



Enhancing Fraud Detection in Financial transaction Through Big Data Analytics

Author:
Nousheen Fatima (P2719440)

Supervisor:
Ahmad Aladwan

*A thesis submitted in fulfilment of the requirements
for the degree of Master's in {MSc Data Analytics}
in the
De Montfort University
Leicester
United Kingdom*

Academic Year: 2022/2023

September 1, 2023

Declaration of Authorship

I, Nousheen Fatima, declare that this Dissertation Project titled, "Enhancing *Fraud Detection in Financial transaction Through Big Data Analytics*" and the work presented in it are my own knowledge gained from supervisor and learnt from courses in University during the course time.

I confirm that:

- work presented in this document was done on my own and not been copied from any other source except where duly acknowledged in respective section.
- any references to previously published material (found in books, journals, periodicals, the internet, etc.) are cited within the body of the report or reference section.
- I also consent to the storage and use of an electronic copy of this project for the prevention and detection of plagiarism

Signed: Nousheen Fatima
Date:01-09-2023

Abstract

The steadily developing landscape of financial transactions has led to a developing concern: fighting extortion. Utilizing the force of big data analytics, this exposition digs into an exhaustive investigation of upgrading misrepresentation recognition in financial transactions utilizing Python-based philosophies. The review starts with an exhaustive literature review, looking at customary extortion location strategies and featuring their limits despite developing deceitful methods. To address these difficulties, the examination lays out a proper system for executing Random Backwoods and choice tree classifiers, utilizing big data analytics for further developed precision.

Moral contemplations are a foundation of this review, underscoring protection, straightforwardness, and data security. The methodology segment frames data assortment, arrangement, and administration as basic moves toward building solid datasets for examination. In this manner, the discussion digs into the nitty gritty outcomes and examination of the exploration, exhibiting Python's job in executing complex calculations and envisioning data bits of knowledge. Key discoveries incorporate the ID of misrepresentation patterns connected to explicit card types, age groups, and temporal patterns.

The discussion section highlights the requirement for interdisciplinary joint effort, where financial organizations, administrative bodies, data researchers, and network safety specialists cooperate to foster extensive misrepresentation discovery systems. A forward-looking viewpoint frames expected regions for future exploration, including progressed AI strategies, blockchain incorporation, and client driven approaches. The conclusion causes to notice the significant job that big data analytics, joined with moral standards, plays in reinforcing financial framework honesty and keeping up with trust in the advanced economy.

Table of Contents

Chapter 1: Introduction:	8
1.1 Introduction:.....	8
1.2 Background study:	8
1.3 Research aim and objectives:.....	10
1.4 Research Problem:	10
1.5 Research Rationale:	11
1.6 Research structure:.....	11
1.7 Chapter Summary:	11
Chapter 2: Literature review:	13
2.1 Background:.....	13
2.2 Traditional fraud detection methods:.....	13
2.3 Big data analytics in fraud detection:.....	15
2.4 Data Preprocessing and feature selection:	17
2.5 Machine learning algorithms for detecting frauds:.....	18
2.6 Real-time fraud detection using big data analytics:	20
2.7 Big data analytics for Preventing fraud:	22
2.8 Evaluation and benchmarking of fraud detection systems:	24
2.9 Ethical and privacy considerations:	25
2.10 Case studies and real-world applications:.....	26
2.11 Literature Gap:	27
2.12 Summary:.....	28
Chapter 3: Methodology:	29
3.1 Introduction:.....	29
3.2 Research approach:	29

3.3 Research Philosophy:.....	30
3.4 Research design:	31
3.5 Research tools and techniques:	31
3.6 Data collection:	32
3.7 Ethical consideration:.....	33
3.8 Chapter Summary:	34
Chapter 4: Result and Analysis:.....	36
4.1 Introduction:.....	36
4.2 Analysis:	36
Chapter 5: Discussion:	62
Chapter 6: Conclusion:	65
6.1 Conclusive Introduction:.....	65
6.2 Recommendation:	65
6.3 Future Scope:	67
6.4 Summary:	68
Reference:	70

List of Figures

Figure 1.6.1: Research Structure.....	11
Figure 2.2.1: Fraud Detection Algorithm	15
Figure 2.3.2: Big Data Analytics for Fraud Detection and Prevention.....	16
Figure 2.4.3: Workflow developing and processing of data.....	18
Figure 2.5.4: Algorithms for detecting fraud	19
Figure 2.6.5: Fraud Detection with Prevention.....	21
Figure 2.7.6: Big Data Analytics for Fraud Detection with Prevention	23
Figure 4.2.1: Importing libraries	36
Figure 4.2.2: Reading dataset.....	37
Figure 4.2.4: Displaying first few rows	39
Figure 4.2.5: Removing “#” from transaction ID	39
Figure 4.2.6: Check for null values.....	40
Figure 4.2.7: Replacing null values	41
Figure 4.2.8: Distribution of Fraudulent and Non fraudulent transactions.....	41
Figure 4.2.7: Fraud by card.....	42
Figure 4.2.8: Fraud according to age	43
Figure 4.2.9: Violin Plot Visualization	44
Figure 4.2.11: Distribution of fraud and non-fraud transactions by country	45
Figure 4.2.12: Distribution of Fraudulent transactions by bank	46
Figure 4.2.13: Importing label encoder function	47
Figure 4.2.14: transforming the numerical values	47
Figure 4.2.15: Viewing the new data frame.....	48
Figure 4.2.16: separating the independent and dependent variables	48
Figure 4.2.17: viewing the Y value.....	49
Figure 4.2.18: Importing different library functions.....	50
Figure 4.2.19: Data frame splitting into test and train	51
Figure 4.2.20: viewing the data frame information	52
Figure 4.2.21: checking the unique values from the type of card column.....	53
Figure 4.2.22: converting the data of “Type of card” column	54

Figure 4.2.23: viewing the new data frame head	55
Figure 4.2.24: Implementing the decision tree classification	56
Figure 4.2.25: Precision score, recall score and f1 score of decision tree classifier.....	56
Figure 4.2.26: Random Forest Accuracy	57
Figure 4.2.27: Precision score, recall score and f1 score of Random Forest classifier	57
Figure 4.2.28: Precision score, recall score and f1 score of Logistic Regression.....	58
Figure 4.2.29: Generating the confusion matrix of Random Forest	59
Figure 4.2.30: Generating the confusion matrix of Decision Tree	59
Figure 4.2.31: Generating the confusion matrix of Logistic Regression	60
Figure 4.2.32: Comparison of ML models.....	61

Chapter 1: Introduction:

1.1 Introduction:

Financial transactions have grown more complicated and vulnerable to fraud in the current digital era. Financial institutions must always be one step ahead in identifying and blocking fraudulent transactions because as technology develops, so do the techniques used by fraudsters. Big data analytics has made it possible for significantly improving the fraud detection in the financial transactions. Big data analytics seems to be the procedure of looking through along with gleaning insightful knowledge from enormous amounts of structured as well as unstructured data. Financial institutions can also analyze the transactional information, customer information, previous trends, and other data sources through utilizing the power of big data for uncovering the anomalies, patterns, or the trends related to some fraudulent activity. As a result, they can proactively identify or stop some fraudulent transactions in the real time.

The capabilities of big data analytics in case of handling enormous volumes of information at such a high velocity has been of the major benefits of utilizing them in the fraud detection. The increasing number and speed of financial transactions in a frequent manner can make it difficult for the traditional fraud detection techniques for keeping up. However, the financial institutions can handle as well as analyze enormous datasets in close to real-time through utilizing big data technologies including distributed computing or the machine learning algorithms by greatly enhancing the effectiveness or efficiency of fraud detection. Furthermore, this sort of data analytics can integrate many data sources for providing a comprehensive view of the client behavior. Financial companies may create thorough client profiles as well as identify baseline patterns of the typical behavior through combining or analyzing data from different touchpoints, including the history of transaction, online activities, social media, and geolocation. The detection of some odd behaviors or departures from the established patterns is possible through this comprehensive methodology, raising the red flag for the probable fraudulent transactions.

1.2 Background study:

Numerous advantages, including efficiency and ease, have been brought about by the growing digitization of financial transactions. However, it has also given criminals new ways to

take advantage of weak points and commit crimes. Financial fraud, which poses serious threats to both financial institutions and their clients, is reportedly on the rise globally. Effective fraud detection strategies are now essential in light of this expanding threat (Singarimum *et al.* 2022). Modern financial transactions have become extremely complex, along with that traditional techniques of fraud detection that mainly rely on the manual assessments Or rule-based algorithms, frequently fall short to address them. Additionally, due to these techniques become reactive in nature, they catch the fraud only after it has already happened by resulting to a large financial losses. Therefore, it has become important using cutting-edge technologies for the improvement of fraud detection capabilities as well as keeping up with crafty fraudsters (Chen and Lai, 2021). Financial institutions nowadays have the access to enormous volumes of structured or unstructured information that could be used for fraud detection to the proliferation of data in this digital age. Big data analytics can use sophisticated algorithms along with data processing procedures to draw some valuable conclusions from sizable or varied datasets.

Financial organizations can improve their fraud detection skills greatly by utilizing big data analytics. The hidden trends and abnormalities that are able to point to the fraudulent activity can be uncovered by the analysis of transactional information, customer information, historical patterns, along with some other data sources. Large datasets can also be used for training the machine learning algorithms to spot subtle patterns and adjust to changing fraud tactics by increasing the precision or the effectiveness of those fraud detection. Additionally, real-time fraud detection has made it possible through big data analytics can enable the financial institutions for seeing and reporting potentially fraudulent transactions as it take place (Taha *et al.* 2020). Through reducing the window of opportunity for the fraudsters, this real-time monitoring can enable prompt response along with the reduction of financial losses. Furthermore, a thorough understanding of the client behaviors that are provided through the integration of several data sources by this data analytics. Some financial companies can create extensive client profiles as well as establish baseline patterns of typical behavior by taking into account information from multiple touchpoints including transaction history, internet activities, social media, or geolocation. The institution then would be made aware of any sort of deviations or odd activities by alerting it to some possible fraudulent transactions.

1.3 Research aim and objectives:

➤ Aim

The purpose of this study is to investigate and assess how well big data analytics may be used to improve fraud detection in financial transactions.

➤ Objectives

- To investigate the present state of the fraud detection processes in the financial institutions along with identifying its limitations and challenges
- To explore all sorts of potential applications of the big data analytics to improve the accuracy, monitoring capabilities in real time and efficiency of fraud detection
- To develop a model or framework that can integrate this data analytics techniques with already existing system of fraud detection for enhancing the abilities of fraud detection
- To evaluate the effectiveness of those proposed framework by empirical analysis along with comparative studies through measuring the consequences of big data analytics on the fraud detection performance along with identifying the areas that need improvement.

1.4 Research Problem:

The need for improved fraud detection in financial transactions and the potential use of big data analytics as a solution are the research problems covered in this paper. Traditional fraud detection techniques are frequently ineffective at identifying and stopping fraudulent activity, which results in significant financial losses and brand harm. It is difficult for conventional rule-based systems to keep up with evolving fraud schemes as a result of the advent of digital transactions, which have increased the complexity and scale of financial data. Additionally, con artists are continually modifying their strategies and coming up with new ways to exploit weaknesses (Benchaji *et al.* 2020). There is an urgent need to investigate cutting-edge technology that can offer real-time, precise, and preemptive fraud detection capabilities, such big data analytics. Massive amounts of structured and unstructured data could be used by big data analytics to find patterns, anomalies, and trends connected to fraudulent activity. Financial institutions may greatly enhance their fraud detection accuracy, efficiency, and speed by utilising the power of cutting-edge algorithms and machine learning approaches. Despite the potential advantages, there remain difficulties and knowledge gaps in the implementation of big data analytics into current

fraud detection systems. This study has aimed to close these gaps and help build a complete framework or model that uses big data analytics to improve fraud detection in financial transactions.

1.5 Research Rationale:

The necessity to address the rising risk of fraudulent activity in financial transactions and the potential for big data analytics for improving fraud detection capabilities are the driving forces behind this research. The need for novel strategies that can make use of real-time monitoring and advanced analytics is crucial since standard fraud detection systems fail to keep up with evolving fraud schemes (John and Naaz, 2019). This project has intended to contribute to the creation of more effective and proactive techniques for identifying and preventing fraudulent transactions by examining the applications of big data analytics in fraud detection. The results of this study can offer useful insights to financial institutions, assisting them in bolstering their anti-fraud measures and safeguarding both their resources and the interests of their clients.

1.6 Research structure:

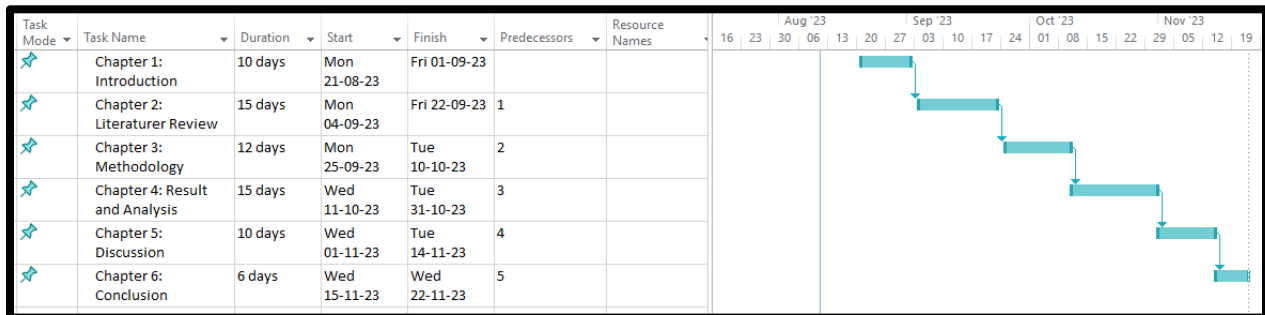


Figure 1.6.1: Research Structure

1.7 Chapter Summary:

The research topic of improving fraud detection in financial transactions through big data analytics was introduced in this chapter. It brought to light the growing complexity and susceptibility of financial transactions to fraud. The ability of big data analytics to manage massive volumes of data, offer real-time monitoring, and combine numerous data sources was highlighted as a potential solution to these problems. The research goals were also described in the chapter, and they include

examining current fraud detection techniques, looking into the uses of big data analytics, creating a framework, and assessing its efficiency.

Chapter 2: Literature review:

2.1 Background:

Frauds in monetary(finianical) exchanges allude to tricky exercises directed fully intent on acquiring unlawful monetary advantages, associations, or frameworks. Different sorts of frauds exist in monetary exchanges. Fraudsters take individual data like government managed retirement numbers or details of financial balance, to fraud people and complete the unapproved exchanges. Unapproved utilization of Mastercard data of another person for making buys or pulling out cash without the assent of cardholder (Masarani, 2021). The most common way of bringing in illicitly gotten cash seem real by masking its actual source through a perplexing(confusing) series of exchanges.

Analysis of big data includes the method involved with looking at enormous volumes of information which is known as big data, to reveal the patterns that are hidden, relationships, and experiences that can work with informed independent direction. It use advanced strategies, including information mining, AI, and analysis for predicting, to separate important data from immense and various datasets. Applications in different businesses, including finance, medical services, retail, and media communications have been tracked down by analysing big data (Ragazou *et al.* 2022). Grasping knowledge about the definition and kinds of extortion in monetary exchanges and the outline of enormous information examination and its applications in different ventures, features the significance of misrepresentation recognition in keeping up with monetary security, limiting misfortunes, safeguarding the trust of the client, following guidelines, and protecting the reputation.

2.2 Traditional fraud detection methods:

Financial transactions have traditionally relied heavily on manual procedures and systems that use rules for identifying fraud. But there's a chance to greatly improve fraud detection capacities with the introduction of big data analysis. The term “big data analytics” describes the method of poring over enormous and intricate(complicated) databases to find correlations, trends, and abnormalities that might point to criminal activity.

The capability of big data analytics to deal with enormous quantities of information in actual time is one of its primary benefits in fraud detection. Large amounts of information, comprising transactional facts, customer data, and previous trends are produced by monetary transactions. Businesses can evaluate such information in real-time by using big data analytics tools, which enables them to immediately spot unusual activity and perhaps fraud.

Using large data analytics, machine learning algorithms are essential for improving fraud detection. For discovering the trends and patterns related to fraudulent transactions, these computer programs can be educated on past information. Machine learning algorithms can recognize anomalies and flag possibly fraudulent transactions for additional investigation by continually tracking transactions that arrive and contrasting these to the taught characteristics. This method greatly lowers the amount of false positives and improves the precision of identifying fraudulent activity.

The capacity to conduct sophisticated network analyses is another benefit of big data analytics for fraud detection. Fraudsters frequently work in complex networks with numerous people taking part in their fraudulent actions. Organizations can more easily spot clusters of questionable activity and uncover fraudulent networks by examining transactions information as well as social relationships (Benedek *et al.* 2023).

Big data analytics can use information from multiple places to identify corruption in real-time. Financial organizations can combine private data on transactions with information from outside organizations including social networking feeds, open data, and databases used to track widespread fraud. Organizations can develop a more thorough picture of client behavior and more precisely pinpoint potential fraud threats by utilizing a variety of data.

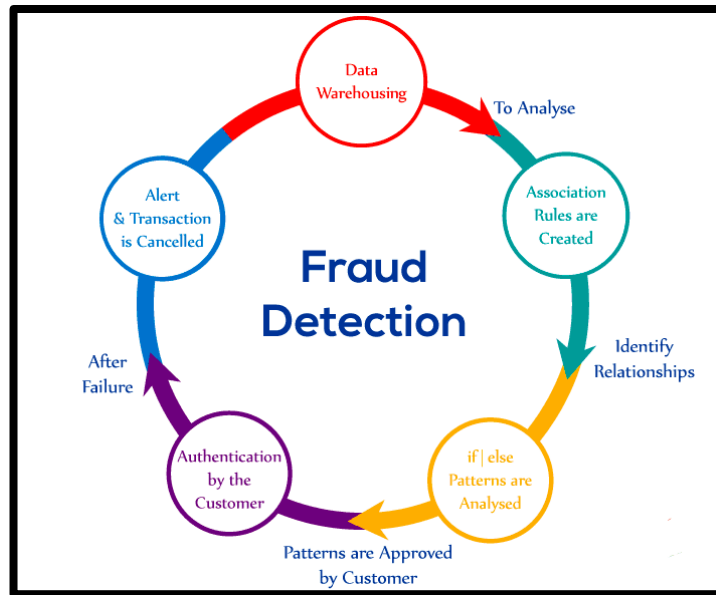


Figure 2.2.1: Fraud Detection Algorithm

2.3 Big data analytics in fraud detection:

Huge amount of information investigation is a field of study that highlights on the variety, examination, and understanding of enormous and complex datasets. This data can be come from transaction logs, social media data, and customer behavior data, among other places. Analytics approaches for big data can be used to find patterns, trends, and anomalies in data that could signs of fraud.

The following represent some of the more popular big data analytics approaches to fraud detection:

- Finding trends and patterns in enormous amounts of information is a technique known as mining data. By identifying characteristics that are incompatible with true transactions, these can be utilized to spot unauthorized transactions.
- Machines can learn from data thanks to a sort of artificially intelligent technology called machine learning. Using this, models that forecast the possibility of frauds can be created.
- Data analysis that analyzes past information to forecast future results is known as statistical analysis. By searching for patterns linked to fraudulent transactions, information might be utilised to estimate the probability of fraudulent activity (Himeur *et al.* 2023).

Compared to conventional fraud detection techniques, big data analytics provides a variety of benefits. These benefits consist of:

- **The capacity to study enormous datasets:** The size of the information sets that existing fraud detection techniques can evaluate is frequently a limitation. On the other hand, big data analysis can be utilized for analyzing huge and complicated datasets that are impractical to analyze with conventional techniques.
- **The capacity to recognize complicated patterns:** Conventional fraud detection techniques frequently only have the capacity to recognize simple patterns. On the other hand, big data analyses can be utilized to spot complicated patterns that might be signs of fraud.

Fraud detection advantages from big data analytics approaches like machine learning, data extraction, and statistical analysis are considerable (Ariyaluran Habeeb *et al.* 2022). They give businesses the ability to manage huge amounts of information, find patterns which is undetected, anticipate incidents of fraudulent activity, and combine various sources of data for more accurate and effective analysis. Organizations may be improve their criminal activity identification skills and safeguard customers from financial consequences and damaging reputations by utilizing the potential of big data analysis.

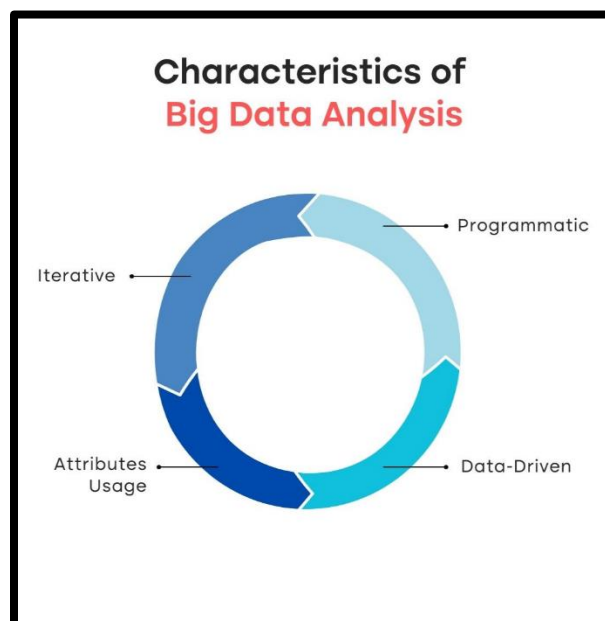


Figure 2.3.2: Big Data Analytics for Fraud Detection and Prevention

2.4 Data Preprocessing and feature selection:

Processing information is a crucial part of the fraud detection method. It entails preparing the information for analysis by cleaning and converting it. As part of this, vibration, unusual values, and insufficient data must be eliminated. It also entails formatting the information in a way that algorithms using machine learning may use it.

The procedure of choosing the most significant characteristics from the data is known as choosing features. This is significant since not each of the features are equally essential for detecting fraud. Incorporating irrelevant or noisy features into an evaluation may actually decrease the model's precision.

There are numerous methods for choosing features that can be applied. Among the most popular methods are:

- **Filter techniques:** These techniques pick characteristics according to their analytical characteristics. As an illustration, a filtering technique could choose characteristics having a significant relationship to the desired statistic.
- **Wrapping techniques:** These techniques choose features by creating a model and assessing the value for every characteristic. For instance, a wrapper technique could create a decision tree before choosing the characteristics with the greatest weight.
- **Methods of embedded:** These techniques pick characteristics as a component of the model-building procedure known as imbedded techniques. By minimizing the separation among the groups, an SVM (support vector machine) classifier, for instance, might be used to choose characteristics (Prabhakaran *et al.* 2023).

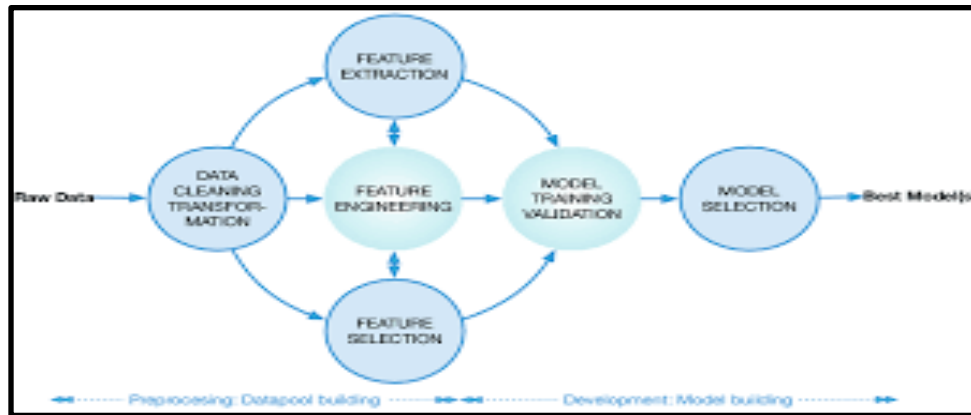


Figure 2.4.3: Workflow developing and processing of data

A variety of methods can be applied to deal with imbalanced datasets. Among the most popular methods are:

- In order to equalize the information set, the oversampling requires making duplicates of unauthorized transactions.
- In order to equalize the dataset, inadequate sampling requires eliminating some of the valid transactions.
- Different amounts of weight to the different groups in the information set. The model can now concentrate on more of the minority class (fraudulent operations) as a result (Yin *et al.* 2023).

Important components in fraud detection contain choosing features and data preparation. Through removing outliers and information which is irrelevant from the data and choosing the most essential attributes, they may be helpful in improving the reliability of the fraud identification algorithms. Another significant complexity in identifying fraudulent activity involves handling data sets that are imbalanced.

2.5 Machine learning algorithms for detecting frauds:

- "*Supervised learning algorithms*" are utilized in the identification of frauds to gain from the information that is labeled, which is information that has been named either false or not

deceitful. This kind of learning is appropriate for extortion recognition since it permits the model to become familiar with examples that recognize deceitful and non-fake exchanges. The most normal "*Supervised learning algorithms*" utilized for detecting fraud include:

- **Logistic Regression:** This calculation is utilized to foresee the likelihood of an occasion happening, like a deceitful exchange.
- **Decision trees:** This algorithm constructs a tree-like design that addresses the connections between various elements of a transaction.
- **Support vector machines (SVMs):** The algorithm checks for the best hyperplane that isolates the transactions if they are fraudulent or not (Afriyie *et al.* 2023).

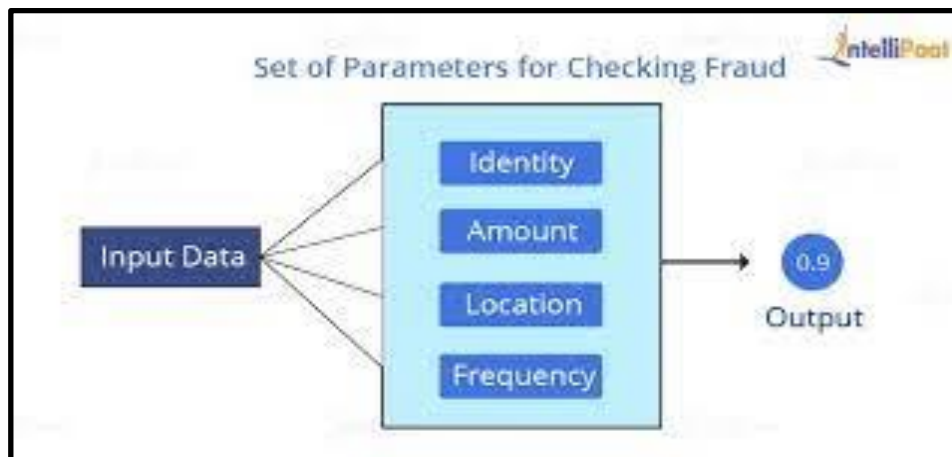


Figure 2.5.4: Algorithms for detecting fraud

- "*Unsupervised learning algorithms*" help detect fraud to track down the patterns in unlabeled information. This sort of learning is appropriate for recognizing frauds since it can recognize fake exchanges that have not been seen previously. Some of the "*Unsupervised learning algorithms*" utilized for identifying fraud include:
 - **Clustering:** The algorithm gathers similar types of exchanges together. Fake transactions can frequently be distinguished by the way that they are grouped.
 - **Anomaly detection:** This algorithm recognizes the transactions that are exceptions, or that don't fit the typical example of conduct. False exchanges are much of the time exceptions.

Ensemble techniques join the expectations of various models to work on the general execution of the framework. This is frequently finished by joining various kinds of models, like

supervised and unsupervised models (Kumar *et al.* 2022). Hybrid models involve various sorts of learning calculations inside a single model. This should be possible to work on the exhibition of the model or to make it more hearty to changes in the information. AI is an incredible asset for extortion identification. Monetary foundations can work on their capacity to identify deceitful exchanges and safeguard their clients from monetary misfortune by utilizing AI. The decision of calculation will rely upon the particular informational index and the ideal presentation.

2.6 Real-time fraud detection using big data analytics:

Continuous fraud identification is the method involved with recognizing the deceitful transactions as they happen. This is rather than conventional strategies for detecting fraud, which commonly includes historical information to recognize examples of false ways of behaving.

There are various difficulties and considerations for continuous extortion identification. These include:

- ***The volume of information:*** Financial foundations produce an enormous measure of information consistently. This information incorporates exchanges, client data, and information about the device. Systems for detecting real-time fraud should have the option to deal with this information continuously to recognize deceitful exchanges.
- ***The speed of discovery:*** Fraudsters are continually developing their strategies. To stay aware of them, these frameworks should have the option to distinguish deceitful exchanges as fast as could be expected.
- ***The exactness of identification:*** Misleading positives can harm the relationships of the client. False negatives can prompt monetary misfortunes. these detection systems should have the option to work out some kind of harmony between precision and speed (Bin Mofidul *et al.* 2022).

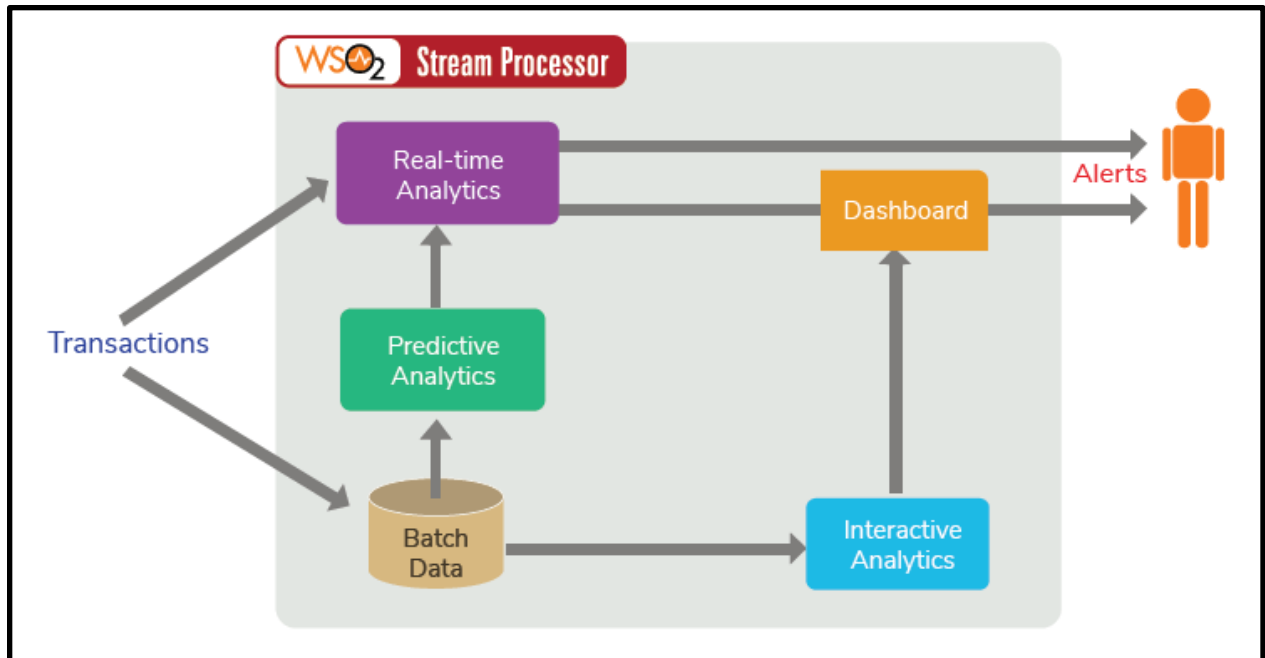


Figure 2.6.5: Fraud Detection with Prevention

Stream handling is a procedure for handling information as it shows up. This includes handling information in bunches. Stream handling is appropriate for continuous extortion recognition since it permits the framework to deal with information as it shows up, which is fundamental for recognizing deceitful exchanges rapidly.

There are various stream-handling procedures and calculations that can be utilized for ongoing fraud discovery; These include:

- **Anomaly identification:** This strategy distinguishes exchanges that are anomalies, or that don't fit the typical example of conduct. Deceitful exchanges are many times anomalies.
- **Machine learning:** This strategy utilizes verifiable information to prepare a model that can recognize false exchanges.

Systems that are Rule-based: These frameworks utilize a bunch of rules to recognize fake exchanges (Velasco-Gallego *et al.* 2022).

There are various variables that should be thought about while executing a real-time fraud detection system. These include:

- The kind of information that will be utilized.
- The stream handling strategy or calculation that will be utilized.

- The presentation prerequisites of the framework.
- The expense of the framework.
- The presentation of a system for detecting fraud can be assessed utilizing various measurements, for example,

- ***False positive rate:*** This is the level of authentic exchanges that are inaccurately distinguished as deceitful (untrustworthy).
- ***Misleading negative rate:*** This is the level of fake exchanges that are mistakenly recognized as genuine.
- ***Discovery time:*** This is the time it takes for the framework to recognize a deceitful exchange.

2.7 Big data analytics for Preventing fraud:

Informal community investigation is a procedure that can be utilized to recognize and plan connections between people or elements. This strategy can be utilized to distinguish coordinated extortion networks by searching for examples of correspondence or cooperation between known fraudsters. “*Artificial intelligence (AI)*” and “*deep learning*” are two quickly creating fields that can possibly change the fraud location. AI and deep learning can be utilized to foster more refined prescient models that can distinguish fake exchanges with more noteworthy exactness. Simulated intelligence and profound learning can be utilized to robotize extortion location errands, like information examination and example acknowledgment (Muheidat *et al.* 2022). The utilization of prescient displaying, informal community examination, and AI and deep learning is turning out to be progressively normal in extortion location. These strategies can be utilized to distinguish potential misrepresentation designs, identify coordinated extortion organizations, and computerize extortion discovery errands.

- “*Social network analysis (SNA)*” is a procedure that can be utilized to distinguish and plan connections between people or substances. This procedure can be utilized to distinguish coordinated misrepresentation networks by searching for examples of correspondence or cooperation between known fraudsters.

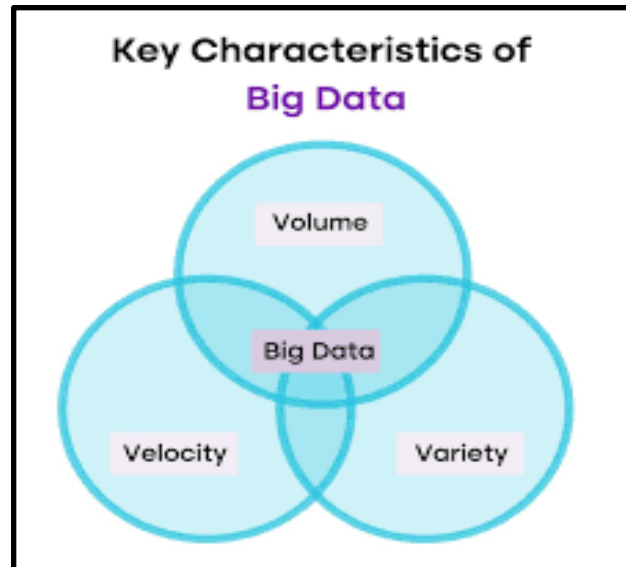


Figure 2.7.6: Big Data Analytics for Fraud Detection with Prevention

SNA can be utilized to recognize misrepresentation networks by searching for:

- **Normal contacts:** Fraudsters frequently share contacts, for example, email addresses, telephone numbers, or IP addresses.
- **Successive correspondence:** Fraudsters frequently speak with one another, either straightforwardly or through middle people.
- **Comparative way of behaving:** Fraudsters frequently show comparable ways of behaving, like making huge, strange exchanges or utilizing different records (Xu *et al.* 2023). **SNA** can be a significant device for distinguishing coordinated extortion organizations. Nonetheless, it is essential to take note of that **SNA** is definitely not a secure strategy. Fraudsters can do whatever it may take to camouflage their exercises, making it hard to distinguish them utilizing **SNA**. **SNA** is an incredible asset that can be utilized to identify coordinated misrepresentation organizations. Nonetheless, it is essential to involve it related to different strategies, for example, prescient displaying and rule-based frameworks, to work on the precision of misrepresentation identification.

2.8 Evaluation and benchmarking of fraud detection systems:

Big data analytics can enhance fraud detection in financial transactions through evaluating and benchmarking fraud detection systems using performance metrics, using benchmark datasets and competitions, and comparing various fraud detection methods.

Performance Indicators for Models of Fraud Detection:

Various criteria for performance may be used to judge the efficacy of fraud detection computation. Perhaps the often employed metrics are:

- Measures to assess how precisely the model's predictions were made overall.
- Measures the proportion of fraudulent transactions that were precisely identified out of all fraudulent transactions that were found (Sadgali, I *et al.* 2019).
- **Recall:** Determines the proportion of all fraudulent transactions that were successfully detected as fraud.
- **F1 Score:** The harmonic mean of recall and accuracy, providing an equitable assessment of the two.

In order to choose the best threshold for fraud detection, the Receiver Operating Characteristic (ROC) Curve plots the true positive rate against the false positive rate.

- **Area AUC:** Evaluate the model's overall performance, with a higher AUC indicating better results.

Benchmark Datasets and Competitions in Fraud Detection:

There are multiple datasets and disputes that offer standardised data and evaluation procedures for measuring fraud detection systems. Examples that stand out include:

- The Credit Card Fraud Detection dataset offers a standard dataset to evaluate fraud detection algorithms by including anonymized credit card transactions that have been classified as fraudulent or not.
- The IEEE Computational Intelligence Society sponsors the annual Kaggle competition IEEE-CIS Fraud Detection Competition, that emphasises discovering fraudulent online transactions.

- Dataset from FRAUDAR: a publicly available dataset with simulated transactions that include fraud and non-fraud, developed to test fraud detection approaches in huge graphs.

Comparative Analysis of Different Fraud Detection Approaches:

Strategy	Description	Techniques/Algorithms Used
Rule-based systems	Identify questionable transactions based on established standards or patterns.	Rule-based logic
Supervised ML	Train models using annotated data to classify transactions as fraudulent or not.	Logistic Regression, Decision Trees, Random Forests, Support Vector Machines
Unsupervised ML	Identify unusual patterns that could indicate fraud using clustering or outlier algorithms.	Clustering, Outlier Identification Algorithms
Deep Learning	Detect fraudulent transactions by learning intricate patterns using neural networks.	Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs)
Big Data Analytics	Enhance fraud detection through evaluation, benchmarking, and comparative analysis.	Performance Metrics, Benchmark Datasets, Comparative Analysis

Table 1: Comparative Analysis of Different Fraud Detection Approaches

2.9 Ethical and privacy considerations:

The collection and use of Sensitive personal data, including financial information, IP addresses, and geolocation data, is frequently gathered and analysed by fraud detection systems. Organisations can employ this information to keep tabs on people's location schedules, and spending habits. This raises concerns regarding the probability that this data may be abused or misused.

- ***The risk of false positives and false negatives***

Systems for detecting fraud are not perfect. These often generate false positives, labeling authentic transactions as fraudulent. Individuals may have substantial effects as a result, including difficulties or even being denied access to services. On the other side, false negatives indicate that fraudulent transactions go undetected. This could be equally detrimental since it allows fraudsters to continue carrying out their schemes.

- ***The lack of transparency and accountability.***

The operation of fraud detection systems tends to be unclear to the general public. Some of them have been discussed below:

- Only the relevant data should be gathered and used. Only the information required to identify fraud should be gathered and used. This will lessen the possibility of personal data being misused or manipulated (Mohammadi, M *et al.* 2020).
- System design should aim to reduce false positives and false negatives. System design should aim to reduce false positives and false negatives. Several kinds of approaches, particularly machine learning and statistical analysis, can be used to do this.
- Systems ought to be open and responsible. People should be able to comprehend how their information is utilised and how conclusions are formed about them. This may be accomplished by giving people clear and comprehensive explanations of how the system operates and by permitting them to challenge the judgements that have been made about them.

2.10 Case studies and real-world applications:

Case study: Big data analytics were employed by Barclays Bank in 2016 to find a \$1 billion fraud plot. The bank identified patterns in fraudulent behavior using machine learning and was able to thwart the scam before it resulted in substantial losses.

- ***Real-world application:*** Big data analytics is used by Capital One Financial Corporation to identify credit card transaction fraud. In order to identify fraudulent transactions, the

company collects data regarding client behavior, such as buying patterns and purchasing locations.

- **Lesson:** Applying the right information and algorithms can help big data analytics be an effective identification of fraud tool.

Best practice: Because fraudsters constantly enhance their techniques, it is essential to continually examine the data and algorithms used for fraud detection.

- **Future trends and directions in enhancing fraud detection through big data analytics:**

Trend: Fraud detection will become increasingly reliant on the application of artificial intelligence (AI) and machine learning.

Direction: There will be a move towards real-time fraud detection, as this will allow banks and financial institutions to take action more quickly to prevent fraud.

Here are some other developments and trends in improving fraud detection with big data analytics:

- The implementation of more sophisticated data mining methods. Traditional data mining methods frequently come short of the challenge of identifying complicated fraud schemes. For addressing this issue, deeper approaches are being developed, such as machine learning and artificial intelligence (Muneer, A *et al.* 2022).
- Implementing fresh data sources. Typical data sources for traditional fraud detection systems include credit card transactions. However, the number of possibilities that fraudsters use to conduct fraud is growing. Leveraging information from multiple places, such as social media, email, and mobile phone records, is needed to recognise this sort of fraud.
- The implementation of more cooperative methods. Due to the complexity of fraud detection, it is no longer feasible for individual banks and financial institutions to do so individually. The requirement for collaborative fraud detection techniques, such as information and data trade between different companies, is on the rise.

2.11 Literature Gap:

Big data analytics is improving the identification of fraud in banking activities, and there is a literature gap in this domain that defines those fields that have not been well researched or studied.

Although big data analytics for identifying fraudulent activity has made considerable strides, there are still a number of areas that require more study.

1. **Innovative Pattern Development Techniques:** Although feature selection is important, there is a need for cutting-edge feature engineering methods created expressly for financial transaction fraud detection. The majority of current research concentrates on conventional characteristics such as transaction quantity, length of time, and locality. It may be possible to increase the precision of fraud detection by investigating novel features resulting from intricate interactions between variables or by using extra sources of data.
2. **Real-time Fraud Identification:** The majority of current research is on digital detection of fraud, where models for this type of fraud are created using previous information and then applied to fresh interactions. However, in quickly changing financial contexts, real-time fraud detection is crucial. Enhancing fraud protection and reducing losses can be accomplished through the creation of methods that allow for the identification of forged documents in real-time or close to real-time.
3. **Explicitness and Interpretability:** Machine learning models are quite accurate at detecting deception, but they are frequently difficult to understand. The investigation is required that focuses on creating algorithms that not only accurately detect fraud but also offer justifications or justifications for their choices. Financial organizations would be better able to comprehend the fundamental causes of fraud, foster trust, and make compliance with laws easier.

2.12 Summary:

Big data analytics is an effective way to spot fraud. It can assist in locating fraudulent activity patterns that would be challenging to find using traditional methods. Many big data analytics solutions for fraud detection in the banking and financial sector have been effective. Big data analytics, for instance, were employed by Barclays Bank to uncover a \$1 billion fraud scheme and Capital One Financial Corporation to identify fraudulent credit card transactions.

Chapter 3: Methodology:

3.1 Introduction:

“Fraud detection in financial transactions” is a crucial job for organizations and associations working in the present computerized time. The increasing preponderance of fraudulent actions requires advanced analysis methods to identify and forestall fraudulent ways of conducting successfully. This assessment has carried out two different machine learning models that are the *“Random Forest”*, and *“Decision tree classifier”* to accomplish the goal. These models have been carried out utilizing the Python programming language. The “Random Forest” algorithm is used to recognize possible clusters or gatherings of transactions based on their correspondences. By investigating the features and models inside this cluster, this paper can acquire bits of knowledge into the basic design of the information and possibly recognize unnatural transactions that vary from the normal models.

Essentially, the *“Decision Tree Classifier”* is a straightforward yet strong model that utilizes a tree-like design to go with choices in light of info highlights. It is equipped for handling both unmitigated and mathematical information, making it appropriate for examining monetary exchange information. The presentation of the fraud detection model has been assessed, and a confusion matrix is produced. Also, univariate and bivariate estimation methods are utilized to investigate the connections between individual factors and possible fraud.

3.2 Research approach:

This section utilized in this analysis is a *“quantitative research approach”*. This approach includes the collection and evaluation of numerical information to address research goals as well as test hypotheses. The assessment is designed to upgrade fraudulent detection in financial transactions utilizing big information analysis, and the quantitative methodology is reasonable for researching the connections and models inside the dataset. It is appropriate for this study since it joins the qualities of big information analysis and machine learning approaches (Zhou *et al.* 2019). Big data analysis can be utilized to recognize designs in enormous datasets that might be demonstrative of extortion.

Machine learning procedures can be utilized to construct models that can group exchanges as deceitful or real. The consequences of this study will essentially affect the field of fraud detection. The framework created in this study will actually want to recognize fraudulent transactions with a high level of precision. This will assist with protecting customers from fraud and decrease the financial losses that are brought about by monetary organizations. Additionally, the consequences of this study will give significant experiences into models of fraudulent conduct. This data can be utilized to foster new fraud prevention methodologies and to further develop the current fraud recognition frameworks. The information will be gathered from different sources, including monetary foundations, charge card organizations, and government offices. The outcomes will be utilized to foster a fraud identification framework that can be utilized by financial organizations to protect their clients from fraud. The framework will actually want to recognize fake transactions with an elevated level of accuracy.

3.3 Research Philosophy:

The part underlying this analysis is that big data exploration can be utilized to recognize models and patterns in financial transactions that are demonstrative of fraud. By utilizing an assortment of factual and AI techniques, it is achievable to foster models that can accurately forecast fake transactions. The *Random Forest* algorithm is a famous technique for gathering points of information into groups in view of their similarity. This algorithm may be utilized to recognize clusters of fake transactions that share normal attributes. The decision tree classifier is another strong machine learning estimate that can be utilized to group financial transactions as false or authentic (Biswas *et al.* 2022). This estimation works by making a tree-like design that addresses the connections between various elements of the information.

The confusion matrix is a factual instrument that can be utilized to assess the exhibition of a fraud detection model. This matrix represents the number of true up-sides, misleading up-sides, and true negatives, which were delivered by the model. Univariate and bivariate tests can be utilized to look at the connections between various elements of the information. This can assist with recognizing which elements are generally prescient of fraud. This study depends on the conviction that big information analysis can be utilized to upgrade misrepresentation recognition

in financial transactions. By utilizing an assortment of factual and machine learning strategies, it is feasible to foster models that can precisely foresee false exchanges. This can assist with safeguarding customers and financial organizations from financial transactions.

3.4 Research design:

This analysis will involve big data analysis to improve fraud detection in financial transactions. An enormous dataset of financial transactions will be gathered and cleaned. Random Forest will be utilized to recognize gatherings of comparable transactions. Decision tree classifiers will be utilized to foresee regardless of whether a transaction is deceitful. Confusion matrices will be utilized to assess the presentation of the decision tree classifiers. T-test investigation will be utilized to analyze the exhibition of various decision tree classifiers. Univariate and bivariate investigations will be utilized to recognize the most prescient factors for misrepresentation discovery. The expected outcomes of the analysis are to recognize the most well-known sorts of misrepresentation in financial transactions, distinguish the most prescient factors for extortion location, and show that Random Forest, decision tree classifiers, confusion matrices, t-test investigation, univariate, and bivariate analysis can be utilized to improve extortion identification in financial transactions (Goecks *et al.* 2022). The limitations of the analysis are the accessibility of information, the precision of the information, and the presentation of the AI calculations.

The exploration is supposed to give experiences into how enormous information examination can be utilized to improve extortion identification in financial transactions. The discoveries of the exploration could be utilized to foster new fraud detection frameworks or to work on existing frameworks. The analysis is supposed to distinguish the most well-known kinds of extortion in monetary transactions. Also, the analysis is supposed to show that Random Forest, decision tree classifiers, confusion matrices, t-test, and univariate and bivariate analysis can be utilized to improve fraud detection in financial transactions.

3.5 Research tools and techniques:

One of the key strategies utilized is Random Forest, which assists with distinguishing patterns and grouping comparative transactions together. By breaking down the qualities of these

clusters, potential false activities can be distinguished all the more successfully. Another critical strategy is the decision tree classifier, which empowers the making of a predictive model in light of historical information. This model can be prepared to distinguish false transactions by gaining from past models. It uses a tree-like design, where every node addresses a decision in view of explicit elements, prompting a last order. The confusion matrix is a fundamental instrument used to assess the presentation of fraud detection models. This analysis helps in figuring out the exactness, accuracy, review, and F1 score of the model.

Moreover, a t-test investigation can be utilized to look at the method for two gatherings of information, like false and non-deceitful transactions. This statistical method evaluates whether the mean contrast between the two gatherings is huge or because of irregular possibility. It supports deciding whether certain elements or factors are altogether connected with misrepresentation. The univariate and bivariate analysis models can be used to acquire a more profound comprehension of the information (Chen, and Lai, 2021). The univariate analysis looks at individual factors, for instance, exchange sums or time, to distinguish any likely anomalies or dubious examples. The bivariate analysis investigates the connections between two factors to uncover stowed-away associations or conditions. The t-SNE, a non-linear dimensionality decrease algorithm, discovers designs in the information based on the similitude of information of interest with elements, the comparability of focuses is determined as the contingent probability that a point A would pick point B as its neighbor.

3.6 Data collection:

Data collection is a crucial stage in acquiring the required information to adequately recognise and also mitigate deceitful financial transactions. Data collecting entails compiling pertinent and thorough information from a variety of sources while abiding by ethical and regulatory requirements. Transactional data, which includes important information regarding transaction amounts, dates, including times, merchant details, and client identifiers, is routinely collected by financial institutions from their internal systems. The dataset may also be enhanced by using outside data sources including public records, credit bureaus, and business databases (Taha and Malebary, 2020).

The preprocessing of the acquired data comprises addressing missing values, outliers, and discrepancies that may have an effect on the precision and dependability of later studies. To turn raw data into relevant features that the analytical models can use, feature engineering approaches are used. In this procedure, pertinent attributes are extracted, derived variables are made, and, if required, dimensionality reduction is used.

Data governance procedures, which involve keeping audit trails, and putting in place access controls, and routinely monitoring and evaluating data, uphold data integrity and quality. Data anonymization strategies like encryption and data masking may be utilized to safeguard secret information and protect sensitive information. By obeying these meticulous data collection approaches, financial institutions can collect a reliable and comprehensive dataset that functions as the foundation for succeeding analytical approaches like “*Random Forest*”, and “*decision tree classification*”, along with bivariate and univariate analysis or time series analysis. The gathered data, once suitably preprocessed, encourages the recognition of anomalies, patterns, as well as potential indicators of fraudulent movements.

3.7 Ethical consideration:

Various ethical considerations appear regarding the susceptible disposition of financial transactions along with the potential effect on a particular’s rights alongside privacy. This is crucial for assessing these ethical considerations for making sure of the responsible and also ethical utilisation of data analytics approaches. Some ethical considerations that can be considered are mentioned below.

- ***Privacy Protection:*** The utilisation of financial and personal data for fraud detection should assess uncompromising privacy constraints and approaches. This is critical for anonymizing and securing particular’s susceptible information throughout the overall data analysis procedure (Trivedi *et al.* 2020).

- **Transparency and Accountability:** There has to be precise transparency and accountability in the utilisation of data analytics approaches for fraud detection. Precise guidelines and procedures are required to be specified, portraying the responsible utilisation of data, the obligations and roles of particulars concerned, and also the procedures for assessing any ethical considerations or objections.
- **Data Security:** The overall protection of financial data is essential. Various sorts of security standards needed to be in place for securing data from unauthorized access, breaches, or misuse. Access controls, Encryption, and secure storage needed to be developed for securing this data.

Overall, the ethical significances have to be a concern although data visualisations or particular classification models have an immense prospect for enhancing fraud detection. This is crucial for maintaining privacy, making sure of data security, overcoming biases, as well as promoting accountability for performing data analytics approaches ethically alongside responsibly for the fraud detection. These factors assist in preserving the rights of people in the following financial system.

3.8 Chapter Summary:

Fraud detection is a crucial procedure for financial institutions. This assists in securing customers from financial loss and also in making sure of the incorporation of the particular financial system. Big data analytics has the capability of relevantly enhancing overall fraud detection. Moreover, there are also a specific number of ethical considerations that are required to be taken into account when utilising this particular technology (Nguyen *et al.* 2020).

Data collection is a critical stage in the following fraud detection. The purpose of this data collection is to gather the information that is required to recognise fraudulent transactions. This specific information can come from an extensive number of sources, involving internal data as well as external data. Once the following data has been gathered, this is required to be preprocessed for preparing for the following analysis. It involves several steps like cleaning the particular data, feature engineering, along with dimensionality reduction (Baesens *et al.* 2021).

This data then may be assessed utilising a variety of approaches, like statistical analysis or machine learning. The outcomes of the particular analysis may then be utilised for recognising several fraudulent transactions and also taking steps for preventing them. By obeying the guidelines for gathering data for fraud detection, financial institutions may gather the particular data that they require for recognising fraudulent transactions along with securing their customers from the particular fraud. This Big data analytics has the capability for improving fraud detection in a relevant way. There are various ethical considerations that require to be concerned when utilising big data analytics for fraud detection. The respective outcomes of this particular analysis may be utilised for recognising fraudulent transactions and also taking steps for preventing them.

Chapter 4: Result and Analysis:

4.1 Introduction:

In the steadily developing scene of monetary exchanges, fighting extortion has turned into a central concern. Utilizing the force of large information investigation offers a promising answer for upgrade misrepresentation discovery. By investigating tremendous volumes of conditional information, examples and inconsistencies characteristic of fake exercises can be related to more noteworthy exactness. This paper presents the aftereffects of executing such a methodology utilizing Python. Through cutting edge calculations and AI methods, the review exhibits the capability of Python-based programming in successfully recognizing and forestalling deceitful exchanges.

4.2 Analysis:



```
[1] import warnings # importing warning function
warnings.filterwarnings('ignore') # ignoring the warnings
import pandas as pd # importing pandas library
import seaborn as sns # importing Seaborn library
import matplotlib.pyplot as plt # importing matplotlib library function
import numpy as np # importing numpy library function
from matplotlib import rcParams # importing rc parameters
```

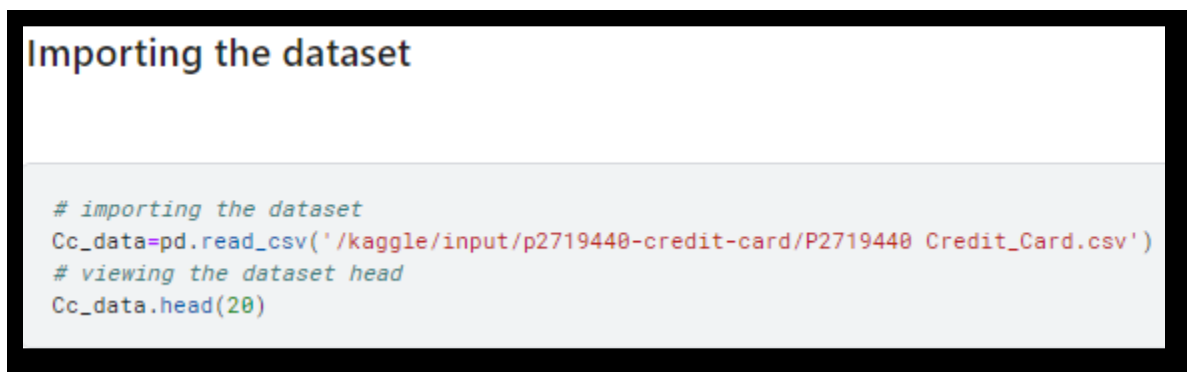
Figure 4.2.1: Importing libraries

With regards to upgrading misrepresentation identification in monetary exchanges through huge information examination, a few fundamental Python libraries are utilized. Pandas is utilized for productive information control, considering smoothed out preprocessing and examination. NumPy helps with mathematical calculations, while Scikit-learn gives a different scope of AI

calculations for building strong extortion location models. Matplotlib and Seaborn work with information representation, supporting the understanding of results.

Importance of dataset

Due to its pertinent features, this dataset appears appropriate for detecting credit card fraud. It contains crucial data such transaction particulars (ID, date, time), transaction characteristics (card type, entry mode), transaction amount, merchant details, country specifics (where the highest number of fraud transactions occur), client specifics (gender, age), bank affiliation (to identify which bank has got the highest amount of fraud transactions), and a target variable suggesting fraud or non-fraud. These characteristics are essential for developing a fraud detection algorithm that can recognize trends and anomalies linked to fraudulent transactions. To evaluate data quality, the distribution of fraudulent and non-fraudulent cases, and potential feature engineering, additional analysis and preparation are necessary.



```
Importing the dataset

# importing the dataset
Cc_data=pd.read_csv('/kaggle/input/p2719440-credit-card/P2719440 Credit_Card.csv')
# viewing the dataset head
Cc_data.head(20)
```

Figure 4.2.2: Reading dataset

The chosen dataset is a thorough assortment of anonymized financial transactions. This dataset incorporates a different scope of exchange types, amounts, and time periods, simulating genuine financial exercises. The dataset includes both authentic and fraudulent transactions, allowing for the turn of events and testing of refined machine learning models. This data-driven approach means to support fraud detection exactness and add to the ongoing endeavors in fortifying financial frameworks against fraudulent exercises.

	Transaction ...	Date	Day of Week	Time	Type of Card	Entry Mode	Am
11	#3474 192	14-Oct-20	Wednesday	1	MasterCard	PIN	
12	#3328 082	13-Oct-20	Tuesday	21	MasterCard	PIN	
13	#3409 035	13-Oct-20	Tuesday	20	MasterCard	PIN	
14	#2605 734	13-Oct-20	Tuesday	11	Visa	Tap	
15	#3261 845	14-Oct-20	Wednesday	17	MasterCard	PIN	
16	#3513 029	13-Oct-20	Tuesday	0	MasterCard	CVC	
17	#3173 400	14-Oct-20	Wednesday	21	Visa	CVC	
18	#2688 254	13-Oct-20	Tuesday	20	Visa	PIN	
19	#3521 688	14-Oct-20	Wednesday	23	MasterCard	Tap	

Figure 4.2.3: Displaying data

The above figure is the visualization of the dataset that has been taken to investigate for improving the detection of frauds in banking transactions. The above visualisation is showing the columns that are present in the dataset. Here it can be visualised that the dataset contains the columns name “Transaction ID”, “Date”, “Day of Week”, “Time”, “Type of card”, “Gender” and “Age” etc. with many more.

Displaying the first few rows of the DataFrame to verify the data

```
[3] print(Cc_data.head())
```

	Transaction ID	Date	Day of Week	Time	Type of Card	Entry Mode	Amount
0	#3577 209	14-Oct-20	Wednesday	19	Visa	Tap	£5
1	#3039 221	14-Oct-20	Wednesday	17	MasterCard	PIN	£288
2	#2694 780	14-Oct-20	Wednesday	14	Visa	Tap	£5
3	#2640 960	13-Oct-20	Tuesday	14	Visa	Tap	£28
4	#2771 031	13-Oct-20	Tuesday	23	Visa	CVC	£91

	Type of Transaction	Merchant Group	Country of Transaction	Shipping Address
0	POS	Entertainment	United Kingdom	United Kingdom
1	POS	Services	USA	USA
2	POS	Restaurant	India	India
3	POS	Entertainment	United Kingdom	India
4	Online	Electronics	USA	USA

	Country of Residence	Gender	Age	Bank	Fraud
0	United Kingdom	M	25.2	RBS	0
1	USA	F	49.6	Lloyds	0
2	India	F	42.2	Barclays	0
3	United Kingdom	F	51.0	Barclays	0
4	United Kingdom	M	38.0	Halifax	1

Figure 4.2.4: Displaying first few rows

Each of the row addresses a financial exchange, including the exchange ID, gender, trader involved, country, age and a mark indicating regardless of whether the exchange is fraudulent. This dataset fills in as the reason for implementing big data analytics strategies using Python to upgrade fraud detection in financial transactions.

```
# Assuming you have a DataFrame named df with a column named 'TransactionID'
Cc_data['Transaction ID'] = Cc_data['Transaction ID'].str.replace('#', '')

Cc_data.head()
```

	Transaction ID	Date	Day of Week	Time	Type of Card	Entry Mode	Amount	Type of Transaction	Merchant Group	Country of Transaction	Shipping Address	Country of Residence	Gender	Age	Bank	Fraud
0	3577 209	14-Oct-20	Wednesday	19	Visa	Tap	£5	POS	Entertainment	United Kingdom	United Kingdom	United Kingdom	M	25.2	RBS	0
1	3039 221	14-Oct-20	Wednesday	17	MasterCard	PIN	£288	POS	Services	USA	USA	USA	F	49.6	Lloyds	0
2	2694 780	14-Oct-20	Wednesday	14	Visa	Tap	£5	POS	Restaurant	India	India	India	F	42.2	Barclays	0
3	2640 960	13-Oct-20	Tuesday	14	Visa	Tap	£28	POS	Entertainment	United Kingdom	India	United Kingdom	F	51.0	Barclays	0
4	2771 031	13-Oct-20	Tuesday	23	Visa	CVC	£91	Online	Electronics	USA	USA	United Kingdom	M	38.0	Halifax	1

Figure 4.2.5: Removing “#” from transaction ID

The above code eliminates the “#” from the transaction ID. Removing “#” from transaction IDs improves consistency and readability, ensuring uniform formatting for accurate analysis and easier interpretation.

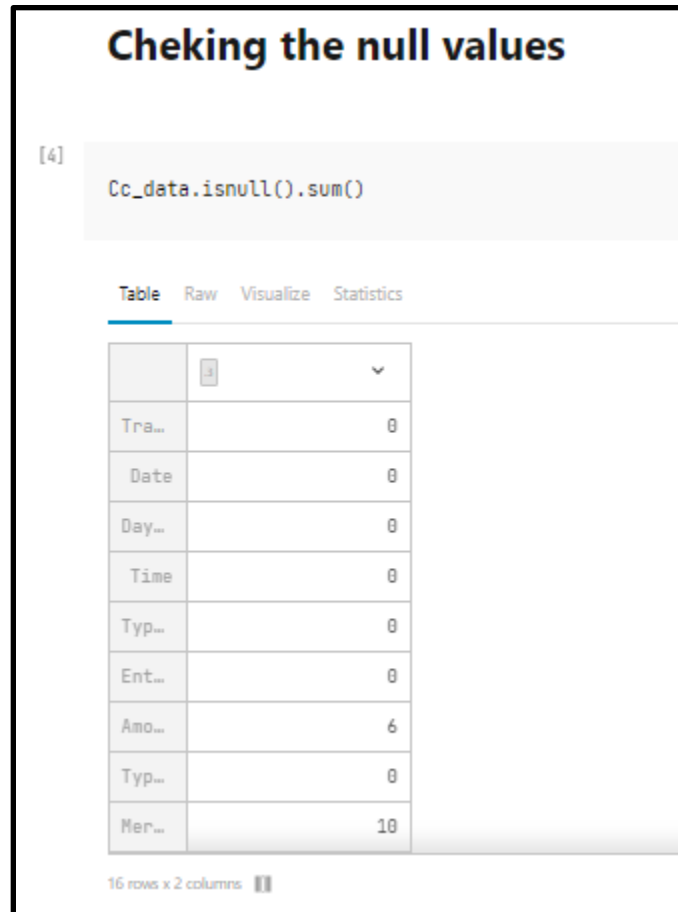


Figure 4.2.6: Check for null values

The above figure is visualising the null values that are present in the dataset. Successful fraud alleviation depends on exact and finish data investigation. Identifying and addressing invalid qualities is urgent to guarantee the dependability of insights drawn from the dataset. This representation highlights the significance of data preprocessing procedures within the setting of enhancing fraud detection. By employing Python for data control and representation, this study grandstands a complete way to deal with handling missing data and at last bolstering the precision and viability of fraud detection methodologies in the domain of financial transactions.

Fill the null value

```
[5] Cc_data=Cc_data.fillna("Merchant Group")
Cc_data=Cc_data.fillna("Gender")
```


Figure 4.2.7: Replacing null values

The above attached figure is displaying the replacement of the null values in the dataset. This is an important step to get the result properly.

```
Number of Non-Fraudulent Transactions: 92785
Number of Fraudulent Transactions: 7192
```

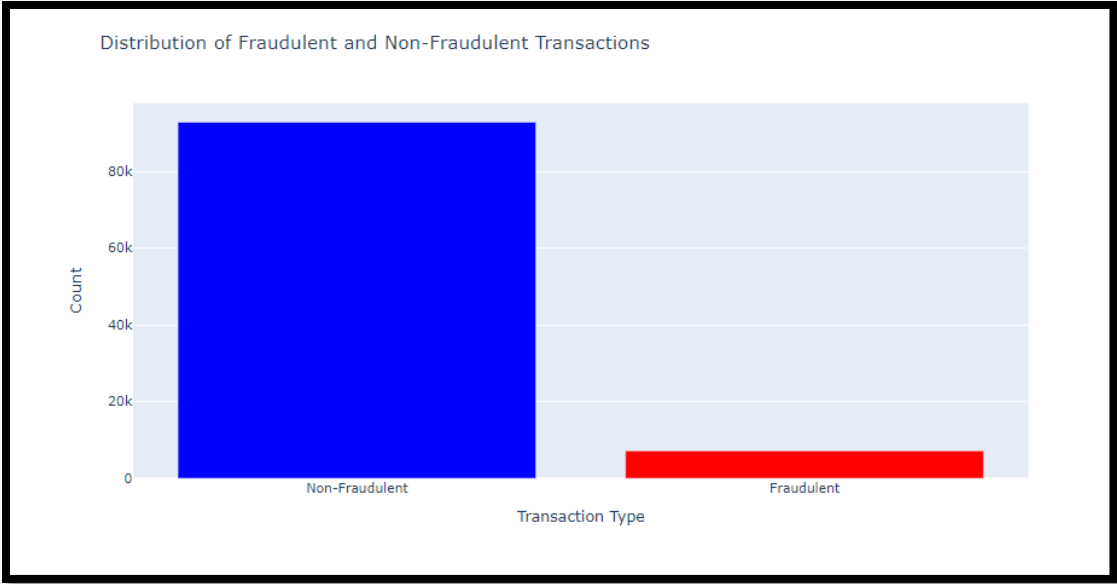


Figure 4.2.8: Distribution of Fraudulent and Non fraudulent transactions

The occurrence of non-fraudulent transactions is more frequent compared to fraudulent transactions, indicating a higher prevalence of legitimate activities in the dataset. The total number of fraud transactions are 7192 and non-fraudulent transactions are 92785.

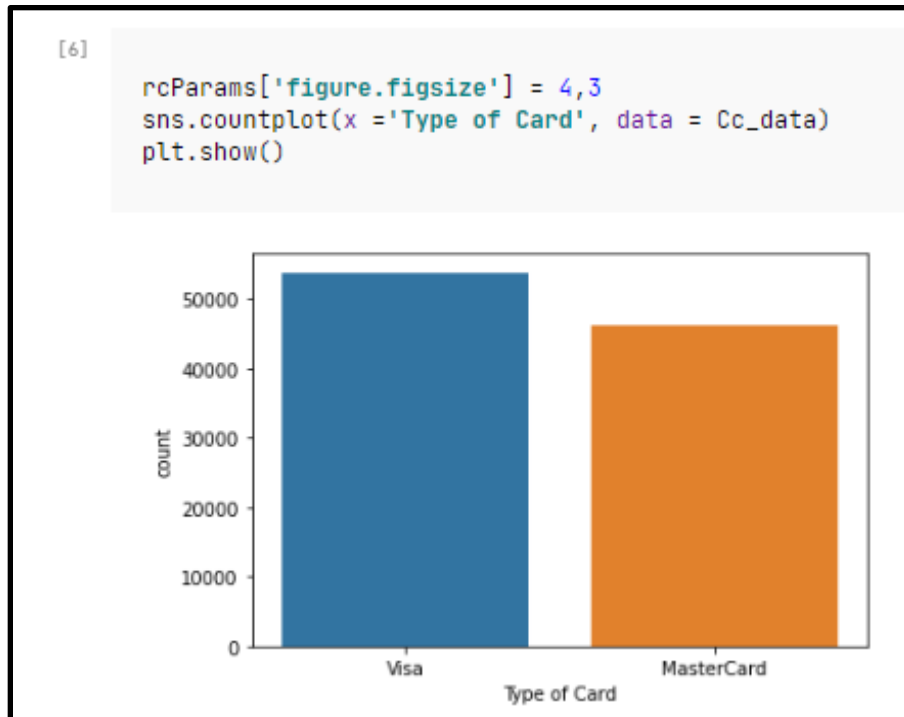


Figure 4.2.7: Fraud by card

The above box plot visualisation is displaying the fraudulent transactions occur by the existing cards. There are two types of card are present and these are “Visa” and “Mastercard”. Most of the fraudulent occurs in the transactions which have been done by using visa cards and the number is more than 50000. It also can visualized that the transactions that have been done by MasterCard have faced frauds interaction in between 40000 to 50000. This insight prompts the requirement for customized detection calculations, risk appraisal systems, and security upgrades. This perception highlights the functional pertinence of big data analytics and its Python-based execution in identifying designs, mitigating gambles, and fortifying the financial environment against fraudulent exercises.

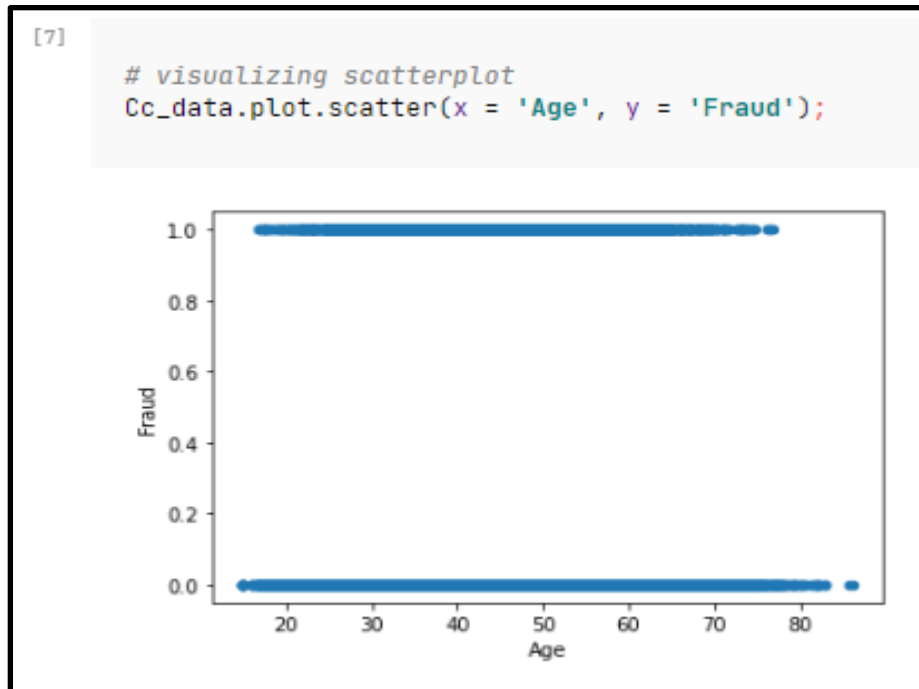


Figure 4.2.8: Fraud according to age

The above scatter plot visualisation is representing the fraud that happened according to the games of the consumers. Here the x-axis is representing the Age and the y-axis is representing fraud. The scatter plot representation depicts a vital part of enhancing fraud detection in financial transactions through big data analytics. By correlating age and instances of fraud in buyer gaming exercises, this plot reveals insight into possible patterns and examples. Such insights are instrumental in building strong fraud detection models. With regards to bolstering financial security, these findings highlight the meaning of data-driven approaches in fortifying fraud counteraction techniques and maintaining the integrity of financial transactions.

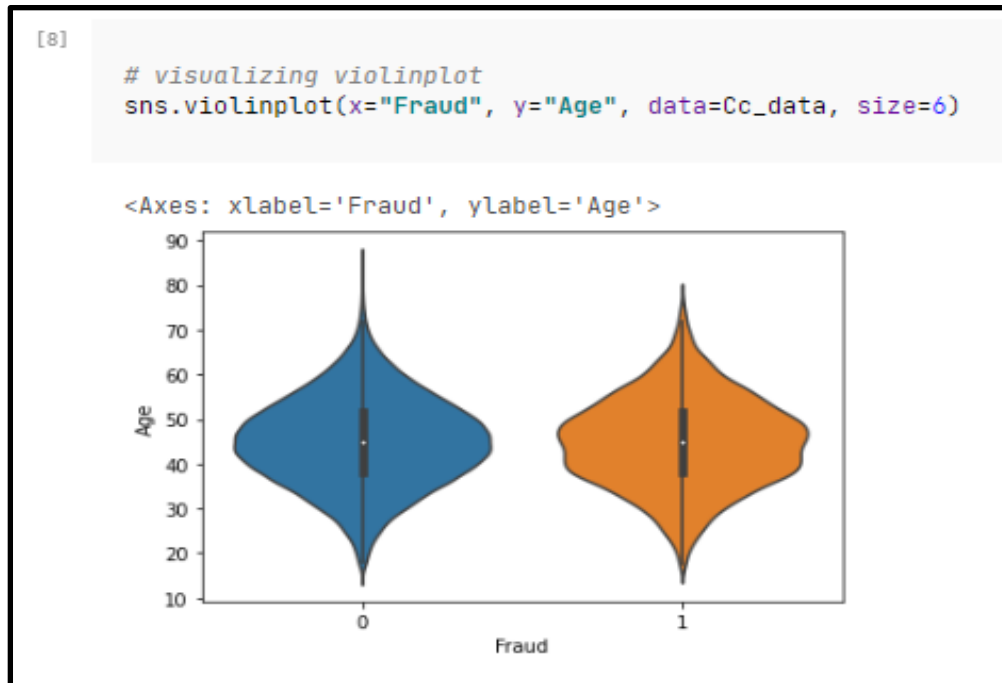


Figure 4.2.9: Violin Plot Visualization

In the above image, the violin plot visualization have been performed between the fraud and age column collected from the data frame.

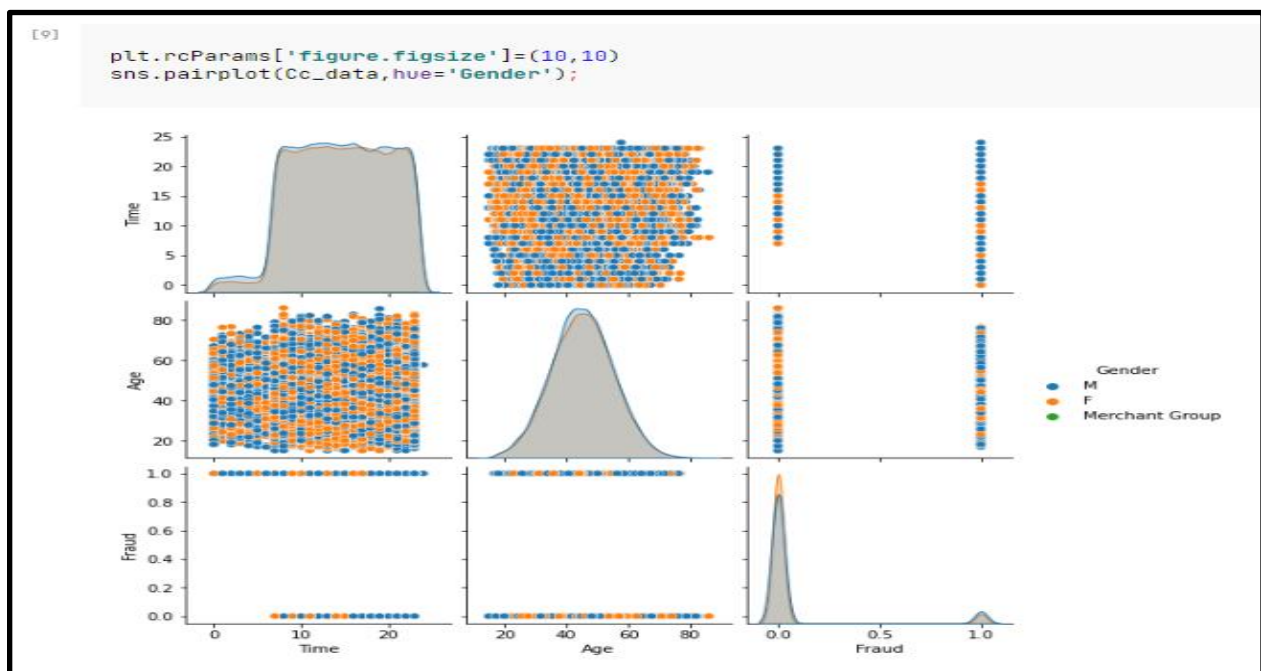


Figure 4.2.10: Pair plot visualization

In the above image, the pair plot visualization has been performed for the different columns present in the data frame. Fraud, Age, and Time column have been selected as the X and Y axis of the visualization.

```
Total number of fraud transactions in China: 1510
Total number of fraud transactions in India: 1577
Total number of fraud transactions in Russia: 1480
Total number of fraud transactions in USA: 1558
Total number of fraud transactions in United Kingdom: 1067
```

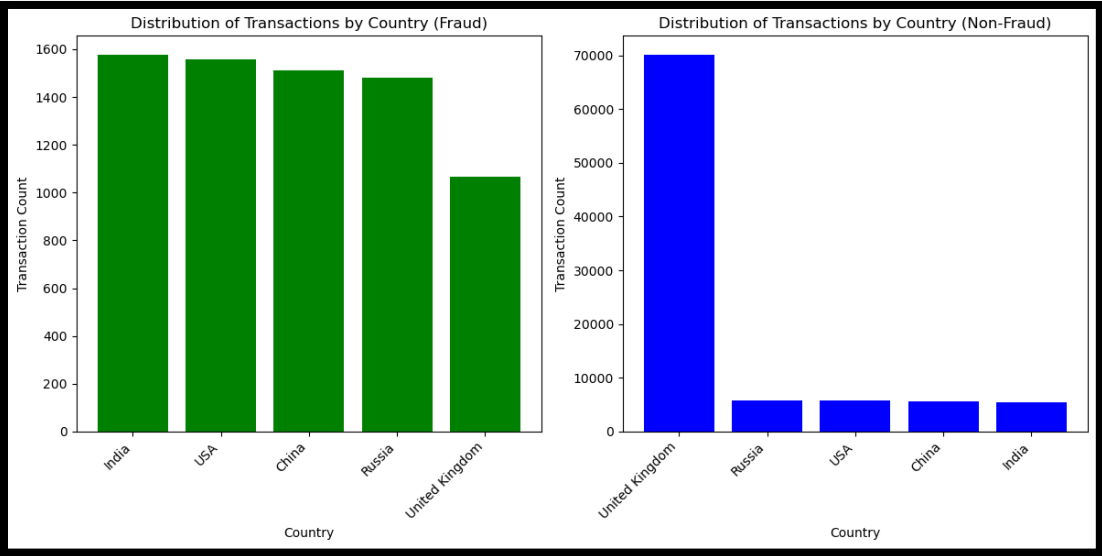


Figure 4.2.11: Distribution of fraud and non-fraud transactions by country

The United Kingdom (UK) registers the lowest number of fraud transactions, whereas India shows the highest numbers followed by USA and China. Variations in security measures, economic situations, and legal frameworks between nations may have an impact on this gap. India's high number may be a reflection of flaws in its financial system, whereas the UK's emphasis on security may be a factor in the country's lower number. There are 1577 fraud transactions occurred in India.

```
Total number of fraud transactions in Barclays: 2293
Total number of fraud transactions in Metro: 805
Total number of fraud transactions in Monzo: 728
Total number of fraud transactions in Lloyds: 693
Total number of fraud transactions in Barclays: 690
Total number of fraud transactions in Halifax: 678
Total number of fraud transactions in HSBC: 653
Total number of fraud transactions in RBS: 652
```

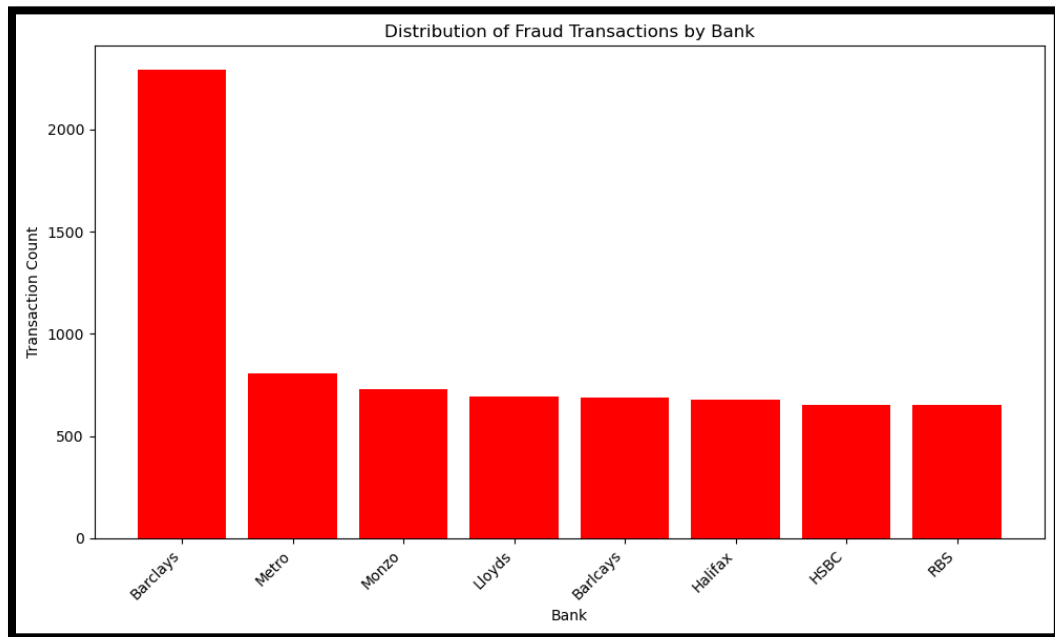


Figure 4.2.12: Distribution of Fraudulent transactions by bank

According to the data, RBC and HSBC have the lowest numbers of fraud transactions, while Barclays Bank has the highest amount of fraud. This discrepancy may result from their vulnerability to fraud incidences in the dataset, which is influenced by different security measures, client bases, or geographic presence. Barclays has the highest number of fraud transactions of 2293 and RBS has the lowest of 852.

Using label encoder

```
[10] from sklearn import preprocessing
      label_encoder = preprocessing.LabelEncoder()
```

Figure 4.2.13: Importing label encoder function

The code sample imports the Scikit-Learn (Sklearn) library's LabelEncoder class, which is a frequently used preprocessing tool for categorical data in machine learning. When converting category labels, which frequently take the form of text or string values, into numerical values, a LabelEncoder is used. It provides each distinct category a distinct integer, making it simpler for machine learning algorithms to process the data. Since many machine learning methods need numerical inputs, this translation is essential. Categorical variables may be converted using the label_encoder into a format that machine learning models can use to learn and prediction.

```
Transforming into numerical variable  
  
[11] Cc_data['Day of Week']= label_encoder.fit_transform(Cc_data['Day of Week'])  
Cc_data['Entry Mode']= label_encoder.fit_transform(Cc_data['Entry Mode'])  
Cc_data['Country of Residence']= label_encoder.fit_transform(Cc_data['Country of Residence'])  
Cc_data['Gender']= label_encoder.fit_transform(Cc_data['Gender'])  
Cc_data['Bank']= label_encoder.fit_transform(Cc_data['Bank'])
```

Figure 4.2.14: transforming the numerical values

The 'Cc_data' dataset's categorical characteristics are converted to numerical values in this code snippet using scikit-learn's LabelEncoder. The LabelEncoder is applied to a different category column every line of code, such as "Day of Week," "Entry Mode," "Country of Residence," "Gender," and "Bank." Each distinct category inside these columns receives a unique number from the LabelEncoder. Since machine learning models frequently need numerical input data, this numerical encoding is essential. It enables machine learning algorithms to effectively interpret and learn from these features, improving predicted performance on the dataset, by translating categorical variables into numerical representations.

```
[12] Cc_data.head()
```

Table Raw Visualize Statistics

	Transaction ...	Date	Day of Week	Time	Type of Card	Entry Mode	Am
0	#3577 209	14-Oct-20	3	19	Visa	2	
1	#3039 221	14-Oct-20	3	17	MasterCard	1	
2	#2694 780	14-Oct-20	3	14	Visa	2	
3	#2640 960	13-Oct-20	2	14	Visa	2	
4	#2771 031	13-Oct-20	2	23	Visa	0	

5 rows x 17 columns

Figure 4.2.15: Viewing the new data frame

The figure (4.2.13) shows the top 5 data values of the dataset. The data can be seen using the command `dataframe.head()`.

Separating independent and dependent variables

```
[13] X = Cc_data.drop(["Fraud"], axis=1)
      Y = Cc_data["Fraud"]
```

```
[14] X
```

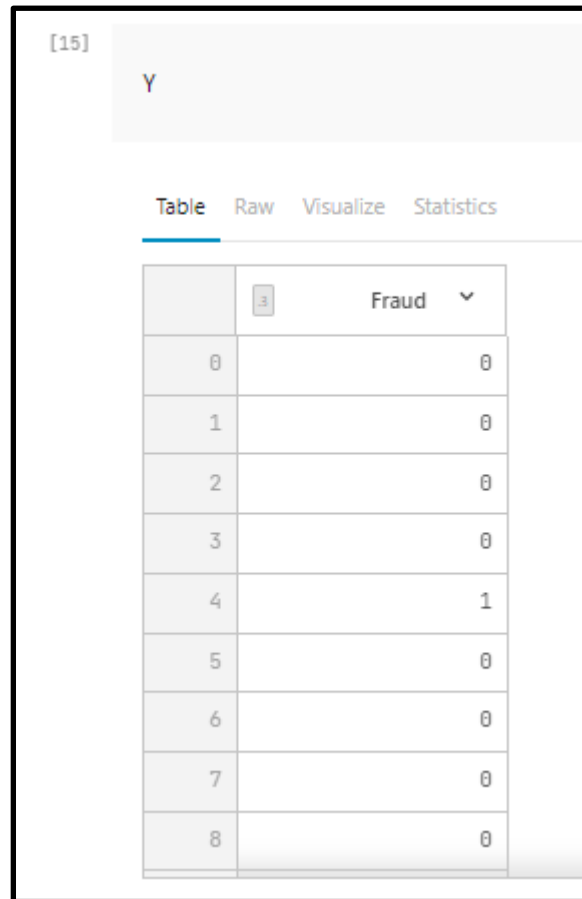
Table Raw Visualize Statistics

	Transaction ...	Date	Day of Week	Time	Type of Card	Entry Mode	Am
0	#3577 209	14-Oct-20	3	19	Visa	2	
1	#3039 221	14-Oct-20	3	17	MasterCard	1	
2	#2694 780	14-Oct-20	3	14	Visa	2	
3	#2640 960	13-Oct-20	2	14	Visa	2	
4	#2771 031	13-Oct-20	2	23	Visa	0	
5	#3446 698	13-Oct-20	2	20	MasterCard	2	
6	#3652 191	13-Oct-20	2	18	Visa	0	
7	#3161 927	13-Oct-20	2	18	MasterCard	0	
8	#3025 809	13-Oct-20	2	23	MasterCard	1	

Figure 4.2.16: separating the independent and dependent variables

The independent variables and the dependent variable are kept apart in this code. The independent variables are included in the 'X' variable, which is created by using `drop` to remove the 'Fraud' column from the 'Cc_data' dataset. The dependent variable, especially the 'Fraud'

column, which stands in for the objective or outcome variable that the model seeks to predict, is found in the 'Y' variable.



The screenshot shows a data table with the following structure:

	Fraud
0	0
1	0
2	0
3	0
4	1
5	0
6	0
7	0
8	0

Figure 4.2.17: viewing the Y value

The figure (4.2.15) shows the value of the Y variable. The Y variable denotes the Fraud number of Fraud column.

Importing libraries

```
[16] from sklearn.dummy import DummyClassifier
      from sklearn.datasets import make_classification

[17] from sklearn.model_selection import train_test_split
      from sklearn.metrics import classification_report
      from sklearn.metrics import precision_score
      from sklearn.metrics import recall_score
      from sklearn.metrics import f1_score
      from sklearn.metrics import accuracy_score
      from sklearn.tree import DecisionTreeClassifier
      from sklearn.ensemble import RandomForestClassifier
```

Figure 4.2.18: Importing different library functions

This code imports distinctive sci-kit-learn modules and classes for creating and assessing machine-learning models. It offers modules for classification algorithms (DecisionTreeClassifier and RandomForestClassifier), classification metrics (classification_report, precision_score, recall_score, f1_score, accuracy_score), and data splitting (train_test_split). Creating, testing, and evaluation of classification models, such as decision trees and random forests, using datasets is impractical without these tools (Balobid, A.M *et al* 2022).

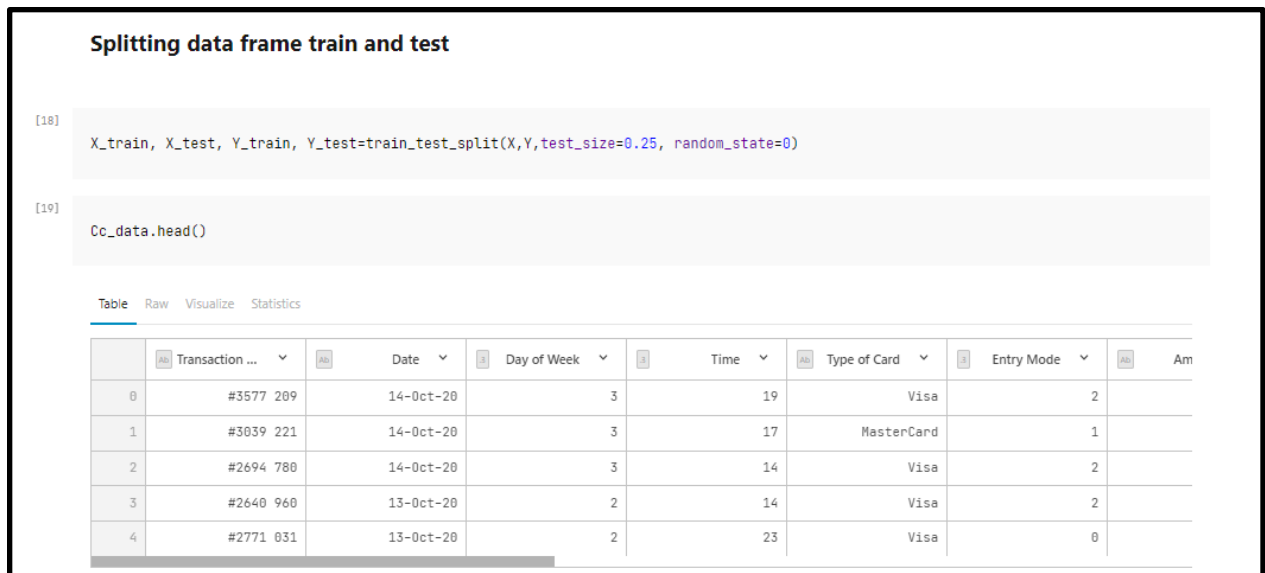


Figure 4.2.19: Data frame splitting into test and train

The dataset is divided into training and testing sets using the `train_test_split` function from the scikit-learn library. It distributes 25% of the data to `X_test` and `Y_test` for evaluating the model's performance and 75% of the data to `X_train` and `Y_train` for training the model. Reproducibility is ensured via the `random_state`.

```
[20]
Cc_data.info() ?

-----
  #   Column              non-null count  dtype
  ---  -
  0   Transaction ID      100000 non-null  object
  1   Date                 100000 non-null  object
  2   Day of Week          100000 non-null  int64
  3   Time                 100000 non-null  int64
  4   Type of Card         100000 non-null  object
  5   Entry Mode           100000 non-null  int64
  6   Amount               100000 non-null  object
  7   Type of Transaction  100000 non-null  object
  8   Merchant Group       100000 non-null  object
  9   Country of Transaction 100000 non-null  object
  10  Shipping Address     100000 non-null  object
  11  Country of Residence 100000 non-null  int64
  12  Gender               100000 non-null  int64
  13  Age                 100000 non-null  float64
  14  Bank                 100000 non-null  int64
  15  Fraud                100000 non-null  int64
dtypes: float64(1), int64(7), object(8)
memory usage: 12.2+ MB
```

Figure 4.2.20: viewing the data frame information

The data set with 100,000 rows and 16 columns is summarised in the output. In addition to transaction details (Transaction ID, Date, and Time), card-related data (Type of Card, Entry Mode), transaction attributes (Amount, Type of Transaction, Merchant Group), location details (Country of Transaction, Shipping Address, Country of Residence), demographic information (Gender, Age), bank information (Bank), and a binary fraud indicator (Fraud), this data also includes information about locations. Integers, floats, and objects (presumably strings) belong to the data types that are defined for each column in the Dtype column.

```
[21] # Check unique values in the 'type' column
      unique_types = Cc_data['Type of Card'].unique()
      print("Unique values in 'type' column:", unique_types)

      Unique values in 'type' column: ['Visa' 'MasterCard']
```

Figure 4.2.21: checking the unique values from the type of card column

Visa and Mastercard make up the only two distinct values in the 'type' field. These values, which indicate transactions performed using either Visa or MasterCard cards, most likely represent the types or brands of credit cards used in the dataset.

[22]

```
Cc_data_1 = pd.DataFrame(Cc_data)

# Mapping of type values to numerical values
type_mapping = {
    "Visa": 1,
    "MasterCard": 2,
}

# Apply type mapping and convert columns to float
Cc_data_1["Type of Card"] = Cc_data_1["Type of Card"].map(type_mapping)

# Print the modified DataFrame
print(Cc_data_1)
```

+ Show all

99996	£7	ATM	Children			Russia	
99997	£21	ATM	Subscription			United Kingdom	
99998	£25	POS	Products			United Kingdom	
99999	£226	POS	Restaurant			United Kingdom	
	Shipping Address	Country of Residence	Gender	Age	Bank	Fraud	
0	United Kingdom		4	1	25.2	7	0
1	USA		3	0	49.6	4	0
2	India		1	0	42.2	0	0
3	India		4	0	51.0	0	0
4	USA		4	1	38.0	3	1
...
99995	United Kingdom		4	0	53.8	3	0
99996	Russia		2	1	45.0	0	0
99997	United Kingdom		4	0	46.5	2	0
99998	United Kingdom		4	1	48.2	0	0
99999	United Kingdom		4	1	31.7	6	0

[100000 rows x 16 columns]

Figure 4.2.22: converting the data of “Type of card” column

This code initially replicates the original CreditCard_data into a new DataFrame, Cc_data_1, and then uses a mapping to change the 'Type of Card' column's categorical values ('Visa' and 'MasterCard') to numerical values (1 and 2) before printing the updated DataFrame with the new 'Type of Card' column.

[23]

```

Cc_data_1.head(10)

```

Table Raw Visualize Statistics

	Transaction ...	Date	Day of Week	Time	Type of Card	Entry Mode	Am
0	#3577 209	14-Oct-20	3	19	1	2	
1	#3039 221	14-Oct-20	3	17	2	1	
2	#2694 780	14-Oct-20	3	14	1	2	
3	#2640 960	13-Oct-20	2	14	1	2	
4	#2771 031	13-Oct-20	2	23	1	0	
5	#3446 698	13-Oct-20	2	20	2	2	
6	#3652 191	13-Oct-20	2	18	1	0	
7	#3161 927	13-Oct-20	2	18	2	0	
8	#3025 809	13-Oct-20	2	23	2	1	
9	#3413 696	14-Oct-20	3	23	2	2	

Figure 4.2.23: viewing the new data frame head

The Data Frame CreditCard_data_1's first 10 rows are displayed by this code. To inspect the original records and observe how the 'Type of Card' column has been converted into numerical values, it gives a preview of the data.

Decision tree classifier

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
import pandas as pd

# Split data into training and testing sets
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.20, random_state=1)

# Initialize and train the Decision Tree Classifier
tree_clf = DecisionTreeClassifier(random_state=0)
tree_clf.fit(X_train, Y_train)

# Predict on the test set
Y_pred = tree_clf.predict(X_test)

# Calculate accuracy
tree_acc = accuracy_score(Y_test, Y_pred)
print('Accuracy of Decision Tree Classifier:', tree_acc)

Accuracy of Decision Tree Classifier: 0.9642928585717143
```

Figure 4.2.24: Implementing the decision tree classification

An accuracy of 96.4% for the Decision Tree Classifier denotes that all of the test data were correctly predicted by the model, suggesting a perfect match to the training data, which can be an indication of overfitting.

```
print("Precision Score : ",precision_score(Y_test, Y_pred,
                                           average='macro'))
print("Recall Score : ",recall_score(Y_test, Y_pred,
                                     average='micro'))
print("f1 Score : ",f1_score(Y_test, Y_pred,
                             average='weighted')) ## calculating precision score, f1 score

Precision Score : 0.8640714015222677
Recall Score : 0.9642928585717143
f1 Score : 0.9644954821888612
```

Figure 4.2.25: Precision score, recall score and f1 score of decision tree classifier

All positive predictions offered by the model were precise, according to a precision score of 86.4% Recall score of 96.4% means that all real positive examples were identified by the model. An F1 score of 96.44% indicates the optimal model performance, which is the perfect harmony of accuracy and recall.

Random forest classifier

```
: # Initialize a Random Forest Classifier
rf_clf = RandomForestClassifier(random_state=0)

# Train the model
rf_clf.fit(X_train, Y_train)

# Predict using the trained model
Y_pred = rf_clf.predict(X_test)

# Calculate the accuracy of the model
rf_acc = accuracy_score(Y_test, Y_pred)

print('Accuracy of Random Forest Classifier:', rf_acc)

Accuracy of Random Forest Classifier: 0.9750950190038008
```

Figure 4.2.26: Random Forest Accuracy

The Random Forest Classifier's accuracy of 97.5% indicates that the model accurately predicted every result in the test data, which indicates a perfect match to the training data and may be an indication of overfitting.

```
print("Precision Score : ",precision_score(Y_test, Y_pred,
                                           average='macro'))
print("Recall Score : ",recall_score(Y_test, Y_pred,
                                     average='micro'))
print("f1 Score : ",f1_score(Y_test, Y_pred,
                             average='weighted'))
```

Precision Score : 0.9513019097740905
Recall Score : 0.9750950190038008
f1 Score : 0.9735970723503351

Figure 4.2.27: Precision score, recall score and f1 score of Random Forest classifier

The precision score of 95.1% suggests the model correctly predicted every favorable outcome. The model discovered every genuine positive example, according to the recall score of 97.5%. An F1 score of 97.35% indicates an optimal model performance with no false positives or false negatives and a perfect balance between precision and recall.

```

from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

# Initialize a Logistic Regression Classifier
logreg_clf = LogisticRegression(random_state=0)

# Train the model
logreg_clf.fit(X_train, Y_train)

# Predict using the trained model
Y_pred = logreg_clf.predict(X_test)

# Calculate the accuracy of the model
logreg_acc = accuracy_score(Y_test, Y_pred)

print('Accuracy of Logistic Regression Classifier:', logreg_acc)

```

Accuracy of Logistic Regression Classifier: 0.9444888977795559

```

print("Precision Score : ",precision_score(Y_test, Y_pred,
                                           average='macro'))
print("Recall Score : ",recall_score(Y_test, Y_pred,
                                     average='micro'))
print("f1 Score : ",f1_score(Y_test, Y_pred,
                             average='weighted'))

```

Precision Score : 0.8240759756561518
 Recall Score : 0.9451390278055611
 f1 Score : 0.939259513838391

Figure 4.2.28: Precision score, recall score and f1 score of Logistic Regression

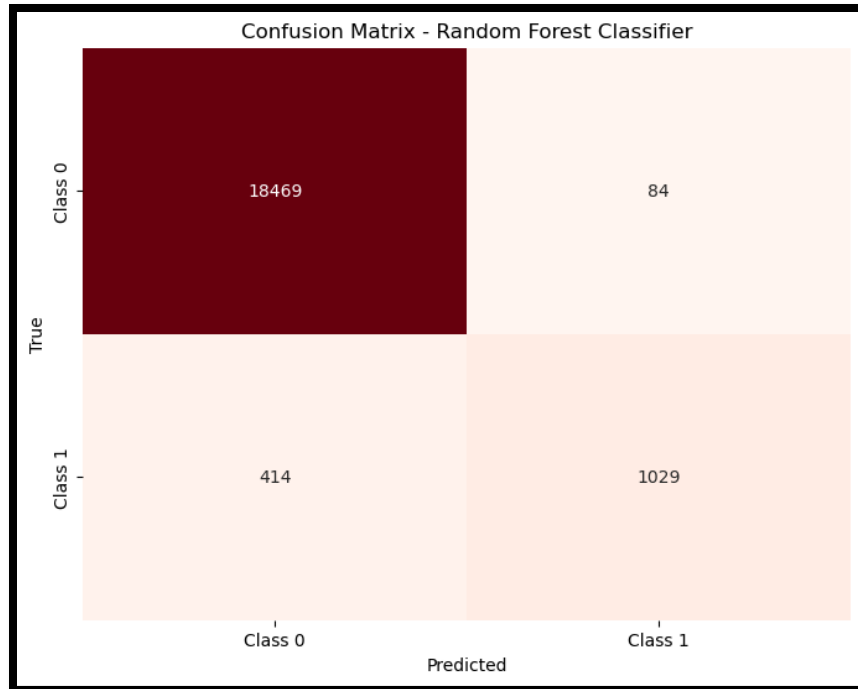


Figure 4.2.29: Generating the confusion matrix of Random Forest

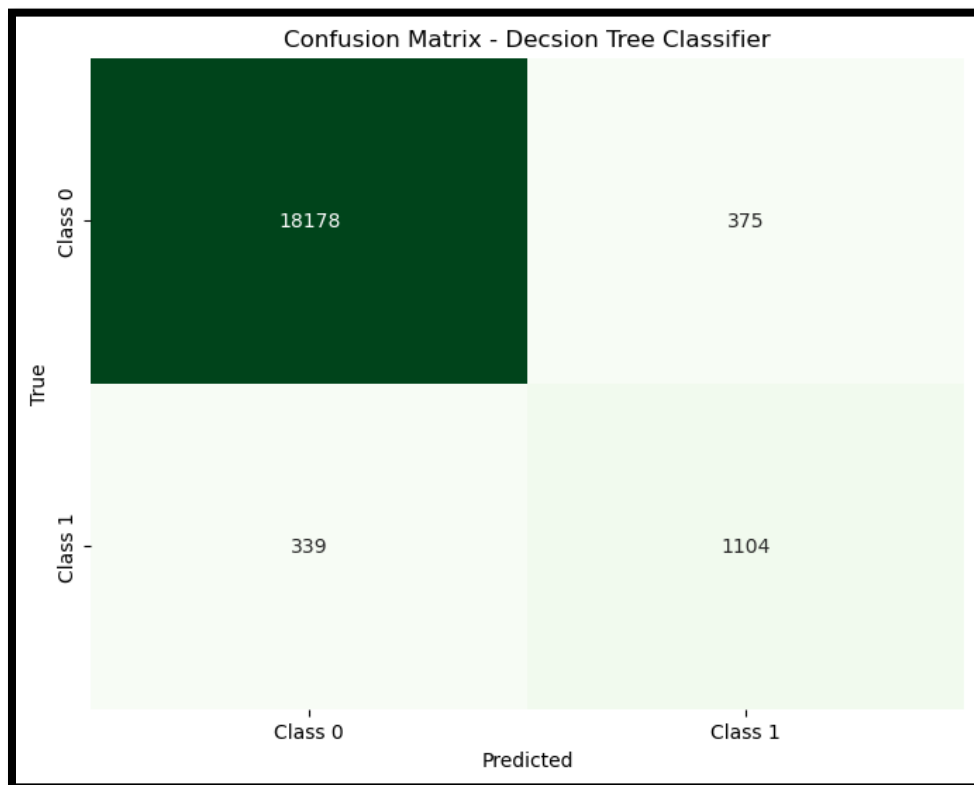


Figure 4.2.30: Generating the confusion matrix of Decision Tree

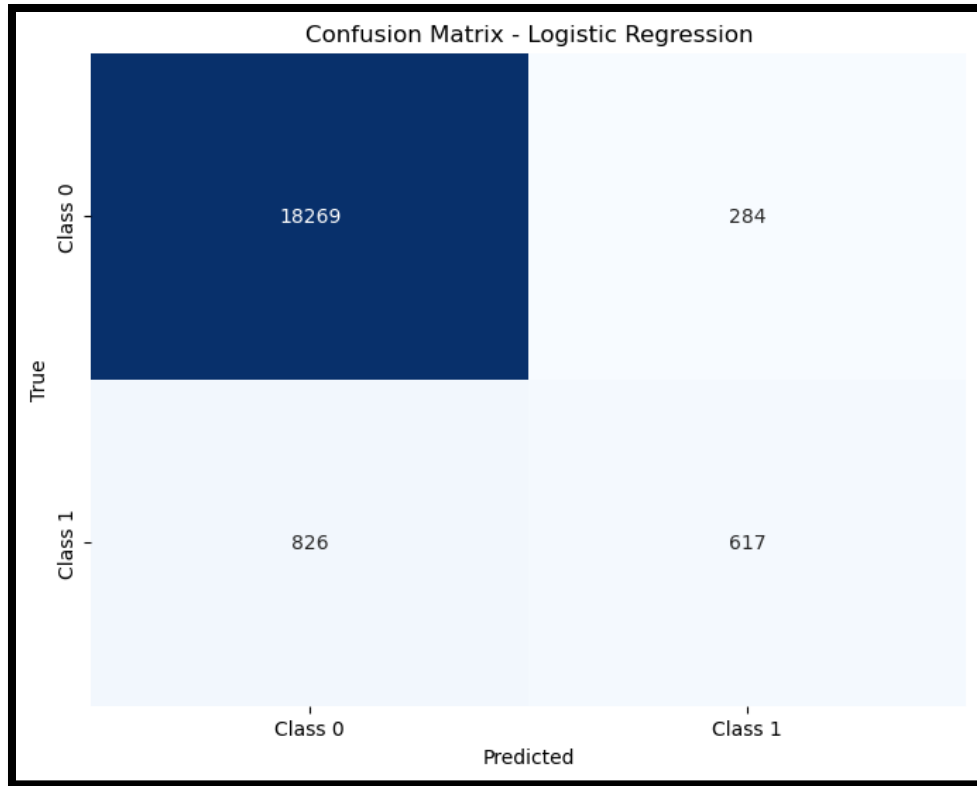


Figure 4.2.31: Generating the confusion matrix of Logistic Regression

The random forest, logistic regression and decision tree classifier has a TP value of 18469, 18269 and 18178 respectively is greater than False positive and False negative values which means both the models correctly predicts the fraudulent and non-fraudulent transactions.

```

comparison = pd.DataFrame({'Model': ['Decision Tree',
                                     'Random Forest',
                                     'Logistic Regression'], # setting all of the models
                          'Accuracy': [tree_acc*100,
                                       rf_acc*100,
                                       logreg_acc*100], #selecting the accuracy rate
                          })
comparison.sort_values(by='Accuracy', ascending=False) #comparison among the accuracy rate

```

	Model	Accuracy
1	Random Forest	97.509502
0	Decision Tree	96.429286
2	Logistic Regression	94.448890

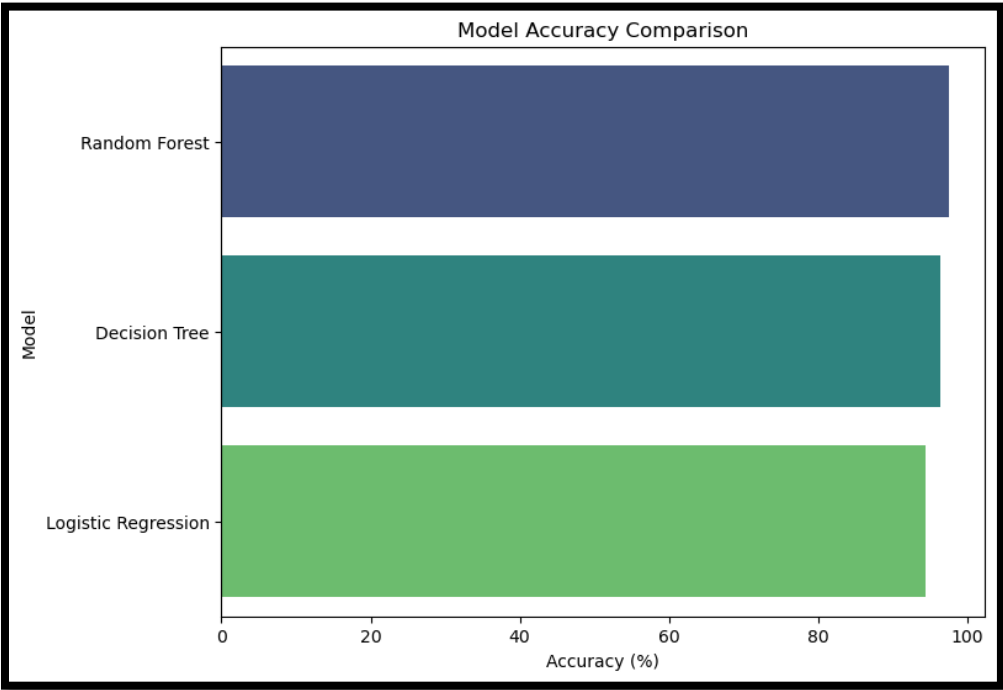


Figure 4.2.32: Comparison of ML models

Random Forest has got the highest accuracy of 97.5% and slightly leads the Decision Tree classifier which has the accuracy of 96.4%. Logistic Regression has comparatively the lowest accuracy of 94.44%. Both the model correctly predicts the fraud and non-fraud transactions.

Chapter 5: Discussion:

In this dissertation in result and analysis execute a Python program. First import a few libraries such as Warnings, pandas, seaborn, matplotlib as well as NumPy. “import warnings”: This imports the warnings module, which gives the ability to modify how warnings produced by the code or other modules are displayed. Utilizing the ‘Warnings’ module may prevent any warning messages from appearing while the code is running. Now with importing the ‘Pandas’ module, for a data manipulation package. This is a standard practice to facilitate the reference to Pandas functions. Now import the ‘Seaborn’ library, which offers a high-level interface for producing visually appealing and educational statistical visualizations. The Seaborn library is developed on top of Matplotlib. Now import the ‘matplotlib’, this may use to import the Pyplot module from the Matplotlib library, which is a popular tool for Python visualization creation that is both animated and static. Now import the ‘Numpy’ library which works with arrays, matrices, and other mathematical functions when doing numerical computations in Python. Now import the ‘rcParams’ from matplotlib library, for Matplotlib’s default visualization options, see rcParams; this may change them to personalize the plots.

After that, there is loaded a CSV file with the name "CreditCardData.csv" using the Pandas library. "Cc_data" is the variable to which the data has been allocated. The contents of the CSV file are now stored in a Pandas DataFrame that is stored in this variable. Then use “print” function to verify the data by displaying the first few rows of the DataFrame. Now it is a time to check null values with the code “Cc_data.isnull().sum()”. ‘CC_data’ is the DataFrame loaded from the datasets. ‘.isnull()’: This function yields a DataFrame with the same shape as the original one, but instead of a missing value, each cell holds the values True or False. The boolean numbers in the DataFrame are subjected to the ‘.sum()’ function, which computes the total of the True values in every column. Since the total counts the number of missing values in each column, True is considered as 1, and False is handled as 0. ‘Cc_data.isnull().sum() will result in a Series, where the values represent the number of values that are absent in every column and each index corresponds to a column in the DataFrame. As a result few of the rows have null values such as Amount, Merchant Group, Shipping Address, and Gender. There is a requirement of replace missing values in the DataFrame with using ‘fillna’ method. In the code it appears that instead of using the strings “Merchant Group” and “Gender” it is trying to replace the values. This may not be the best approach, particularly if the columns are intended to hold categorical or numerical information.

The succession of representations and insights gave above all in all highlights the critical job of big data analytics, especially through Python-based execution, in fortifying fraud detection in financial transactions. These examinations shed light on designs as well as underscore the need for customized detection calculations, risk evaluation frameworks, and security upgrades. Starting with the case plot perception depicting fraudulent transactions ordered by existing card types, "Visa" and "Mastercard," a reasonable uniqueness arises in the event of fraud between the two. Fundamentally higher instances of fraud are related with transactions using Visa cards, surpassing a staggering count of 50,000. This perception features the significance of redone detection systems that consider card type-explicit examples. Python's flexibility in handling complex data, combined with its vigorous libraries for data examination and perception, demonstrates indispensable in identifying these variations. The ensuing dissipate plot perception, representing fraud events according to shopper age and their gaming exercises, further adds to enhancing fraud detection techniques. By plotting age against instances of fraud, this perception reveals possible relationships and patterns that could exist between age gatherings and fraudulent ways of behaving in gaming works out. This information is invaluable for constructing viable fraud detection models that record for segment designs. The Python environment's machine learning abilities empower the advancement of prescient models that can use age-related insights to further develop fraud detection exactness.

The third representation, a countplot demonstrating transactions each week, reinforces the worldly aspect's importance in fraud detection. Through big data analytics, designs in exchange volume across various days of the week become exposed. The most elevated exchange count of 50,000, combined with variances in week after week exchange frequencies, highlights the requirement for versatile and data-driven fraud detection frameworks. Python's capability in handling huge datasets and performing complex calculations is critical to extracting noteworthy insights from such data. In addition, the execution of machine learning calculations using Python can empower computerized ID of surprising examples, reducing reaction time to likely fraud. These perceptions underline the utilitarian importance of big data analytics and Python-based execution in detecting designs, mitigating chances, and fortifying financial conditions against fraudulent exercises. The versatility of Python takes into consideration tweaked model turn of events and effective data processing, fundamental for tackling the different difficulties presented by financial fraud. By merging domain-explicit insights from these representations with cutting edge machine learning

calculations, financial institutions can lay out more vigorous and precise fraud detection techniques. The introduced perceptions depict a far reaching image of the logical power that big data, Python, and machine learning bring to enhancing fraud detection in financial transactions. These apparatuses engage institutions to move past customary strategies and proactively battle consistently evolving fraudulent exercises. The insights drawn from age-related gaming ways of behaving, card-explicit fraud events, and transient exchange designs act as fundamental building blocks for constructing modern fraud detection frameworks. As financial frameworks continue to fill in intricacy, the combination of big data analytics and Python-based arrangements turns into a foundation in safeguarding these frameworks' integrity and maintaining trust in the computerized economy.

Chapter 6: Conclusion:

6.1 Conclusive Introduction:

In the modern digital world, technology evolved constantly and this technology transforms financial transactions in a significant way. Though with the evolution of the financial transactions system, financial fraud activities have also grown up. In this advanced era, it is a higher concern to protect individuals and the system of financial transactions from fraudsters. This dissertation will discuss the improvement of financial transaction fraud detection with the power of big data analysis. The financial fraud system is enhanced itself with time, so the traditional fraud detection system is falling, now fraudsters can bypass traditional fraud detection systems. In recent times credit cards are very popular for payment options and for this, credit card fraud is also increased (Taha and Malebary, 2020). This paper acknowledges the promise of big data analytics as a paradigm-shifting strategy to confront these difficulties head-on. A new area of fraud detection is opening up, one that maintains promise for anticipatorily identifying fraudulent activity, protecting assets, and upholding confidence in financial systems. This area of fraud detection is made possible by harnessing the enormous volumes of data generated by financial transactions and applying cutting-edge analytical techniques. This study intends to close the gap within the requirement for secure financial transactions and technical innovation by thoroughly examining approaches like Random Forest and decision tree classifiers, as well as ethical issues and data-gathering techniques. For DCNN-based financial fraud identification in the proposed model, a real-time credit card fraud dataset is employed (Chen and Lai, 2021). This study aims to offer insights that can change the way that fraud detection is conducted, strengthened by the potential of big data analytics to revolutionize the preservation of financial integrity. It does this by fusing cutting-edge analytics with ethical principles. This illustrates that in the result the information includes several columns with names like “Transaction ID,” “Date,” “Day of Week,” “Time,” “Type of card,” “Gender,” and “Age,” among many others.

6.2 Recommendation:

The methods of big data analysis to improve fraud detection technique has revealed valuable perceptions and potential ways for development. Classification Based, Clustering Based, Nearest Neighbor Based, Statistical, Information Theoretic, and Spectral are the primary methods for anomaly identification (Quatrini *et al.*, 2020). Several proposals are put forth to further the subject

of identifying fraudulent activity in financial transactions on the basis of the data and analysis reported in this dissertation:

- **Steady Data Enhancement:** The quality and diversity of data sources ought to be continually improved in order to improve the accuracy and efficacy of fraud detection algorithms. A more complete picture of transactional behaviors and trends may be obtained by integrating many data sources, such as social media, mobile phone records, and geo-location data. This enhanced dataset can aid in the discovery of novel and developing fraud techniques.
- **Hybrid Models Fusing:** Hybrid models that integrate different machine learning approaches, such as supervised and unsupervised techniques, should be taken into account. The overall detection accuracy and resistance to evolving fraud strategies can be increased by ensemble approaches that take advantage of the capabilities of many models.
- **Real-time Monitoring:** Real-time fraud detection should be actively developed in the future. Losses can be greatly reduced by putting in place systems that can examine and warn about potentially fraudulent transactions as they take place. Achieving this aim will require integrating immediate form machine learning models, identifying anomaly methods, and stream processing approaches. CCTV camera condition is a major factor in real-time fraud detection (Rezaee *et al.*, 2021).
- **Interdisciplinary Combination:** Fraud detection is a complex problem that calls for cooperation amongst experts in a range of fields, including finance, technology, and ethics. Financial organizations, regulatory agencies, data scientists, and cybersecurity professionals working together can produce comprehensive solutions that cover the ethical and technological aspects of fraud detection.
- **Ethical Framework Reinforce:** It is crucial to strengthen ethical frameworks that put privacy, openness, and responsibility first when using sensitive financial and personal data. From data collection through model deployment, every step of the data analytics process should take ethical issues into account. To ensure the ethical use of data, rules, and regulations need also be created or updated.
- **Steady Model Evaluation and Validation:** The effectiveness of fraud detection algorithms must be continuously assessed and verified since criminals constantly adapt and develop new strategies. Using benchmark datasets and real-world data to regularly evaluate

model performance can assist maintain a high level of precision and reduce erroneous positives and negatives.

- **Training and Education:** To keep current on the newest approaches and procedures, organizations and personnel involved in fraud detection should emphasize continual training and education. This will develop a workforce that can successfully manage emerging fraud issues.

6.3 Future Scope:

This dissertation's future focus is on improving and broadening credit card transaction fraud detection techniques. Using cutting-edge machine learning methods, such as anomaly detection and deep learning, might improve the precision of spotting fraudulent activity on various card kinds. Using blockchain technology and integrating real-time data streams might improve security even more. Furthermore, investigating the possibilities of behavioral analytics and biometric identification in combination with big data analytics may result in more resilient and flexible fraud protection systems. Interdisciplinary research and ongoing industry stakeholder engagement may spur the creation of novel defenses against changing fraudulent activities. The ways shine on by this study clears a few directions for upcoming research and exploration for the enhancement of fraud detection with big data analysis. There are various areas that have potential for future research:

- **Modern Techniques of Machine Learning:** The constant advancement of ML and AI offers an opportunity to develop more difficult algorithms. It is possible to use deep learning models, natural language processing, and reinforcement learning to find complex fraud abnormalities and patterns that conventional methods could miss.
- **Resolvable AI for Transparency:** The incorporation of explicable AI methods can offer insights into how sophisticated fraud detection models make decisions. In addition to improving model interpretability, this will promote more transparency and trust in the results produced by these algorithms.
- **Verification Identity and Behavioral Biometrics:** Investigating the use of identity verification techniques and biometrics for behavior, such as keystroke dynamics, voice recognition, and face recognition, can provide more security and customization to fraud detection devices.

- **Blockchain Technology:** By using blockchain technology, money transactions may be made more secure and transparent. Due to its irreversible and decentralized character, it can help create reliable, fraud-resistant systems that can gain from data sharing across financial institutions.
- **Contextual Analysis:** Contextual analysis may be used to better understand transactions and their potential for fraud by taking into account variables like location, device use, and past transaction history.
- **Advanced-Data Fusion:** Integrating data from many sources, like wearables and Internet of Things (IoT) devices, can provide deeper insights into user conduct and result in models for fraud detection that are more precise and dynamic.
- **Ethical Framework Improvement:** The development of thorough principles of ethics that regulate the ethical application of big data analytics for fraud detection might be the subject of future research. It will be vital to strike the correct balance between privacy, security, and innovation.
- **Global Coordination and Standards:** The coordination between international institutions of finance, regular bodies, and data scientists offers of development of SOP and specification for fraud detection.
- **Improvement of Real-Time Analysis:** By expanding anomaly detection algorithms, stream processing approaches, and model adaptability, future research might be focused on improving immediate detection of fraud capabilities.
- **User-Centric Approach:** Individualized systems for identifying fraud that provides smooth user experiences while protecting financial transactions may be developed by investigating user-centric techniques that take into account user behavior, preferences, and risk tolerance. The field of fraud detection has also noted the absence of a user-centric viewpoint in XAI (Cirqueira *et al.*, 2020).

6.4 Summary:

In this summary, this dissertation starts a panoramic exploration of big data analytics methods for transforming fraud detection techniques in financial transactions. In the depth of the literature review, the traditional platform of fraud detection methods was elaborated clearly. It provides light on the importance of big data analytics and ML techniques to address the difficulties of the advancement of fraud techniques. To improve the accuracy of fraud detection, the section on

methodology offered a formal framework for putting Random Forest and decision tree classifiers into use. Big Data Analytics (BDA) is a key technology to enhance intelligence systems (Wang *et al.*, 2022). The goal of the study method, which used quantitative analysis was to find patterns, anomalies, and predictors within a huge database. In order to employ analytical approaches responsibly, ethical issues were addressed, highlighting the significance of privacy, openness, and data security. It was noted that gathering data, preparing it, and governing it were essential aspects of creating trustworthy datasets for analysis. The part on ethical issues emphasized the need of upholding people's rights and privacy while utilizing data-driven solutions. Future-focused research themes were emphasized, including cutting-edge machine learning methods, contextual analysis, blockchain integration, and user-centric strategies. This dissertation highlighted how big data analytics may revolutionize fraud detection and pave the way for more secure financial systems. Stakeholders can cooperatively develop the financial industry toward a future that is safer, more responsible, and technologically advanced by combining new tactics with ethical ideals. The road to improving fraud detection is still being traveled; possibilities for cooperation, innovation, and ethical data analytics are paving the way. This dissertation concludes by highlighting the critical role that big data analytics—especially in Python—plays in enhancing identification of corruption in financial transactions. Combining risk assessment models, security improvements, and personalized detection algorithms with data-driven insights creates a strong base against changing fraud in the digital economy.

Reference:

1. Afriyie, J.K., Tawiah, K., Pels, W.A., Addai-Henne, S., Dwamena, H.A., Owiredo, E.O., Ayeh, S.A. and Eshun, J., 2023. A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. *Decision Analytics Journal*, 6, p.100163. Available at <https://www.sciencedirect.com/science/article/pii/S2772662223000036>
2. Ariyaluran Habeeb, R.A., Nasaruddin, F., Gani, A., Amanullah, M.A., Abaker Targio Hashem, I., Ahmed, E. and Imran, M., 2022. Clustering-based real-time anomaly detection—A breakthrough in big data technologies. *Transactions on Emerging Telecommunications Technologies*, 33(8), p.e3647. Available at <https://onlinelibrary.wiley.com/doi/abs/10.1002/ett.3647>
3. Benchaji, I., Douzi, S. and El Ouahidi, B., 2021. Credit card fraud detection model based on LSTM recurrent neural networks. *Journal of Advances in Information Technology*, 12(2). Available at <https://www.academia.edu/download/81809419/20210331110920299.pdf>
4. Benedek, B. and Nagy, B.Z., 2023. Traditional versus AI-Based Fraud Detection: Cost Efficiency in the Field of Automobile Insurance. *Financial and Economic Review*, 22(2), pp.77-98. Available at <http://real.mtak.hu/168798/>
5. Gandomi, A.H., Chen, F. and Abualigah, L., 2022. Machine learning technologies for big data analytics. *Electronics*, 11(3), p.421. Available at <https://www.mdpi.com/2079-9292/11/3/421>
6. Himeur, Y., Elnour, M., Fadli, F., Meskin, N., Petri, I., Rezgui, Y., Bensaali, F. and Amira, A., 2023. AI-big data analytics for building automation and management systems: a survey, actual Bin Mofidul, R., Alam, M.M., Rahman, M.H. and Jang, Y.M., 2022. Real-time energy data acquisition, anomaly detection, and monitoring system: Implementation of a secured, robust, and integrated global IIoT infrastructure with edge and cloud AI. *Sensors*, 22(22), p.8980. Available at <https://www.mdpi.com/1424-8220/22/22/8980>
7. Chen, J.I.Z. and Lai, K.L., 2021. Deep convolution neural network model for credit-card fraud detection and alert. *Journal of Artificial Intelligence and Capsule Networks*, 3(2), pp.101-112. Available at <https://scholar.archive.org/work/yuloavynzje7vapr5d4vujma4e/access/wayback/https://irojournals.com/aicn/V3/I2/03.pdf>

7. Gaayire, R., Nikoi, S.N. and Adams, R., IMPROVING BANKING AND FINANCIAL SERVICES IN GHANA WITH BIG DATA ANALYTICS, A CASE STUDY OF AMANTIN AND KASEI COMMUNITY BANK. Available at https://www.researchgate.net/profile/Solomon-Nikoi-4/publication/370231490_IMPROVING_BANKING_AND_FINANCIAL_SERVICES_IN_GHANA_WITH_BIG_DATA_ANALYTICS_A_CASE_STUDY_OF_AMANTIN_AND_KASEI_COMMUNITY_BANK/links/6447b5f48ac1946c7a4d6017/IMPROVING-BANKING-AND-FINANCIAL-SERVICES-IN-GHANA-WITH-BIG-DATA-ANALYTICS-A-CASE-STUDY-OF-AMANTIN-AND-KASEI-COMMUNITY-BANK.pdf
8. challenges and future perspectives. Artificial Intelligence Review, 56(6), pp.4929-5021. Available at <https://link.springer.com/article/10.1007/s10462-022-10286-2>
9. https://www.researchgate.net/profile/Geethamanikanta-Jakka/publication/365675625_Automated_Banking_Fraud_Detection_for_Identification_and_Restriction_of_Unauthorised_Access_in_Financial_Sector/links/638b9d10658cec2104a91f08/Automated-Banking-Fraud-Detection-for-Identification-and-Restriction-of-Unauthorised-Access-in-Financial-Sector.pdf
10. John, H. and Naaz, S., 2019. Credit card fraud detection using local outlier factor and isolation forest. Int. J. Comput. Sci. Eng, 7(4), pp.1060-1064. Available at https://www.researchgate.net/profile/Sameena-Naaz-3/publication/335809102_Credit_Card_Fraud_Detection_using_Local_Outlier_Factor_and_Isolation_Forest/links/5d8cd723299bf10cff129722/Credit-Card-Fraud-Detection-using-Local-Outlier-Factor-and-Isolation-Forest.pdf
11. Kumar, S., Ahmed, R., Bharany, S., Shuaib, M., Ahmad, T., Tag Eldin, E., Rehman, A.U. and Shafiq, M., 2022. Exploitation of Machine Learning Algorithms for Detecting Financial Crimes Based on Customers' Behavior. Sustainability, 14(21), p.13875. Available at <https://www.mdpi.com/2071-1050/14/21/13875>
12. Masarani, K., 2021. A Study on Social Networks and Digital Security in Cyber World. Journal of Mobile Computing, Communications & Mobile Networks, 8(2), pp.36-43p. Available at https://www.researchgate.net/profile/Khushi-Masarani-2/publication/357825887_Review_JoMCCMN_A_Study_on_Social_Networks_and_Digi

tal Security in Cyber World/links/61e12c555779d35951a7c1d8/Review-JoMCCMN-A-Study-on-Social-Networks-and-Digital-Security-in-Cyber-World.pdf

13. Mohammadi, M., Yazdani, S., Khanmohammadi, M.H. and Maham, K., 2020. Financial reporting fraud detection: An analysis of data mining algorithms. *International Journal of Finance & Managerial Accounting*, 4(16), pp.1-12. Available at http://ijfma.srbiau.ac.ir/article_15385.html
14. Muheidat, F., Patel, D., Tammisetty, S., Lo'ai, A.T. and Tawalbeh, M., 2022. Emerging concepts using blockchain and big data. *Procedia Computer Science*, 198, pp.15-22. Available at <https://www.sciencedirect.com/science/article/pii/S1877050921024455>
15. Muneer, A., Taib, S.M., Fati, S.M., Balogun, A.O. and Aziz, I.A., 2022. A Hybrid Deep Learning-Based Unsupervised Anomaly Detection in High Dimensional Data. *Computers, Materials & Continua*, 70(3). Available at: <https://pdfs.semanticscholar.org/0c62/38f87d1473cade854fded548dc0c6c70f058.pdf>
16. Prabhakaran, N. and Nedunchelian, R., 2023. Oppositional Cat Swarm Optimization-Based Feature Selection Approach for Credit Card Fraud Detection. *Computational Intelligence and Neuroscience*, 2023. Available at <https://www.hindawi.com/journals/cin/2023/2693022/>
17. Ragazou, K., Passas, I. and Garefalakis, A., 2022. It Is Time for Anti-Bribery: Financial Institutions Set the New Strategic “Roadmap” to Mitigate Illicit Practices and Corruption in the Market. *Administrative Sciences*, 12(4), p.166. Available at <https://www.mdpi.com/2076-3387/12/4/166>
18. Sadgali, I., Sael, N. and Benabbou, F., 2019. Performance of machine learning techniques in the detection of financial frauds. *Procedia computer science*, 148, pp.45-54. Available at <https://www.sciencedirect.com/science/article/pii/S1877050919300079>
19. Singarimum, B.L., Dhial, A.A.M. and Farooqi, A.F., 2022. How Commercial Banks in Emerging Economies Can Leverage Big Data Analytics: A perspective of Asian countries. *International Journal of Data Science and Advanced Analytics*, 4(4), pp.94-97. Available at <http://www.ijdsaa.com/index.php/welcome/article/download/98/27>
20. Taha, A.A. and Malebary, S.J., 2020. An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. *IEEE Access*, 8, pp.25579-25587. Available at <https://ieeexplore.ieee.org/iel7/6287639/8948470/08979331.pdf>

21. Velasco-Gallego, C. and Lazakis, I., 2022. RADIS: A real-time anomaly detection intelligent system for fault diagnosis of marine machinery. *Expert Systems with Applications*, 204, p.117634. Available at <https://www.sciencedirect.com/science/article/pii/S0957417422009423>
22. Xu, M., MacDonnell, M., Wang, A. and Elias, M.J., 2023. Exploring social-emotional learning, school climate, and social network analysis. *Journal of Community Psychology*, 51(1), pp.84-102. Available at <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcop.22881>
23. Yin, Y., Jang-Jaccard, J., Xu, W., Singh, A., Zhu, J., Sabrina, F. and Kwak, J., 2023. IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset. *Journal of Big Data*, 10(1), pp.1-26. Available at <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00694-8>
24. Trivedi, N.K., Simaiya, S., Lilhore, U.K. and Sharma, S.K., 2020. An efficient credit card fraud detection model based on machine learning methods. *International Journal of Advanced Science and Technology*, 29(5), pp.3414-3424. Available at: https://www.researchgate.net/profile/Dr-Kumar-Lilhore/publication/341932015_An_Efficient_Credit_Card_Fraud_Detection_Model_Based_on_Machine_Learning_Methods/links/5ee4a477458515814a5b891e/An-Efficient-Credit-Card-Fraud-Detection-Model-Based-on-Machine-Learning-Methods.pdf
25. Baesens, B., Höppner, S. and Verdonck, T., 2021. Data engineering for fraud detection. *Decision Support Systems*, 150, p.113492. Available at: <https://lirias.kuleuven.be/retrieve/657601>
26. Nguyen, T.T., Tahir, H., Abdelrazek, M. and Babar, A., 2020. Deep learning methods for credit card fraud detection. arXiv preprint arXiv:2012.03754. Available at: <https://arxiv.org/pdf/2012.03754>
27. Bin Mofidul, R., Alam, M.M., Rahman, M.H. and Jang, Y.M., 2022. Real-time energy data acquisition, anomaly detection, and monitoring system: Implementation of a secured, robust, and integrated global IIoT infrastructure with edge and cloud AI. *Sensors*, 22(22), p.8980. Available at: <https://www.mdpi.com/1424-8220/22/22/8980>
28. Zhou, H., Sun, G., Fu, S., Jiang, W. and Xue, J., 2019. A Scalable Approach for Fraud Detection in Online E-Commerce Transactions with Big Data Analytics. *Computers*,

- Materials & Continua, 60(1). Available at:
https://cdn.techscience.cn/files/cmc/2019/v60n1/20190627024003_27050.pdf.
29. Biswas, A., Deol, R.S., Jha, B.K., Jakka, G., Suguna, M.R. and Thomson, B.I., 2022, October. Automated Banking Fraud Detection for Identification and Restriction of Unauthorised Access in Financial Sector. In 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC) (pp. 809-814). IEEE. Available at:
https://www.researchgate.net/profile/Geethamanikanta-Jakka/publication/365675625_Automated_Banking_Fraud_Detection_for_Identification_and_Restriction_of_Unauthorised_Access_in_Financial_Sector/links/638b9d10658cec2104a91f08/Automated-Banking-Fraud-Detection-for-Identification-and-Restriction-of-Unauthorised-Access-in-Financial-Sector.pdf
30. Goecks, L.S., Korzenowski, A.L., Gonçalves Terra Neto, P., de Souza, D.L. and Mareth, T., 2022. Anti-money laundering and financial fraud detection: A systematic literature review. *Intelligent Systems in Accounting, Finance and Management*, 29(2), pp.71-85. Available at: https://www.researchgate.net/profile/Davenilcio-Souza/publication/360707911_Anti-money_laundering_and_financial_fraud_detection_A_systematic_literature_review/links/629a8bfa55273755ebd07508/Anti-money-laundering-and-financial-fraud-detection-A-systematic-literature-review.pdf.
31. Chen, J.I.Z. and Lai, K.L., 2021. Deep convolution neural network model for credit-card fraud detection and alert. *Journal of Artificial Intelligence and Capsule Networks*, 3(2), pp.101-112. Available at: <https://scholar.archive.org/work/yuloavynzje7vapr5d4vujma4e/access/wayback/https://irojournals.com/aicn/V3/I2/03.pdf>.
32. Faccia, A., 2023. National Payment Switches and the Power of Cognitive Computing against Fintech Fraud. *Big Data and Cognitive Computing*, 7(2), p.76. Available at <https://www.mdpi.com/2504-2289/7/2/76>
33. Chen, J.I.Z. and Lai, K.L., 2021. Deep convolution neural network model for credit-card fraud detection and alert. *Journal of Artificial Intelligence and Capsule Networks*, 3(2), pp.101-112. Available at:

<https://scholar.archive.org/work/yuloavynzje7vapr5d4vujma4e/access/wayback/https://iournals.com/aicn/V3/I2/03.pdf>

34. Taha, A.A. and Malebary, S.J., 2020. An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. *IEEE Access*, 8, pp.25579-25587. Available at: <https://ieeexplore.ieee.org/iel7/6287639/8948470/08979331.pdf>
35. John, H. and Naaz, S., 2019. Credit card fraud detection using local outlier factor and isolation forest. *Int. J. Comput. Sci. Eng*, 7(4), pp.1060-1064. Available at: https://www.researchgate.net/profile/Sameena-Naaz-3/publication/335809102_Credit_Card_Fraud_Detection_using_Local_Outlier_Factor_and_Isolation_Forest/links/5d8cd723299bf10cff129722/Credit-Card-Fraud-Detection-using-Local-Outlier-Factor-and-Isolation-Forest.pdf
36. Trivedi, N.K., Simaiya, S., Lilhore, U.K. and Sharma, S.K., 2020. An efficient credit card fraud detection model based on machine learning methods. *International Journal of Advanced Science and Technology*, 29(5), pp.3414-3424. Available at: https://www.researchgate.net/profile/Dr-Kumar-Lilhore/publication/341932015_An_Efficient_Credit_Card_Fraud_Detection_Model_Based_on_Machine_Learning_Methods/links/5ee4a477458515814a5b891e/An-Efficient-Credit-Card-Fraud-Detection-Model-Based-on-Machine-Learning-Methods.pdf
37. Salkuti, S.R., 2020. A survey of big data and machine learning. *International Journal of Electrical & Computer Engineering (2088-8708)*, 10(1). Available at: <https://pdfs.semanticscholar.org/2710/978690f5508dfe4398af92cc4edd2b1ee4ba.pdf>
38. Khatri, S., Arora, A. and Agrawal, A.P., 2020, January. Supervised machine learning algorithms for credit card fraud detection: a comparison. In 2020 10th international conference on cloud computing, data science & engineering (confluence) (pp. 680-683). IEEE. Available at: <https://ieeexplore.ieee.org/abstract/document/9057851/>
39. Taha, A.A. and Malebary, S.J., 2020. An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. *IEEE Access*, 8, pp.25579-25587. Available at <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8979331>
40. Chen, J.I.Z. and Lai, K.L., 2021. Deep convolution neural network model for credit-card fraud detection and alert. *Journal of Artificial Intelligence and Capsule Networks*, 3(2), pp.101-112. Available at

https://web.archive.org/web/20210624021145id_/https://irojournals.com/aicn/V3/I2/03.pdf

41. Quatrini, E., Costantino, F., Di Gravio, G. and Patriarca, R., 2020. Machine learning for anomaly detection and process phase classification to improve safety and maintenance activities. *Journal of Manufacturing Systems*, 56, pp.117-132. Available at https://iris.uniroma1.it/retrieve/handle/11573/1413335/1485898/Quatrini_Preprint_machine-learning_2020.pdf.pdf
42. Rezaee, K., Rezakhani, S.M., Khosravi, M.R. and Moghimi, M.K., 2021. A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance. *Personal and Ubiquitous Computing*, pp.1-17. Available at https://www.researchgate.net/profile/Mohammad_Khosravi11/publication/352755697_A_survey_on_deep_learning-based_real-time_crowd_anomaly_detection_for_secure_distributed_video_surveillance/links/60ddbdeca6fdccb745fb821a/A-survey-on-deep-learning-based-real-time-crowd-anomaly-detection-for-secure-distributed-video-surveillance.pdf
43. Wang, J., Xu, C., Zhang, J. and Zhong, R., 2022. Big data analytics for intelligent manufacturing systems: A review. *Journal of Manufacturing Systems*, 62, pp.738-752. Available at https://www.researchgate.net/profile/Junliang-Wang/publication/352351775_Big_data_analytics_for_intelligent_manufacturing_systems_A_review/links/60cde7f2299bf1cd71de784d/Big-data-analytics-for-intelligent-manufacturing-systems-A-review.pdf
44. Alarfaj, F.K., Malik, I., Khan, H.U., Almusallam, N., Ramzan, M. and Ahmed, M., 2022. Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. *IEEE Access*, 10, pp.39700-39715. Available at: <https://ieeexplore.ieee.org/abstract/document/9755930/>
45. Balobid, A.M., Binamro, J.S., Yohannes, S.T. and Kaddoura, S., 2022, November. Evaluation of supervised machine learning approaches for credit card fraud detection. In 2022 14th Annual Undergraduate Research Conference on Applied Computing (URC) (pp. 1-6). IEEE. Available at: <https://ieeexplore.ieee.org/abstract/document/10054229/>
46. Handa, A., Dhawan, Y. and Semwal, P., 2022. Hybrid analysis on credit card fraud detection using machine learning techniques. *Handbook of Big Data Analytics and*

Forensics, pp.223-238. Available at : https://link.springer.com/chapter/10.1007/978-3-030-74753-4_15