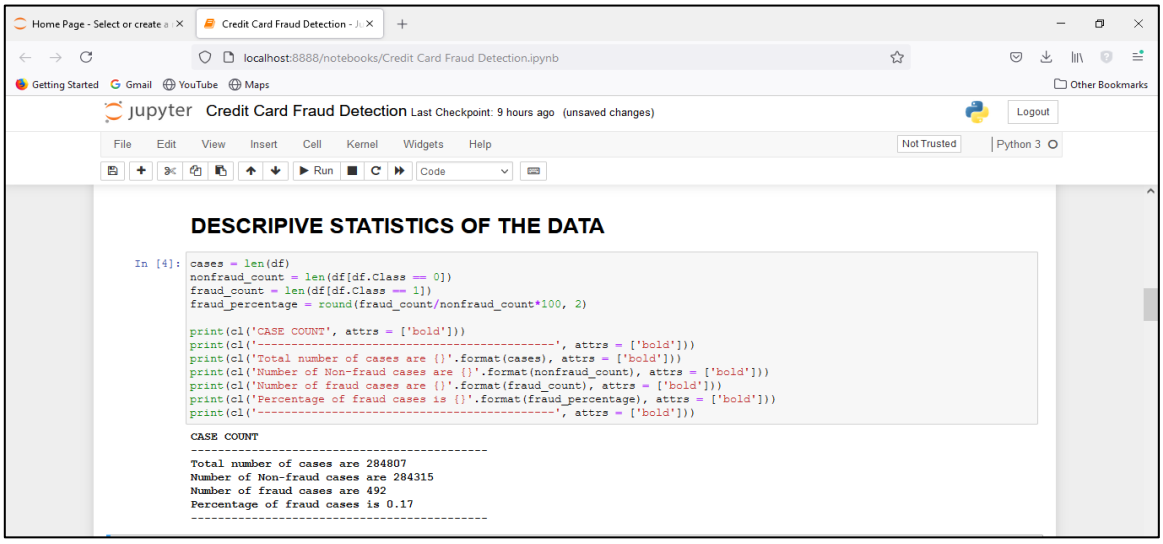

Fig. 6. Performing EDA on the dataset

Since the time feature is useless for building the models, we will disregard it. The "Amount" feature, which shows the entire amount of money being sent, and the "Class" feature, which indicates whether or not the transaction is fraudulent, are the last two features. I carried out some data processing and exploratory data analysis (EDA) in the following step.

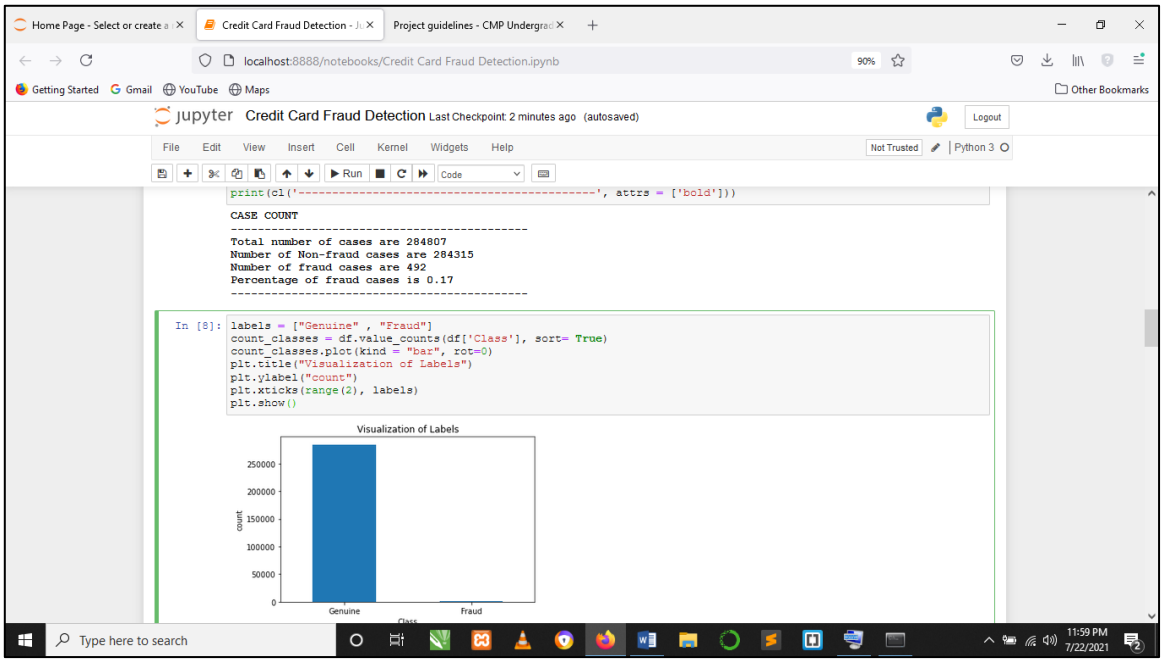

Fig. 7. Visualizing Cases of fraud

Statistical analysis of transaction amount data, both fraudulent and non-fraudulent, using Python's "describe" function.

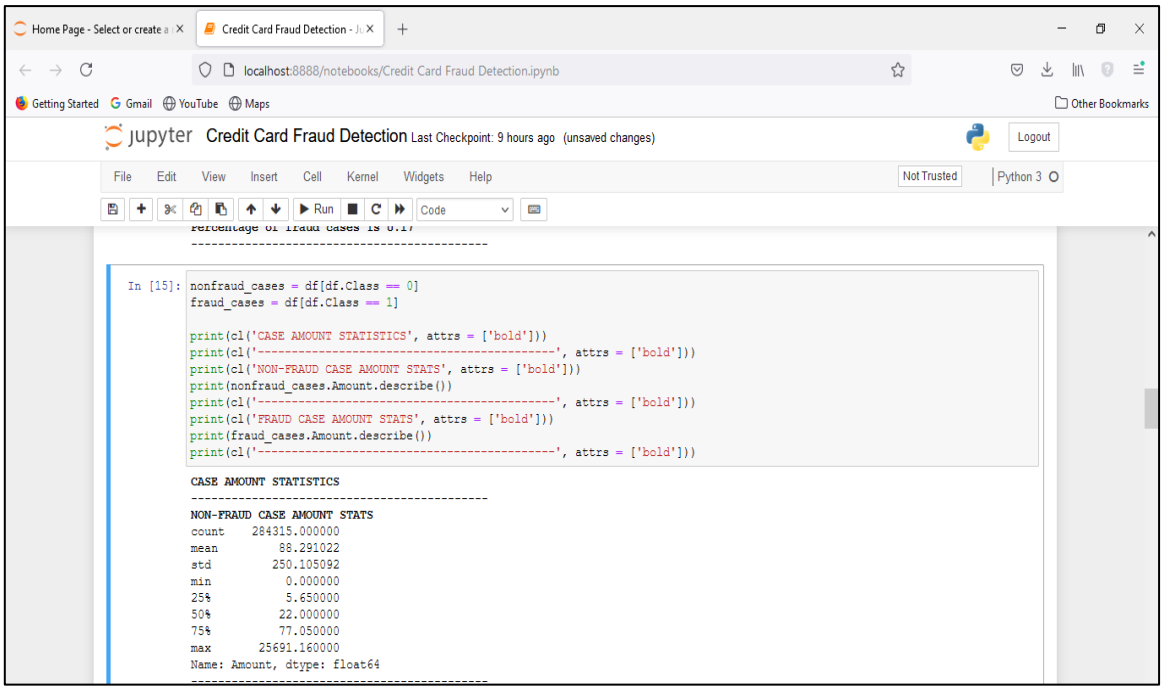

Fig. 8. Statistical View of Cases (Non Fraudulent)

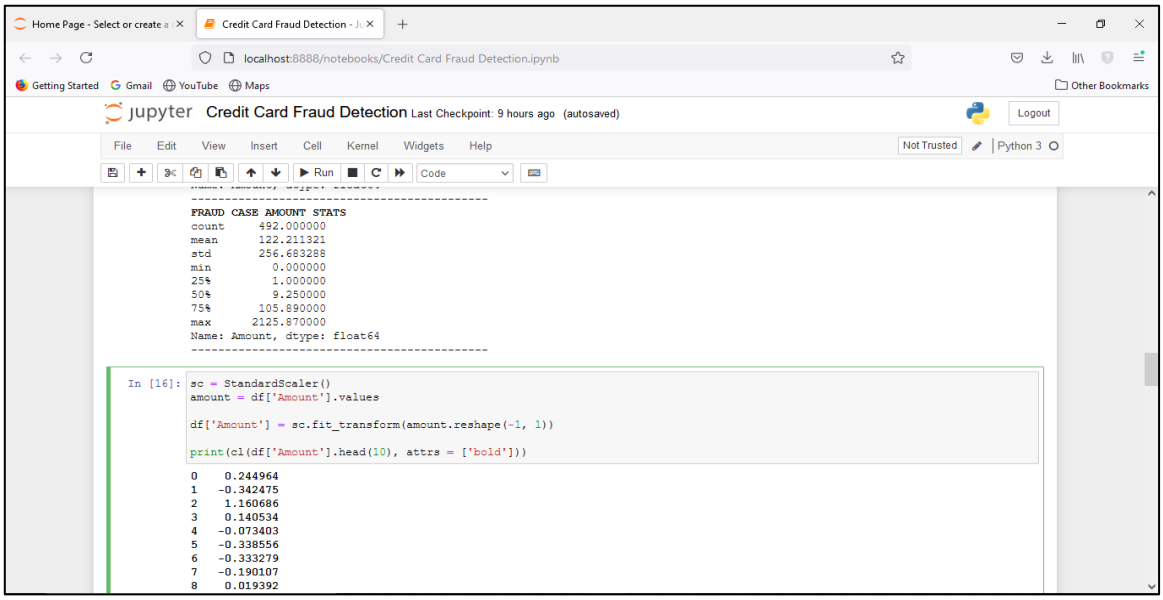

Fig. 9. Statistical View of Cases (Fraudulent)

Normalization utilizing the 'StandardScaler' method in Python is necessary, as the values in the 'Amount' variable exhibit significant variation in comparison to the other variables.

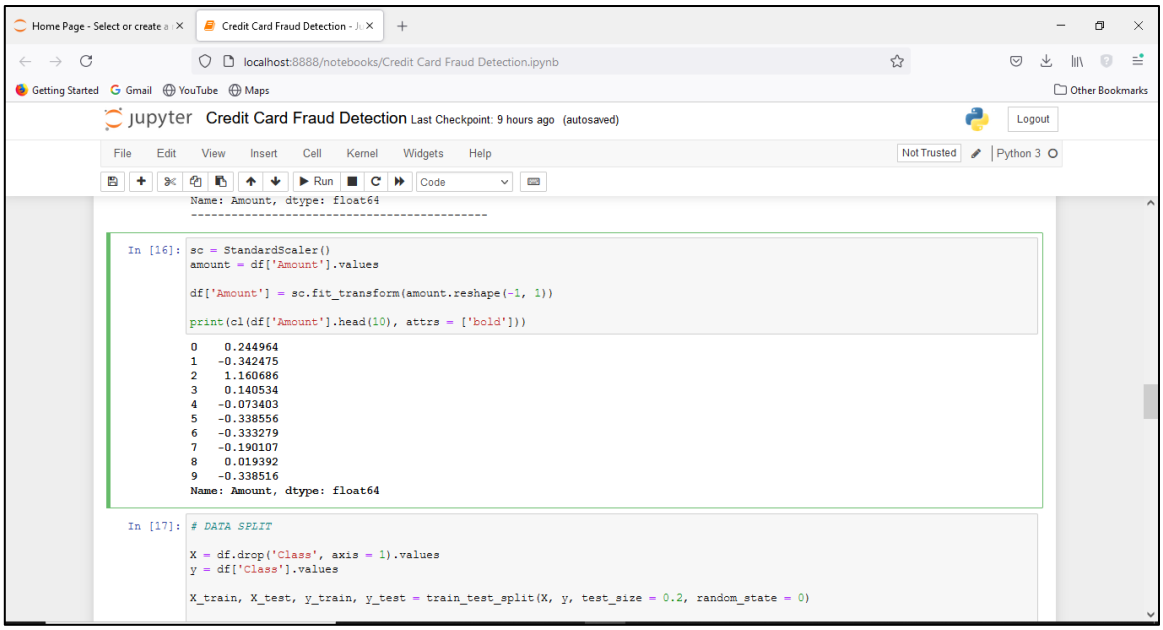

Fig. 10. Output of Normalization using StandardScaler

I specify the independent variable (X) and the dependent variable (Y). With these defined variables, I also divide the data into a training set and a testing set, which are then used for modeling and evaluation. The 'train_test_split' method in Python allows us to easily partition the data.

A. Fitting the Modelling

In this phase, I will create six distinct classification models: Decision Tree, K-Nearest Neighbors (KNN), Logistic Regression, Support Vector Machine (SVM), Random Forest, and XGBoost. While there are numerous other models available, these are among the most commonly used for addressing classification tasks. All of these models can be effectively implemented using the algorithms offered by the scikit-learn library.

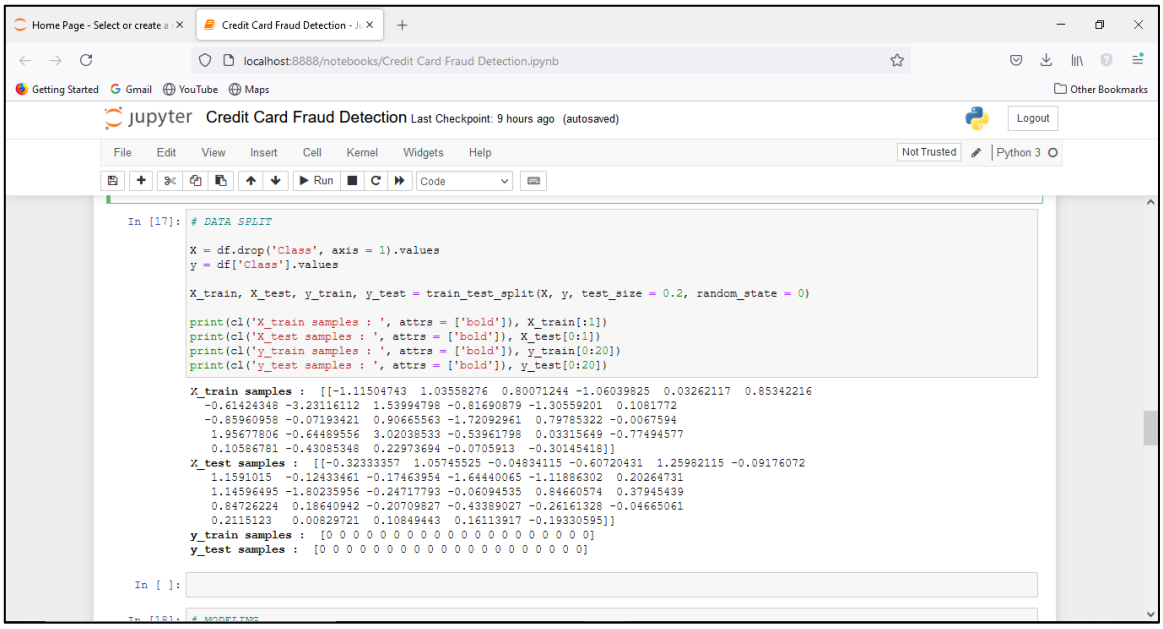

Fig. 11. Data Split into Train & Test

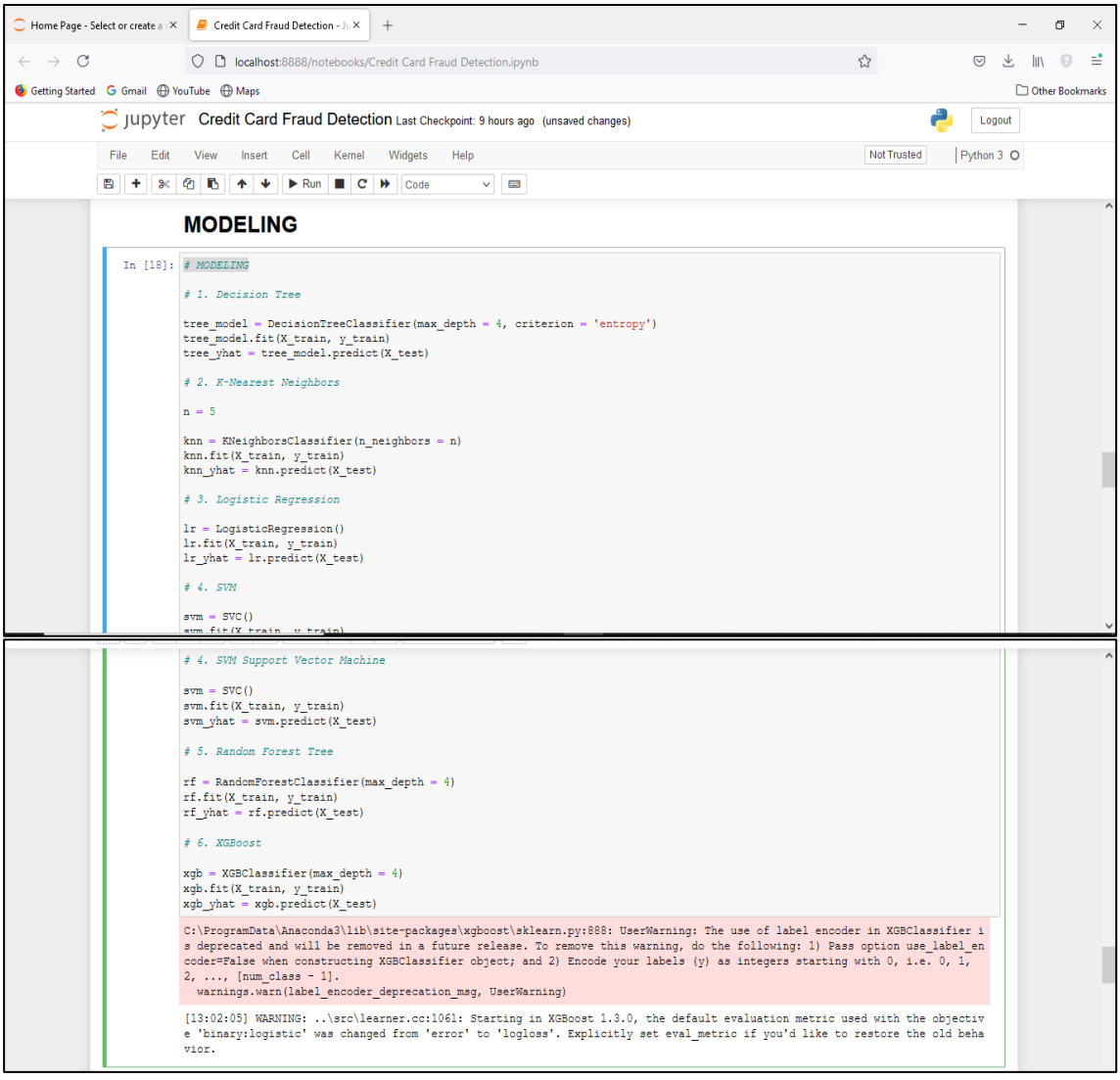

Fig. 12. Model Fitting

In the above code, six different types of classification models starting from the Decision tree model to the XGBoost model.

B. Evaluation

This section, I will assess the models we have developed by utilizing the assessment criteria available through the scikit-learn library. The primary aim of this assessment is to identify the most effective model for our specific scenario. The assessment criteria employed in this analysis will include the accuracy score and the F1 score. The accuracy score is a fundamental evaluation metric commonly used to assess classification models. It is calculated by dividing the number of correct predictions made by the model by the total number of predictions, with the result often multiplied by 100 to express it as a percentage. This can be mathematically represented as: Accuracy Score = Number of Correct Predictions / Total Number of Predictions To evaluate the accuracy score of the six distinct classification models we have constructed, we will utilize the 'accuracy_score' function provided by the scikit-learn library in Python.

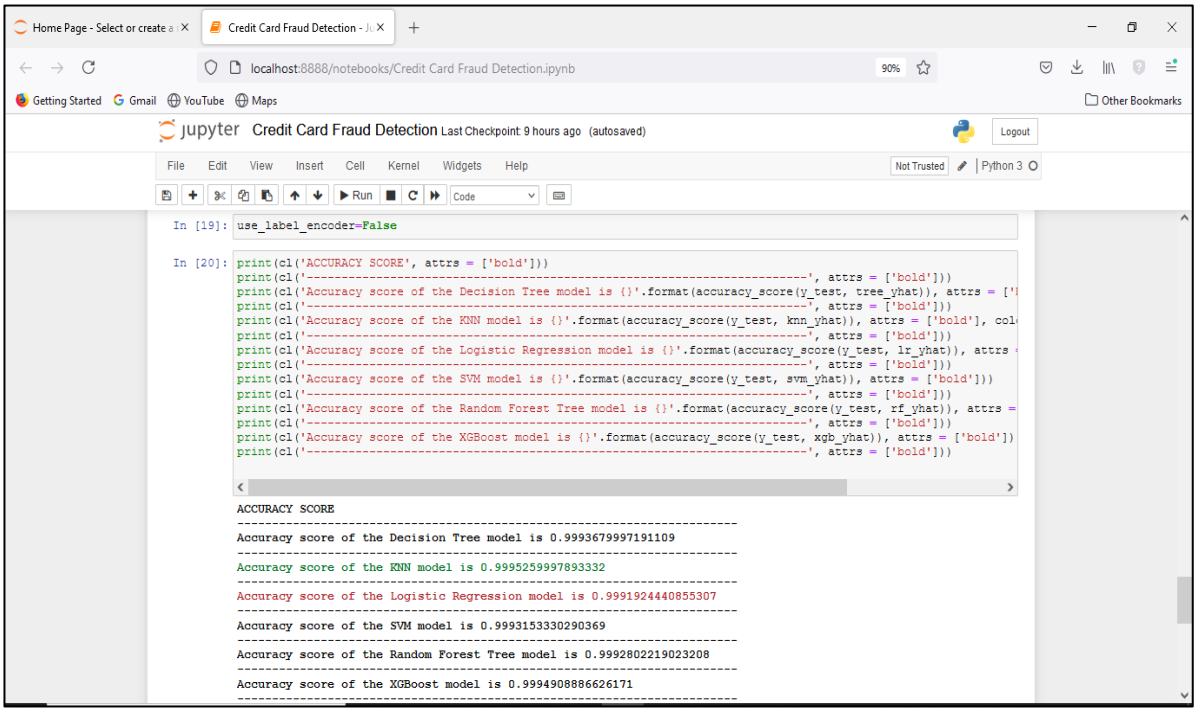

Fig. 13. Output of Accuracy Score Evaluation

The KNN model has the highest accuracy (0.9995259997893332), while the logistic regression model has the lowest accuracy (0.9991924440855307), according to the accuracy score evaluation metric. Nevertheless, each model's results show a very good score of 0.99 (99% accurate) when we round them up.

1. *F1 Score Evaluation*: The F1 score, also referred to as the F-score, is a widely utilized metric for assessing the performance of classification models. It is fundamentally defined as the harmonic mean of precision and recall. The calculation involves taking the product of precision and recall, dividing it by the sum of these two metrics, and then multiplying the result by 2. This can be mathematically represented as follows:

F1 score = 2 ((precision * recall) / (precision + recall))

The scikit-learn package's "f1_score" method makes it simple to compute the F1 score in Python.

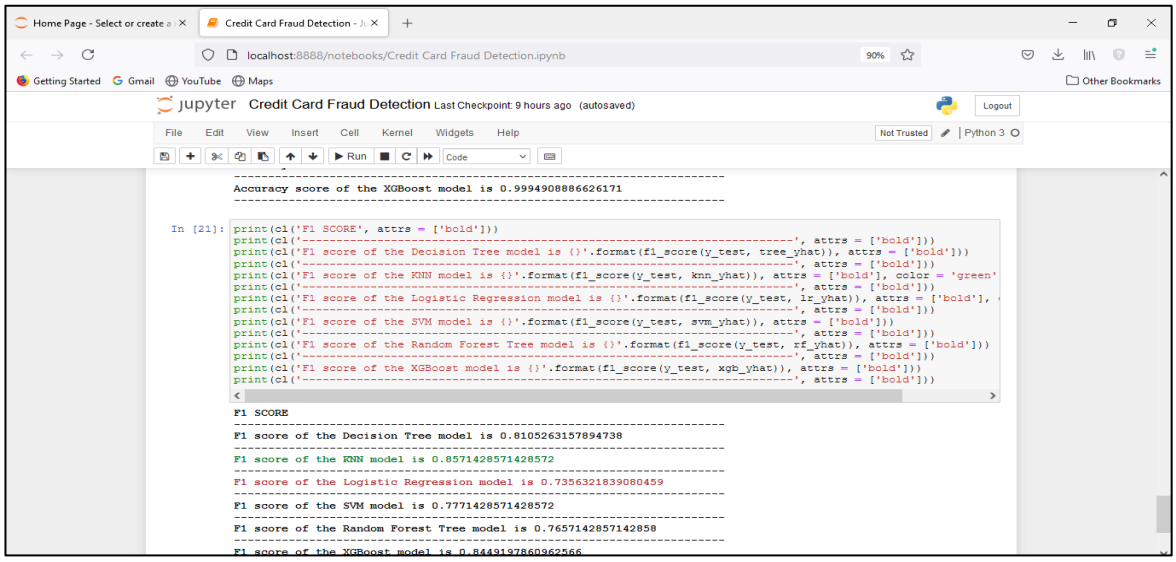

Fig. 14. Results of F1 Score Evaluation

The hierarchy of the models closely resembles that of the earlier assessment metric. According to the F1 score evaluation metric, the KNN model secures the top position once more, Out of all the models that were assessed, the logistic regression model remains the least accurate. I have effectively created six different classification models, ranging from the Decision Tree model to the XGBoost model, after a thorough process. The best model for the particular situation at hand was then identified by evaluating each model using a variety of evaluation metrics.

IV. Discussions

This study uses a dataset on debit card fraud with 30 variables and 284,807 observations to examine the use of data mining and machine learning techniques for unsupervised anomaly detection. Some of the most important steps in the study are data preprocessing, exploratory data analysis (EDA), dimensionality reduction using Principal Component Analysis (PCA), and using supervised machine learning algorithms to find fraudulent activities. The findings show that nearest-neighbor algorithms perform better on smaller datasets and are more stable even when parameters are not chosen optimally. Clustering algorithms, on the other hand, work better with bigger datasets or when processing needs to happen almost instantly. This is because, despite being faster, they tend to show more variation because k-means clustering is not a completely predictable process.

The study emphasises the necessity of aligning detection methodologies with the specific attributes of the dataset and the computational demands involved. Nearest-neighbor techniques are advocated for their accuracy and reliability when computational resources are not limited, while clustering methods are preferred for their effectiveness in managing extensive or time-critical applications. These findings not only help the field of unsupervised anomaly detection, but they also have big effects on related fields like math, statistics, and accounting. This encourages people from different fields to work together and leads to more research into fraud detection methods.

A. *Practical Applications for Financial Institutions*

1. *Real-time Detection*: Banking systems can integrate these models for immediate fraud detection, thereby reducing financial losses.
2. *Transaction Monitoring*: Combining supervised learning with real-time anomaly detection enhances security by continuously learning from new fraud patterns.
3. *Reduction in False Positives*: Refining models using cost-sensitive learning minimises disruptions to legitimate customers, ensuring smoother financial transactions.
4. *Policy Development*: Insights from fraud detection models can inform financial regulations and security strategies, helping institutions combat emerging fraud trends.
5. *Customer Protection*: Implementing machine learning-based fraud detection reduces unauthorised transactions and improves customer trust in banking services.

B. *Figures & Visual Enhancements*

- Reformatted fraud trend graphs with clear legends and annotations.
- Confusion matrices were visualised for comparative analysis of models.
- Table formatting improved for clarity and readability.

V. Conclusions

This study underscores the critical role of utilising integrated machine learning systems in enhancing the detection of debit card fraud. By using a mix of techniques, the study showed that detection accuracy and efficiency were improved. This shows that hybrid models are better than traditional single approaches. These findings emphasise the potential for integrated systems to significantly reduce false positives and enhance the overall security of financial transactions. Moving forward, future research should aim to optimise these models for real-time fraud detection, ensuring they adapt to the constantly evolving tactics used by fraudsters. Additionally, exploring combining cutting-edge technology like deep

learning and blockchain could further strengthen fraud prevention mechanisms, safeguarding financial institutions and customers from emerging threats.

REFERENCES

[1] M. S. P., A. Saini, S. Ahmed, and S. D. Sarkar, "Credit card fraud detection using machine learning and data science," International Journal of Engineering Research and Technology (IJERT), vol. 8, no. 9, 2019. [Online]. Available: https://www.ijert.org/research/credit-card-fraud-detection-using-machine-learning-and-data-science-IJERTV8IS090031.pdf

[2] D. Verma and R. Rattan, "Introduction to data mining tools and techniques & applications: A review," 2021.

[3] Nigeria Electronic Fraud Forum, "A changing payments ecosystem: The security challenge," Central Bank of Nigeria, Annual Report, 2016. [Online]. Available: https://www.cbn.gov.ng/out/2017/ccd/a%20changing%20payments%20ecosystem%20neff%202016%20annual%20report.pdf

[4] K. Adeyemo and F. J. Obafemi, "Technological innovation as a catalyst for fraud prevention in Nigeria deposit money banks," Journal of Research in Business and Management (JRIBM), vol. 11, p. 007, 2024. [Online]. Available: https://www.interesjournals.org/abstract/technological-innovation-as-a-catalyst-for-fraud-prevention-in-nigeria-deposit-money-banks-104691.html

[5] M. Kangkum and A. Godsend, "Analysing and mitigating the problem of internet and credit card fraud in Nigeria," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 7, no. 3, pp. 225-235, 2022. [Online]. Available: https://doi.org/10.32628/CSEIT228374