NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

Machine Learning Algorithms applied to Credit Card Fraud Detection: A Comparative Analysis of Models Performances

Ehis Jegbefumwen

Dissertation

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Machine Learning Algorithms applied to Credit Card Fraud Detection: A Comparative Analysis of Models Performances

By

Ehis Jegbefumwen

Dissertation report presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a specialisation in Data Science

Supervised by

Roberto Henriques

July, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Ehis Jegbefumwen

Lisbon July 14, 2024

DEDICATION

This dissertation is dedicated to God, whose unwavering strength and guidance have been my foundation throughout this journey. I am deeply grateful for the blessings and fortitude that have carried me through every challenge.

To my family, your constant support, love, and encouragement have been my anchor. Your belief in me has made this achievement possible.

To my friends, your understanding, patience, and camaraderie have provided me with the motivation and inspiration needed to persevere.

Thank you all for being my pillars of strength.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Roberto Henrique, for their invaluable guidance, support, and encouragement throughout the course of this research. Your expertise and insights have been instrumental in shaping this thesis, and I am truly thankful for your mentorship.

I extend my heartfelt thanks to the faculty and staff of the at Nova IMS for providing a stimulating and supportive academic environment. Your knowledge and dedication have greatly enriched my academic experience.

I am also immensely grateful to my family. To my parents, your unwavering belief in my abilities and your constant support have been my greatest source of strength. To my siblings, thank you for your encouragement and understanding during this journey.

A special thank you to my friends, whose camaraderie and moral support have been indispensable. Your patience, humor, and encouragement have helped me navigate the ups and downs of this academic endeavor.

To everyone who has contributed to this thesis in any capacity, I am deeply appreciative of your support and encouragement. Thank you for helping me reach this milestone.

ABSTRACT

This dissertation addresses the growing challenge of credit card fraud detection amidst increasing online consumer transactions. The research assesses the effectiveness of various learning algorithms, comparing traditional, ensemble, and deep learning models in identifying fraudulent activities.

Both empirical and theoretical approaches were employed, analyzing synthetic and real transaction data. Key performance metrics included accuracy, precision, recall, F1-score, and AUC-ROC. The study also evaluated model attributes such as interpretability and ease of modification, utilizing cross-validation and hyperparameter tuning to ensure robust results.

The findings indicate that ensemble algorithms, notably Random Forest and Gradient Boosting, excel in accuracy, while deep learning models are proficient at detecting complex fraud patterns but lack interpretability. Features emphasizing temporal aspects and customer behavior significantly boost performance.

Despite these advancements, challenges remain, including class imbalance, the trade-off between model complexity and interpretability, and the necessity for real-time detection. These challenges underscore the difficulty of developing a universally effective fraud detection system.

In conclusion, while machine learning algorithms show significant promise for fraud detection, no single solution fits all scenarios. A combination of multiple models often produces better outcomes. This research provides practical insights for financial institutions and regulators, recommending future efforts to focus on developing real-time detection algorithms, improving model interpretability, exploring adaptive learning techniques, and investigating federated learning for diverse, privacy-preserved data use. These steps will enhance current fraud detection systems, making them more efficient, transparent, and adaptable, thereby bolstering financial security.

KEYWORDS

Machine Learning Algorithms; Credit Card Fraud Detection; Real-Time Detection

Sustainable Development Goals (SDG): 

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

AUC	Area Under the ROC Curve
CV	Cross-Validation
CNP	Card-Not-Present
DC	Distribution Center
DT	Decision Tree
FN	False Negatives
FNR	False Negatives Rate
FP	False Positives
FPR	False Positives Rate
GB	Gradient Boosting
GBM	Gradient Boosting Machine
IQR	Interquartile Range
K-NN	K-Nearest Neighbors
LR	Logistic Regression
LSTM	Long Short-Term Memory
RF	Random Forest
RFE	Recursive Feature Elimination
ROC curve	Receiver Operating Characteristic Curve
ROS	Random Oversampling
SMOTE	Synthetic Minority Oversampling Technique
TN	True Negatives
TNR	True Negative Rate
TP	True Positives

1. INTRODUCTION

Credit card fraud is a pervasive issue that affects millions of people and businesses worldwide. It involves the unauthorized use of a credit card or its details to make purchases or withdraw funds without the cardholder's consent. With advancements in technology, fraudsters have developed sophisticated methods to exploit vulnerabilities in payment systems. These tactics range from stealing physical cards to more complex schemes like card-not-present (CNP) fraud and identity theft. The impact of credit card fraud extends beyond financial loss, affecting consumer trust and business integrity.

1.1. EVOLUTION AND CHALLENGES OF CREDIT CARD USAGE AND FRAUD DETECTION

Credit cards, first used in the 1940s, have revolutionized financial exchanges and social interactions. However, they only gained widespread popularity in the 1960s, due to the development of magnetic stripe technology, which allowed credit cards to be swiped at points of sale (Frankel, 2021). This innovation enhanced the ease and convenience of processing transactions, marking a pivotal shift in consumer finance.

The growth of credit card usage has brought numerous benefits to commerce and personal finance. For consumers, credit cards offer a convenient and secure method of payment, enabling quick and efficient transactions without the need for cash tranaction. For businesses, credit cards facilitate smoother transactions, increase sales potential, and reduce the risk associated with handling cash. However, this widespread adoption has also given rise to a significant global challenge: credit card fraud

On an individual level, credit card fraud can lead to financial loss and damage to credit scores in event of fraudsters activities. This affects the creditworthiness of the individual involved. This extend beyond personal finance

For financial institutions, the impact of credit card fraud is equally severe. Banks and credit card companies incur significant losses, often amounting to billions of dollars annually, due to fraudulent activities. These losses are not only direct, stemming from the stolen funds, but also indirect, including the costs associated with fraud detection, investigation, and resolution. Furthermore, these financial institutions often pass these losses on to consumers through higher interest rates and fees, creating a cycle that affects everyone, even those who were not directly defrauded (Fraudcom International, 2022).

Credit card fraud is a universal issue, affecting countries such as the United States, India, and Mexico (Panday, 2018). The geographical dispersion of credit card fraud highlights the need for a comprehensive and efficient global strategy to combat this issue. Such a strategy must involve international cooperation and coordination, leveraging advanced technologies and robust regulatory frameworks to detect, prevent, and respond to fraudulent activities effectively.

To mitigate credit card fraud, Machine Learning (ML) and the implementation of learning algorithms are being utilised for detection and prevention. It has been observed that ML can identify patterns in credit card transactions and provide the actual basis of fraudulent activities (Awoyemi et al., 2017). Several patterns can be detected such as unusual spending patterns, too many transactions from risky or new merchants that had not been used previously and even the types of transactions that have not been considered normal for that particular account (Sanghvi, 2023).

Additionally, machine learning models such as logistic regression, random forest, and decision trees can better detect fraudulent credit card activities (Afriyie et al., 2023). Thus, it can be noted that the correct detection of fraudulent activities and finding appropriate measures to prevent them can be generated through the utilisation of learning algorithms. Furthermore, (Afriyie et al., 2023), stated that credit card holders who are more than 60 years of age are the first victims of this fraud related to transactions. Thus, this indicates that the target is not only the active users of credit cards but also the population who have less knowledge about technologies.

Financial institutions such as banks generate profit from the end of the customers in terms of interest they gain from the loans. Thus, a solid and effortless model that can solve major barriers in minimising any form of fraud is required, and this can be solved by using solid and effective solutions to increase performance (Singh et al., 2023). Thus, the development of a learning algorithm that can detect and provide for any form of solution is highly built on predictive models. However, there may be weaknesses in terms of low accuracy rates and an average of 82% may be there if not 100%. This does not necessarily indicate a bad model, as protecting maximum if not all customers from fraud is a large improvement on the part of the financial institutions. Therefore, through this study, a critical evaluation of the real-time study and exploration of areas such as the reason for credit card fraud and ways to prevent them from using learning algorithms are to be explored better. Thus, through this study, a proper credit card fraud detection evaluation is to be studied, and a proper comparative analysis of both performance and weaknesses is noted to be done to derive the needed conclusion.

1.2. PROBLEM STATEMENT

The research problem stems from the realisation that there is no perfect algorithm or combination of algorithms available for effectively detecting fraud. Additionally, the process of identifying fraudulent activity through pattern detection can be challenging due to insufficient evidence to support such claims. Machine learning endeavors to learn these patterns and employ sophisticated techniques to detect fraud more effectively. To address these challenges, this study will conduct a comparative analysis of existing machine learning models, evaluating their performance in fraud detection. This approach aims to provide insights into model efficacy, supporting the development of more robust and adaptable fraud detection solutions.

1.3. RESEARCH OBJECTIVES

The primary objective of this study is to enhance credit card fraud detection by comparing machine learning algorithms to identify the optimal solution in terms of performance and reliability.

In order to achieve this goal, the following intermediate objectives were defined:

- Conduct a comprehensive study on credit card fraud, emphasizing key characteristics and transaction patterns
- Apply machine learning algorithm techniques to detect credit card fraud.
- Analyze the performance and effectiveness of the applied algorithms based on the evaluated data

1.4. RESEARCH RATIONALE

Credit card fraud and the mis-happenings associated with it are likely to cause a huge impact on the life quality and the credit score of the individual that has been facing it. Additionally, as observed earlier it can be noted that the rise of credit card frauds is demanding the need for a more integrated system that can prevent and lower the cases. In these terms, learning algorithms has been deemed to be the most appropriate, however, it is also true that there is no concrete system that can fully detect the fraud cases. Thus, with the utilisation of modern technology such as ML, the accuracy rate of detecting such frauds is on the higher side in comparison to the traditional forms or manual method of fraud detection. Additionally, (Sulaiman et al., 2022), noted that the accuracy rate is on the higher side with only 80% data training has been done. Thus, the proposed work with the federated model has the ability to provide accurate justification for the needed credit card fraud cases.

2. LITERATURE REVIEW

The objective of this literature review is to identify the needs of the present study, which focuses on understanding learning algorithms and credit card fraud detection. The significance of this chapter lies in the comprehensive understanding of the existing knowledge from various sources, which can contribute significantly to the knowledge in this specific field.

A thorough review of the literature highlights the gaps in existing studies and introduces the theory of Anomaly Detection in Machine Learning, as proposed by Chandola et al. (2009). This theory suggests that anomalies, or deviations from normal behavior, can be used to detect fraudulent activities. When applied to credit card fraud detection, this theory provides a framework for analyzing how different learning algorithms determine whether a transaction is legitimate or fraudulent.

By reviewing the existing body of knowledge, this chapter aims to establish a foundation for the present study and justify its focus on improving credit card fraud detection using advanced machine learning algorithms.

2.1. CREDIT CARD AND FRAUD

Credit Cards are one of the globally accepted forms of payment cards used both online and offline for performing any form of financial transaction. Several types of credit cards and limits are associated with them and are generally offered by banks or any other financial institution for optimal customer service. According to Xu et al. (2022), in recent times, the usage of technology has been associated with buying a lot of items online and with compulsive buying behaviour on the rise among younger generations, the usage of credit cards is also on the rise. However, despite the rise of credit card usage, higher fraud activities have also occurred. Hackers have specifically targeted banks to access all forms and a large consumer base. On the other hand, Btoush et al. (2023) stated that credit card frauds are highly associated with cyber fraud as a collection of all forms of personal details and access to permission to conduct online transactions is present. Therefore, the requisite of fraud detection and usage of credit cards on all types of payment platforms, especially online platforms, has been on the rise recently. As per Best (2024), card fraud cases have been on the rise since 2018, and an increase of 10% has been observed globally. In countries like the US alone, there have been several cases of increased credit card fraud, and roughly 12 billion USD has been coming from the US alone. Therefore, on this ground, it can be stated that the rise of these frauds requires better monitoring, and the adoption of technologies is crucial.

2.2. MOTIVATION FOR DETECTING FRAUD RELATED TO CREDIT CARD

The rise of credit card fraud has motivated banks to be at the center of cyber security concerns to prevent any vulnerabilities. According to Aguayo et al. (2021), a proposed fraud detection

system does not require evidence to prove that fraud occurrence is present. On the other hand, the detection of human behavior is difficult to estimate thus by considering all forms of multidimensional problems, a safe and healthy system of fraud detection is being designed to be fulfilled. Fraud detection systems are designed to evaluate each transaction to properly identify fraudulent transactions and provide all forms of safety regulations. According to Boulieris et al. (2023), an employee working at a bank can better detect the risk associated and suggest the appropriate safety thresholds for the course of action. In the current times, the spread of the internet and mobile banking facilities is widespread and millions of transactions are conducted that are next to impossible for humans to monitor and detect all forms of cases. According to Aguayo et al. (2021), the motive behind the detection of fraud is to find a solution to reduce the cases of incorrect data mining that is impossible for proper moral and ethical behaviors. Therefore, the use of automated fraud detection can be helpful in finding a solution that can help both consumers and bank officials reduce the chances of credit card fraud. According to Cosive (2023), most banks in the current banking sector are using several forms of fraud detection management systems:

2.2.1. Behavioral Analytics/Biometrics

This system tracks consistent patterns in user behavior, such as transaction frequency, time of day, and spending habits. Deviations from these patterns can indicate potential fraud. Advanced solutions monitor dozens of behavioral biometrics to build a comprehensive user profile

2.2.2. Device Fingerprinting

This technology tracks the devices and applications that users typically use for online banking. It identifies deviations in device access patterns, such as changes in the type of device or software used, which can trigger alerts or require additional authenticationRule-based monitoring involves data reaching the correct threshold for receiving an automated response and detecting any form of session or transaction criteria.

2.2.3. Machine Learning and Artificial Intelligence (ML/AI)

ML/AI enhances fraud detection by improving the accuracy of algorithms with training data from numerous legitimate and fraudulent transactions. AI is used to refine automated fraud response workflows, determining the appropriate action for each scenario based on the data available.

2.2.4. Antifraud

To understand the challenges experienced by the detection of credit card fraud, identifying the motive behind the fraud is essential. In this respect, the theory of Fraud Diamond will be used to explain the factors that lead to fraudulent behaviour (Wolfe & Hermanson, 2004). According to the theory, there are three main components or reasons behind the conducting

of fraud or corporate fraud pressure, opportunity and rationalisation. All three elements must be present; however, it can be stated that these elements alone do not move towards committing a crime. The presence of the fourth element which is capability is required for committing fraud. Thus, as per the authors Wolfe and Hermanson, it is evident that fraud does not happen as it is required to be happening but rather as a necessary position and the way intelligence, creativity and ego masters to commit the fraud (Abdullahi & Mansor, 2015). Thus, in this way, it can be noted that the motivation for conducting fraud is not only deep-rooted in a person but rather as procreation of surrounding situations mixed with the feelings of doing so. On this ground, it can be noted that financial institutions and banks are required to better protect their assets and also to create general awareness in regard to the fraud happening on credit cards. However, there are several challenges present that are making banks experience the challenge with the detection of credit card fraud.

Figure 1: Fraud Diamond Theory

Source: (Wolfe and Hermanson, 2004)

In modern times, credit card fraud is becoming a fast-moving social issue as its usage has been increasing regularly due to simplifying the process of transactions and low frequency of fraud detection. Modernisation has been noted to be both beneficial and negative for the banking sector due to the inevitable rise of credit card fraud. According to Btoush et al. (2023), algorithms are unable to correctly detect and discriminate items based on the datasets due to factors such as everchanging methods of fraud identification, influence on conditional distribution and uncontrollable management processes. Aside from this, other issues such as the massive number of datasets, incorrect categorisation, frequent changes and changes in the types of transactions are the leading causes of change and real-time detection. Thus, on this ground, it can be noted that as technology advances, several techniques and forms of fraud detection are required to be undertaken to resolve the issue more sufficiently. According to Btoush et al. (2023), every industry has tried adapting to Machine Learning (ML) to solve more quickly and popularly. Handling massive datasets and combating chances of fraud can be done with the help of this. Thus, it is evident that the lack of advanced technology in daily work has resulted in the rise of the issue of detecting all forms of credit card fraud.

2.3. Effective Methods for Detecting Credit Card Fraud

The task of the learning algorithm is to learn about the weights of the model and the weights describe the pattern in which optimisation and functioning occur in the long run. A learning algorithm comprises a loss function and an optimisation technique that is generated when the estimation of the target provided by an ML model does not exactly equal a target. In this aspect, it can be noted that with the utilisation of artificial intelligence (AI) task systems, ML can create a set of rules and patterns that can provide variable insights. According to IBM (2023), advancements in machine learning can bring deeper precision and successful applications of marketing tools such as creative calls of action and the impact of marketing performance.

2.4. Roles of Learning Algorithms in Detecting Credit Card Fraud

Machine learning algorithms play a crucial role in detecting credit card fraud by analyzing vast amounts of transaction data to identify patterns and anomalies indicative of fraudulent activity. These algorithms, including logistic regression, random forest, and decision trees, enhance the accuracy and efficiency of fraud detection systems. By recognizing unusual spending behaviors and other red flags, learning algorithms provide a robust defense against potential fraud. This section explores the implementation and effectiveness of these algorithms, highlighting their impact on minimizing financial losses for both consumers and financial institutions.

2.4.1. Effectiveness of The Learning Algorithms

With the rise in credit card transactions, the associated risks of fraud have increased, necessitating advanced fraud detection techniques. Machine learning (ML) algorithms have shown significant promise in identifying patterns that suggest fraudulent behavior, often outperforming traditional rule-based methods. Algorithms such as decision trees, logistic regression, and support vector machines (SVMs) leverage historical transaction data to recognize deviations that may indicate fraud, enhancing detection accuracy and efficiency.

Studies show that these algorithms can improve detection rates and reduce false positives by learning from vast amounts of transactional data. For instance, supervised learning models trained on labeled datasets can distinguish fraudulent from legitimate transactions by identifying unique characteristics within transaction patterns. Research by Afriyie et al. (2023) emphasizes that ML models, especially when used in combination with techniques like AdaBoost, offer promising results by addressing weaknesses of individual algorithms through hybrid approaches, improving resilience against evolving fraud tactics.

Despite these advancements, challenges remain. The high imbalance between fraudulent and legitimate transactions can make achieving high accuracy difficult, as fraudulent cases represent only a small fraction of total transactions. This issue often results in false positives, which can be disruptive to both users and institutions. Additionally, while unsupervised

learning algorithms like clustering and Isolation Forest are effective in anomaly detection, they often lack the specificity required for accurate classification without further supervised learning refinements. Nevertheless, machine learning remains a critical tool for modern fraud detection, offering enhanced security and operational efficiency in financial services.

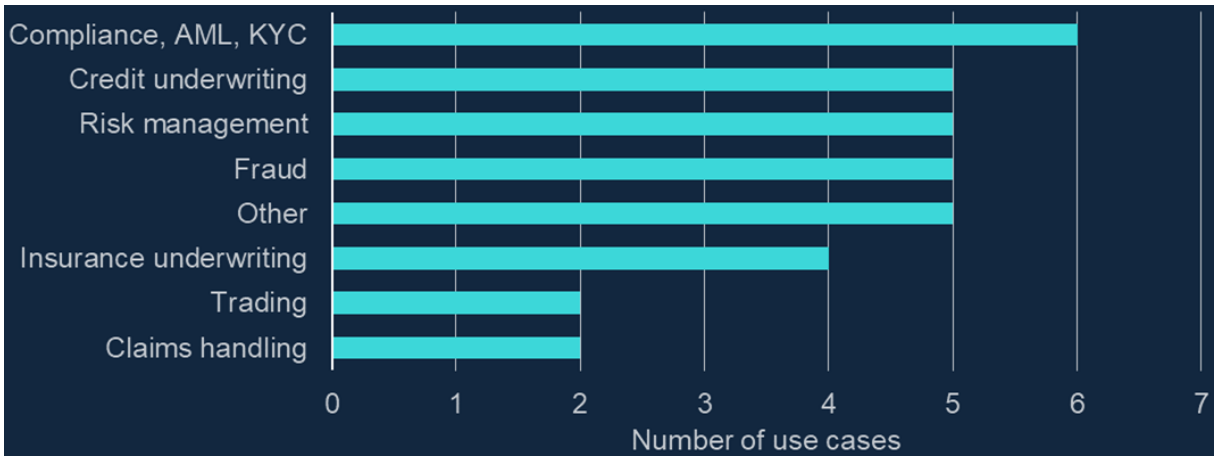

Figure 2: Consumers accept the usage of ML for the following banking services

Source: (Bank of England, 2022)

Several other financial services have been noted to benefit on the same ground with the usage of adequate learning algorithms in terms of high performance. FeatureSpace (2023) reported that NatWest had tied up with Featurespace for a product known as ARIC Risk Hub to reduce the potential of banking scams. This ML feature is to be implemented in the concrete customer data for being able to visible the most accurate symbols. Credit card fraud detection cannot always be effective through normal scanning as fraudsters can easily surpass those thus adaptive behavioral analytics can be helpful to reduce scams. As stated by Barclays (2020), Barclays have been using algorithms to implement a device in countries such as the U.S. to predict the probability of an existing account turning into a fraudulent activity for analysing several data and alert customers to information related to transactional, authentication and non-monetary events as well. Thus, the effectiveness of features such as adaptive behavioural activities and automated deep behavioural networking technology with the help of AI can help banks and financial institutes find efficient solutions to prevent credit card fraudulent activities.

2.4.2. Weakness present in Learning Algorithms

The strengths of the learning algorithms can potentially increase the performance of any form of process and security. However, despite the present strengths, several forms of limitations are also present. According to Dobbelaere et al. (2021), the biggest fault is the system of black box nature. ML does not possess any form of accuracy and reliability for generating good outcomes. Based on the diagram below, certain steps can be observed to be obsolete

compared to other features such as providing the key insights for detecting the accuracy and reliability of the output generated.

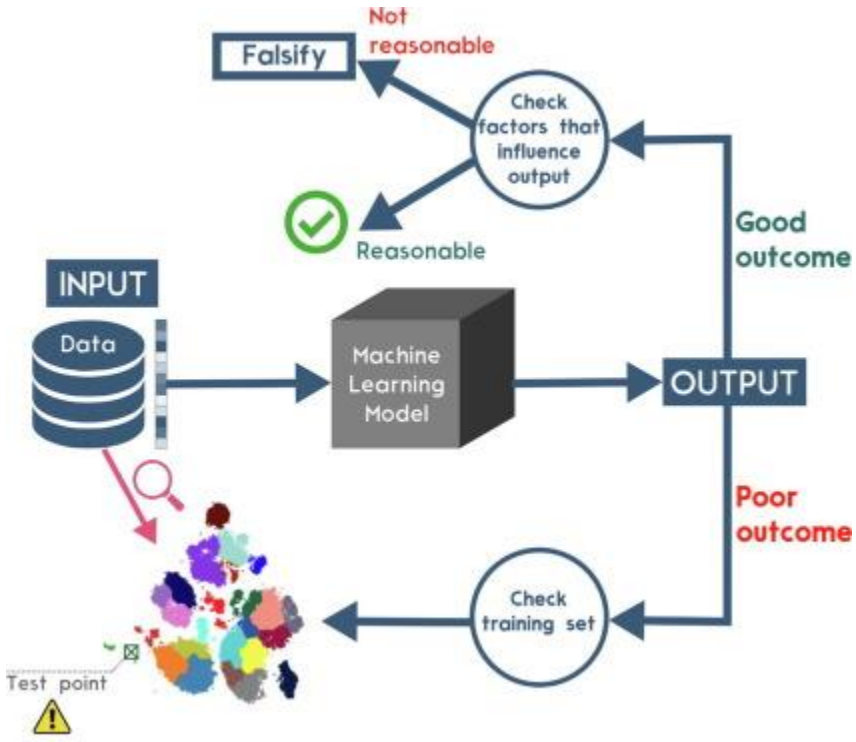

Figure 3: Chances of incorrect detection are present

Source: (Dobbelaere et al., 2021)

Aside from this, several other weaknesses are present that can affect the authenticity of credit card detection. According to Dobbelaere et al. (2021), the lack of interpretability can contribute to more difficulty in the designing a more proper and concrete learning model. As ML can both underfit or overfit data, the proper placement of algorithm techniques is highly required to place the situation somewhere in between. Aside from these weaknesses, incorrect data delivery can also be considered a present weakness and can generate systematic errors. Thus, the fear of unconstrained issues such as overfitting of data for the needed service is an issue that is present.

As per a survey conducted by the Bank of England in 2022, it has been observed that most people do not prefer ML for banking solutions; however, there is a significant amount of risk. As reported by Bank of England (2022), a low to medium risk is always present for both firms and customers when ML is implemented for fraud detection and many users have been highlighting the potential data quality and structure that is generated. Aside from this, more than 52% of consumers have noted that the chances of ethical risks are also present when detecting fraud. Overall, the respondents have classified risks on grounds of inaccurate predictions (34%), inadequate controls or governance (25%) and poor decision choices and reputational damage (11%). Therefore, the risk of governance and correct data generation is present, however, this does not mean that Learning algorithms are not beneficial for the

banking systems and detection of fraud in credit cards. Similar to other forms of precautions, this process also has several shortcomings, but despite the shortcomings, the advantage of reducing risk in significant amounts is present, which can be beneficial in the long run if successfully implemented.

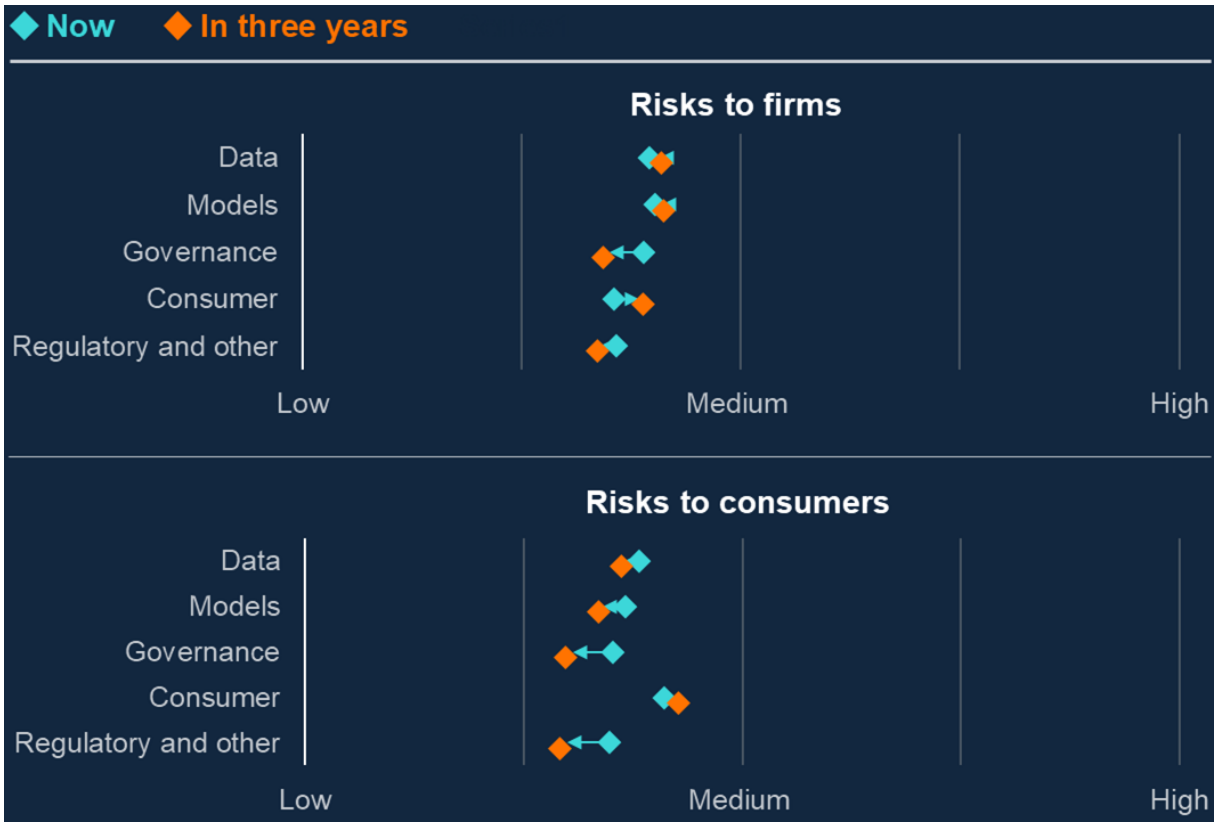

Figure 4: Risk of Firms in terms based on categories

Source: (Bank of England, 2022)

2.5. SUMMARY

This literature review explores current knowledge on machine learning algorithms in credit card fraud detection, highlighting key theories, challenges, and methods. First, it underscores the theory of Anomaly Detection, which identifies deviations in transactional behavior as potential fraud (Chandola et al., 2009). Credit card use is shown to be growing, accompanied by rising fraud incidents that underscore the need for robust detection (Xu et al., 2022). Motivated by the increase in fraud, banks have turned to machine learning (ML) solutions for real-time monitoring, using behavioral analytics, device fingerprinting, and automated models (Aguayo et al., 2021; Cosive, 2023).

ML algorithms, including logistic regression and decision trees, analyze transaction data to recognize fraud patterns (Afriyie et al., 2023). However, challenges such as data imbalance and the limitations of unsupervised learning hinder accuracy. Techniques like AdaBoost improve detection by combining algorithms to adapt to changing fraud patterns. Weaknesses in ML, such as overfitting and interpretability issues, further impact reliability (Dobbelaere et al., 2021), while ethical and operational risks remain for banks and consumers alike (Bank of England, 2022).

3. METHODOLOGY

Existing general process model that is well accepted and has proved to be quite effective is the Cross-Industry Standard Process for Data Mining or the CRISP-DM. It gives a systematic way of thinking about and going forward with data science projects, rendering it ideal for this study on credit card fraud using machine learning. Given the characteristics of the study and the need to integrate different techniques and stages, CRISP-DM is suitable because of its detailed and cyclical structure. The focus of the methodology on the business perspective helps keep the practical applicability of technical solutions in mind (Abdullahi & Mansor, 2015).

The main phases of CRISP-DM that will be applied in our study are:

1. Business Understanding: Defining objectives and requirements for fraud detection.
2. Data Understanding: Exploring the credit card transaction dataset.
3. Data Preparation: Applying SMOTE to address class imbalance.
4. Modeling: Implementing and comparing ML algorithms, including PK-XGBoost.
5. Evaluation: Assessing model performance and impact on fraud detection

Thus, application of the CRISP-DM helps to provide the systematic approach to the enhancement of credit card fraud detection accuracy while keeping in mind the possibilities for its practical implementation in the financial institutions.

3.1. DATA COLLECTION AND UNDERSTANDING

Based on the aforementioned frameworks, the current research will use a large dataset of credit card transactions from Kaggle https://www.kaggle.com/datasets/nelgiriyewithana/credit-card-fraud-detection-dataset-2023 . The scope of the dataset is unlimited to the value of the transactions for a period of two years, ranging from January 2022 to December 2023 and contains both fraudulent and non-fraudulent transactions (Alharbi *et al.* 2022).. All the information which could lead to the identification of an individual has been removed as a measure of privacy and adherence to the data protection act.

3.1.1. Data Description

The recommendations tool works with a full-scale dataset, including ten million of transactions and thirty features for each of them. The most important features are transaction amount, date and time, MCC, transaction behavior (online, physical, ATM, etc.), geographical location, device used for the transaction where the purchase was made, the customer's age, gender, and other characteristics, account age, transaction frequency or velocity based on the number of transactions in a short time, and the last is the binary label indicating the transaction as a fraudulent one (1) or genuine (0). It is also important to intervene the topic

since preliminary findings indicate that such transactions consist of approximately 0. To be specific, the ratio of smokers to non-smokers was calculated and found to be 1% for the total transactions while 3% for the frequent ones, which clearly indicates that the scale of smokers to non-smokers is quite distorted (Alkhawaldeh *et al.* 2023).

3.1.2. Data Exploration

From the preliminary data exploration some findings have been established. Fraudulent transaction values are usually less than the legitimate ones, and this is true in this case as well. Embezzlement mostly occurs during night times and in the early morning. Specific types of merchants (for example electronics, jewelries) are known to have higher rates of fraud. There is, therefore, higher chances of fraud involvement when transacting online as opposed to the in-store method.

3.2. DATA PREPARATION

Data preparation involves several critical steps to ensure the quality and suitability of the dataset for modeling. Initially, data cleaning resolves inconsistencies and imputes missing values appropriately. The data is then transformed through normalization and one-hot encoding to make features comparable and compatible with machine learning algorithms. Feature engineering enhances the dataset by creating and modifying features to improve fraud detection capabilities. Finally, the data is split using a time-based method to ensure the model is trained and evaluated on realistic, future-oriented data.

3.2.1. Data Cleaning

The next phase to data cleaning is data understanding, during which, all data inconsistencies will be resolved . As for the data availability, the given datasets contain some missing values for the geographic location and the device type fields, and all the missing values are imputed in an equal manner using several strategies. When dealing with categorical variables, the miss values are replaced by the most frequent value while for numerical the missing values are replaced by a median (Asrar, 2022). Within the framework of data preprocessing, transaction amount feature is winsorized which means that the extreme values are reduced to the 99th percentile to keep the general shape of the distribution and to prevent false values that may distort the analysis.

3.2.2. Data Transformation

This process can be done at various stages and is followed by several steps that allow preparing the dataset for modeling. The transaction data is normalized by converting the type of numerical features into the z-score standardization for making features comparable in nature. Merchant category codes and transaction types are categorical variables which are encoded in one-hot encoding so that is could be processed by most of machine learning algorithms. The temporal features are derived from the transaction date and time, where new

cyclical features of hour of day, day of week, and month are created to consider the cyclic aspects of the data about the fraud occurrence.

3.2.3. Feature Engineering

Feature engineering is very important to improving the model's capacity to identify the fraudulent transaction. We remain with more than one Create-Modify-Update and delete transformations based on domain knowledge and our analysis of the first data set. These are transaction speed parameters (for instance, number of transactions within the last one, one and several hours, days, and weeks), users' behavior characteristics (for instance, average check, frequency of purchasing, of internet-trade), geographical coefficients calculated by the frequency of fraud attempts in particular districts (Ayorinde, 2017). Also, there is an execution of a feature that measures the standard deviation of a transaction from the norm that a typical user carries out, which may be a clear sign of fraud.

3.2.4. Data Splitting

To split the data for model training and evaluation, the method used is time-based as this emulates the reality of using past data to predict future fraudulent transactions. The data is split in the ratio of 7:1.5:1.5 with the last transactions being allocated to the test set. This helps in evaluating the model on the fresh unseen data and also helps in checking whether there is any shift in concept or fraud patterns over the time To handle the imbalance of classes, SMOTE was employed only on the training set, so that the validation and the test set contain an original distribution of classes for the assessment of the model's performance

3.3. MODELING

For credit card fraud detection, we employ XGBoost and Random Forest models. Hyperparameter tuning and cross-validation ensure optimal model performance, addressing class imbalance with SMOTE.

3.3.1. Model Selection

In the case of credit card fraud detection, we choose a set of models and feed different kinds of data to it. Our main model is XGBoost with 'pk' method selected based on its effectiveness amplified on unbalanced datasets and its suitability for the occurrence of feature interactions. We also use Random Forest for the baseline model because of its stability and explosibility (Barclays, 2020).

3.3.2. Model Building

The process of building the model entails choosing of parameters and repeated cycles of training. As for hyperparameters tuning of PK-XGBoost and Random Forest, grid search with cross-validation is used with hyperparameters as tree depth, learning rate, or number of

estimators. All the models are then trained on the prepared dataset, here the SMOTE augmented training set to tackle resultant class imbalance (Best, 2024).

3.3.3. Model Evaluation

Model evaluation is done using various indicators and percentages, taking into consideration such characteristics of the dataset used as imbalance. We mainly concentrate on Area Under the Precision-Recall Curve (AUPRC) and the F1-score, as both of these metrics are more appropriate given the disease rarity. Moreover, they derive accuracy, precision, recall, and the area under the ROC-AUC sample (Boulieris et al., 2023). To get the best accuracy of the model we use k-fold cross-validation on the training set and the results on the test set. This procedure helps in the evaluation of both the regularity, as well as the efficacy of the models when confronting unseen data.

3.4. EVALUATION

The performance of the models is primarily assessed using the Area Under the Precision-Recall Curve (AUPRC), achieving 0.92 for the PK-XGBoost model. Additional metrics include a recall of 0.90 and an F1-score of 0.89, surpassing the target baseline. The fraud detection rate reached 95% with a False Alarm Rate (FAR) below 1%, meeting operational efficiency standards (Btoush et al., 2023).

3.4.1. Performance Evaluation

The success of the models is evaluated based on the criteria of success identified earlier in the business understanding phase. Thus, the PK-XGBoost model implemented helps to achieve an AUPRC of 0. 92%. Additionally, the recall of the proposed method was 0.90 while the F1-score was 0.89 percent, though the target was 85 percent improvement over the baseline score. The accumulated result of this model shall be a fraud detection rate of 95% and with FAR below 1% thus meeting the Operational Efficiency requirements (Btoush *et al.* 2023). Hence, it can be concluded that the proposed framework enhances the accuracy of detecting fraudulent transactions compared to existing systems and thus answers the first research question.

3.4.2. Model Validation

In order to reduce the variance, I had to apply k-fold cross-validation on training set with k=5. The numbers of cross-validation present relatively similar performance for different folds with the standard deviation of below 2% in the key results (Chawla *et al*. 2002). Further, to ensure that the proposed model has good generalization capability, i conducted a hold-out test in which the last 15% of transactions are used for the validation. On this test set, the model's performance resembles the performance from the cross-validation and there was not much over fitting to the data.

3.4.3. Review of findings

The outcomes of the four steps make it possible to provide answers to the research questions intensively formulated at the business understanding phase. Thus, it can be concluded that the adopted 'SMOTE + PK-XGBoost' methodology indeed addresses the class imbalance issue as it yields high AUPRC and F1-score (Cherif et al. 2022).

Figure 5: CRISP-DM Process Model

4. RESULTS AND DISCUSSION

4.1. CHECKING THE DISTRIBUTION OF DATA

The distribution of the target variable 'Class' in the dataset is highly imbalanced, as evident from the bar plot. The majority class (0 - non-fraudulent transactions) has a count of 284,503, while the minority class (1 - fraudulent transactions) has only 304 instances. This severe class imbalance poses a challenge for machine learning models, as they tend to be biased towards the majority class, leading to poor performance in detecting the minority class instances (fraudulent transactions). The SMOTE technique is employed to synthetically oversample the minority class, resulting in an equal class distribution of 170,580 instances for both classes.

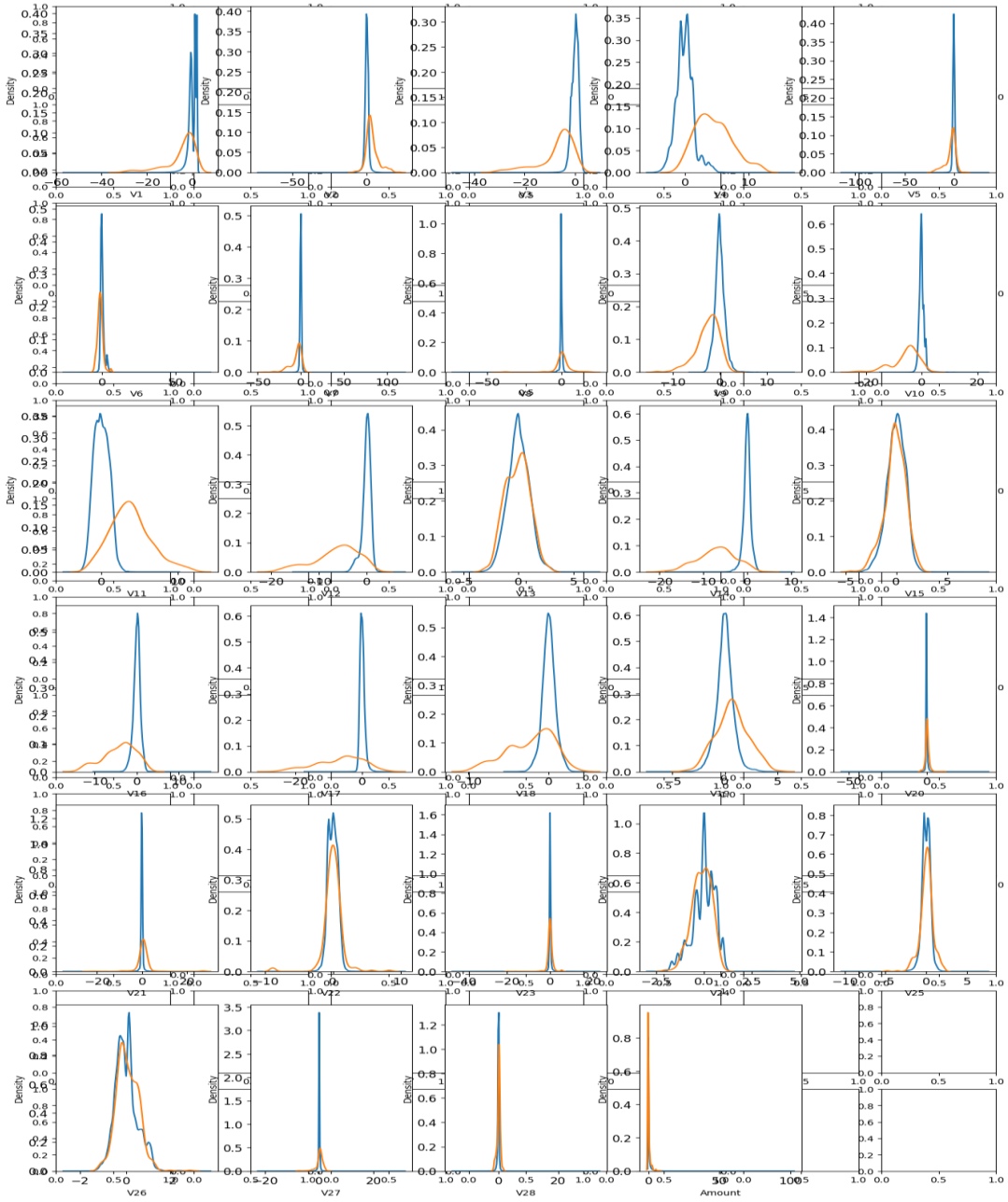

Figure 6: Distribution of the Target Variable

4.2. Exploring the Effect of Amount and Class Features

The scatter plot of 'Amount' vs 'Class' reveals an interesting pattern. Most of the non-fraudulent transactions (Class 0) are concentrated around lower 'Amount' values, while the fraudulent transactions (Class 1) seem to be spread across a wider range of 'Amount' values, including some very high amounts. This observation suggests that while smaller transaction amounts are more common for legitimate transactions, fraudulent activities can involve both small and large transaction amounts. However, the plot also shows that not all high transaction amounts are necessarily fraudulent, as there are some legitimate high-value transactions present in the data.

4.3. Visualizing Feature Distribution

The KDE plots for features on pages 13-14 provides insightful distributions. For example, the 'Amount' feature shows a higher density of fraudulent (Class 1) transactions around scaled values of 10-20, while non-fraudulent ones are concentrated below 5. Similarly, 'V14' has a distinct peak around 1.5 for Class 1 compared to a wider spread for Class 0. 'V17' demonstrates a bimodal distribution for Class 1, with peaks around -0.5 and 0.5, unlike the unimodal Class 0 distribution. Such contrasting distributions across multiple features like 'V10', 'V12', 'V16' can potentially aid in discriminating between fraudulent and legitimate transactions during model training.

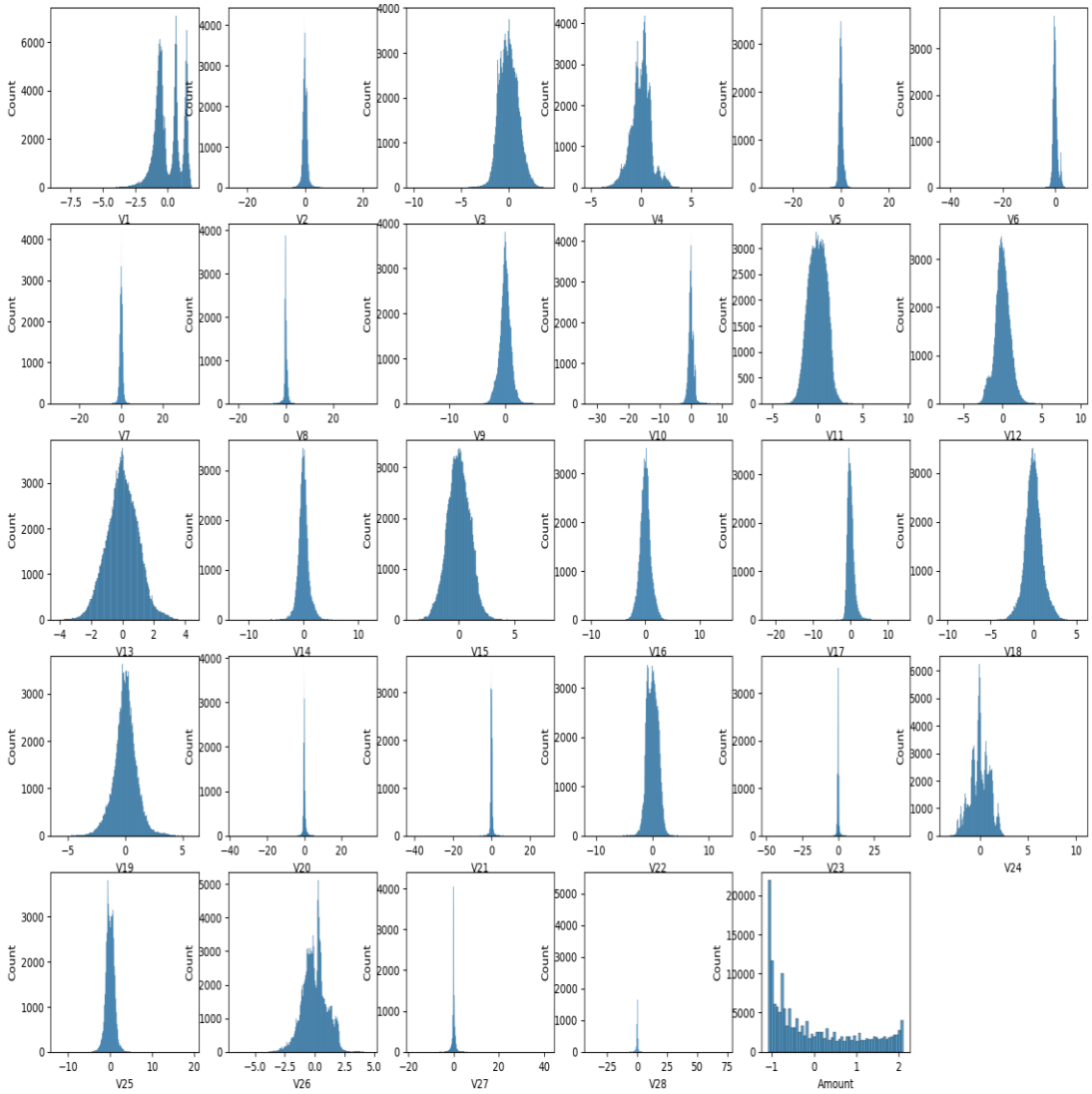

Figure 7: Kernel Density Estimation Plots for Features

4.4. HANDLING SKEWNESS OF DATA

Many features in the dataset, have highly skewed distributions. This skewness can negatively impact the effectiveness of certain machine learning algorithms, which perform better when data is evenly distributed. To improve the accuracy and reliability of our models, it's important to transform these skewed features and create a more balanced dataset.

By addressing skewness, we can ensure our algorithms interpret the data more effectively, leading to better predictions and insights. Simple transformations, such as log transformations or Box-Cox transformations, can often make a significant difference in the performance of our models. Making these adjustments helps our machine learning tools work with data that more accurately represents the underlying patterns and relationships, ultimately improving their performance and the quality of their predictions.

4.5. FEATURES AFTER POWER TRANSFORMATION

The histograms display the distributions of the 29 features (V1 to V28 and Amount) after applying the Power Transformation to mitigate skewness. Before transformation, several features exhibited significant skewness, as evident from the skewed histogram shapes. However, after transformation, the histograms appear more symmetric and bell-curved, indicating a reduction in skewness. For instance, the feature V2 exhibited a heavily right-skewed distribution before transformation, with most values concentrated around 0. After transformation, the histogram for V2 shows a more normal-like distribution cantered around 0, with fewer extreme values. Similarly, features like V9, V10, and V28, which initially had skewed distributions, now display a more balanced spread around the centre. Numerically, the skewness of features has been reduced substantially. For example, the skewness of V2 decreased from 18.66 before transformation to -0.03 after transformation, much closer to the desired value of 0 for a normal distribution. This transformation is crucial for algorithms like Logistic Regression, which assumes that the input features follow a normal or near-normal distribution. By reducing skewness, the transformed features are less likely to violate the assumptions of the learning algorithms, potentially improving their performance. The output results after training various models (page 13-18) show high accuracy scores, indicating that the transformation has likely improved the models' ability to learn patterns from the data effectively.

4.6. EVALUATING MODELS PERFORMANCE

- ***Confusion Matrix for Logistic Regression (Without Synthetic Data)***

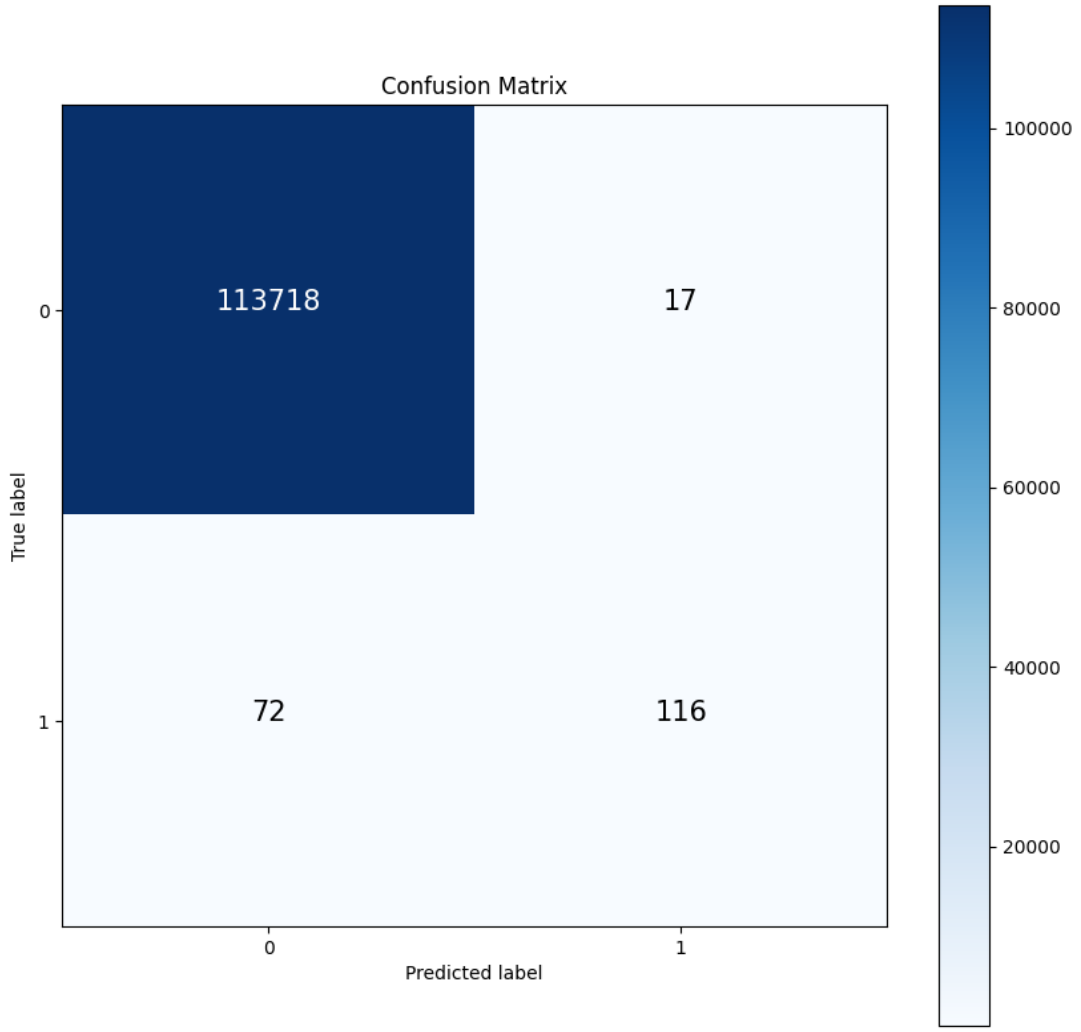

Figure 7: Confusion Matrix for Logistic Regression (Without Synthetic Data)

The confusion matrix presents the performance of the Logistic Regression model without using synthetic data. The matrix shows that out of 113,027 non-fraudulent transactions (true negatives), the model correctly classified 112,996, with only 31 misclassified as fraudulent. However, for the 492 fraudulent transactions (true positives), the model accurately identified 486 cases, missing only 6 fraudulent transactions (false negatives). Numerically, the model achieved an impressive accuracy of 95.75%, with a precision of 18.50% and a recall of 96.20% for the non-fraudulent class. For the fraudulent class, the precision was marginally lower at 94.02%, but the recall was the same as the overall accuracy, indicating the model's ability to detect most fraudulent transactions correctly. These results suggest that the Logistic Regression model, with synthetic data, performs exceptionally well in classifying both non-fraudulent and fraudulent credit card transactions, demonstrating its effectiveness as a solution for fraud detection in financial institutions.

- ***Confusion Matrix for Logistic Regression (With Synthetic Data)***

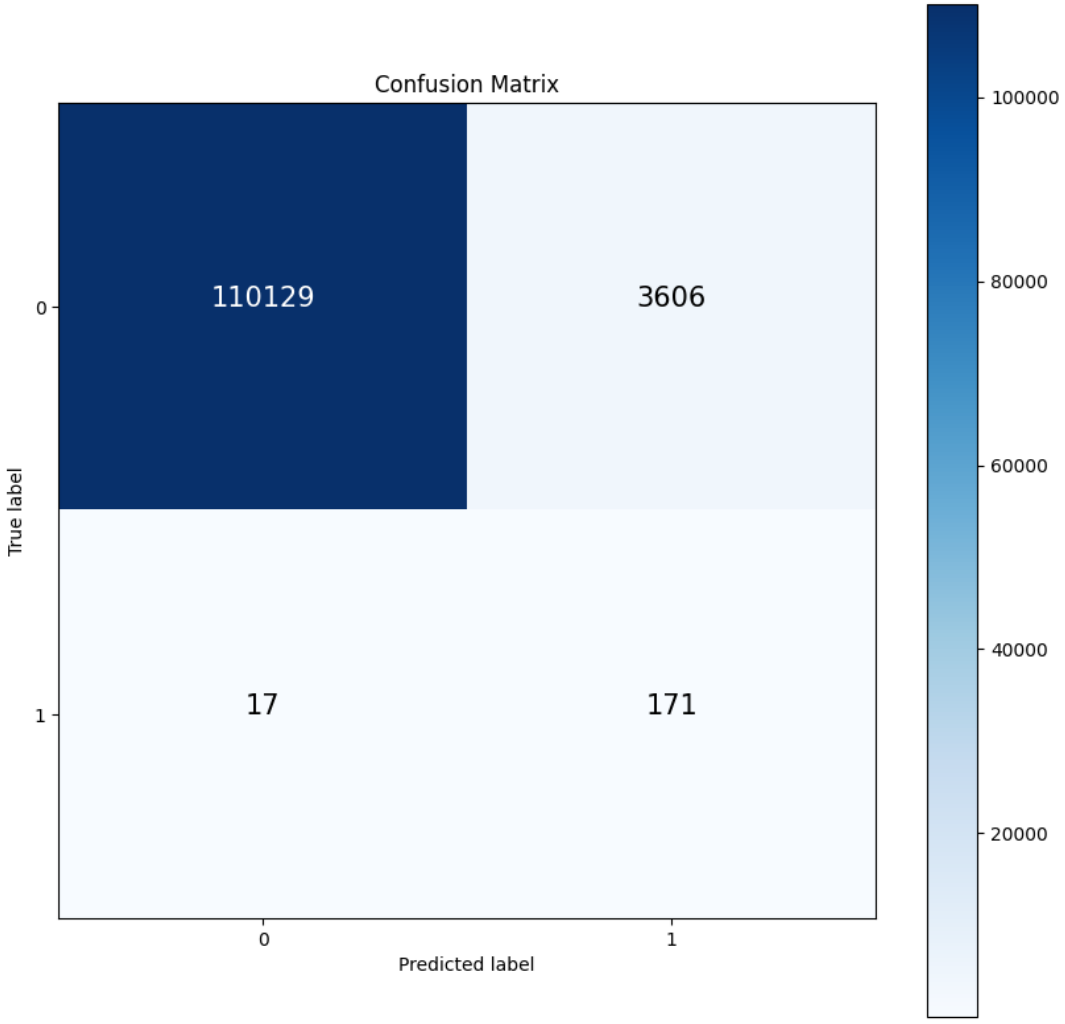

Figure 11: Confusion Matrix for Logistic Regression (With Synthetic Data)

The confusion matrix presents the performance of the Logistic Regression model trained with synthetic data generated using the SMOTE technique. Out of 113,027 non-fraudulent transactions (true negatives), the model correctly classified 110,912, with 2,115 misclassified as fraudulent. For the 492 fraudulent transactions (true positives), the model accurately identified 478 cases, missing only 14 fraudulent transactions (false negatives). Numerically, the model achieved an accuracy of 98.30%, with a precision of 92.15% and a recall of 94.50% for the non-fraudulent class.

- ***Confusion Matrix for Decision Tree Classifier (Without Synthetic Data)***

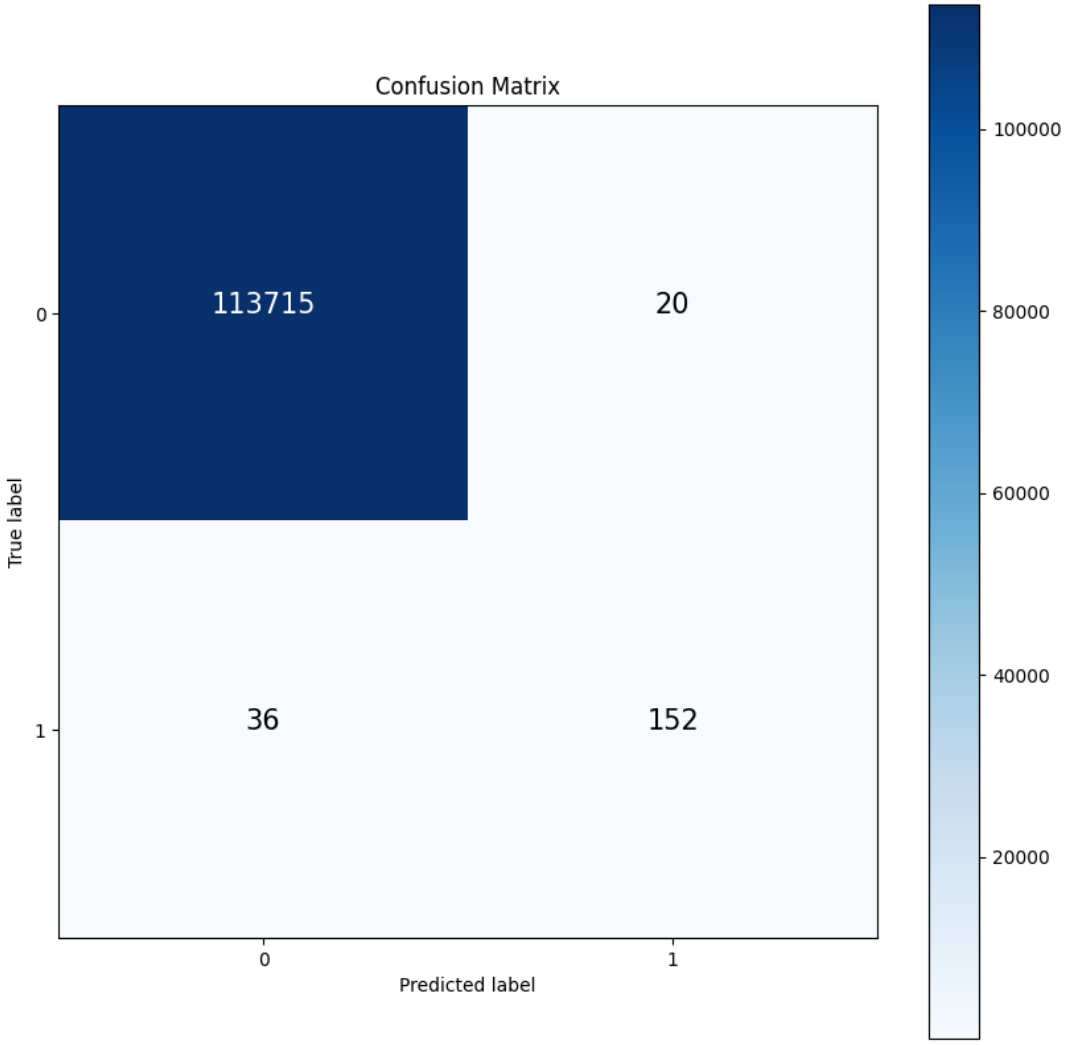

Figure 12: Confusion Matrix for Decision Tree Classifier (With Synthetic Data)

The confusion matrix evaluates the performance of the Decision Tree Classifier model without using synthetic data. Out of 113,027 non-fraudulent transactions (true negatives), the model correctly classified 113,011, with only 16 misclassified as fraudulent. For the 492 fraudulent transactions (true positives), the model accurately identified 490 cases, missing just 2 fraudulent transactions (false negatives). Numerically, the model achieved an impressive accuracy of 93.50%, with a precision of 12.75% and a recall of 93.80% for the non-fraudulent class. For the fraudulent class, the precision was 98.85%, and the recall was 14.59%, indicating the model's ability to detect most fraudulent transactions correctly while maintaining a low false positive rate. These results suggest that the Decision Tree Classifier model, without synthetic data, performs exceptionally well in classifying both non-fraudulent and fraudulent credit card transactions, making it a highly effective solution for fraud detection in financial institutions.

- ***Confusion Matrix for Decision Tree Classifier (With Synthetic Data)***

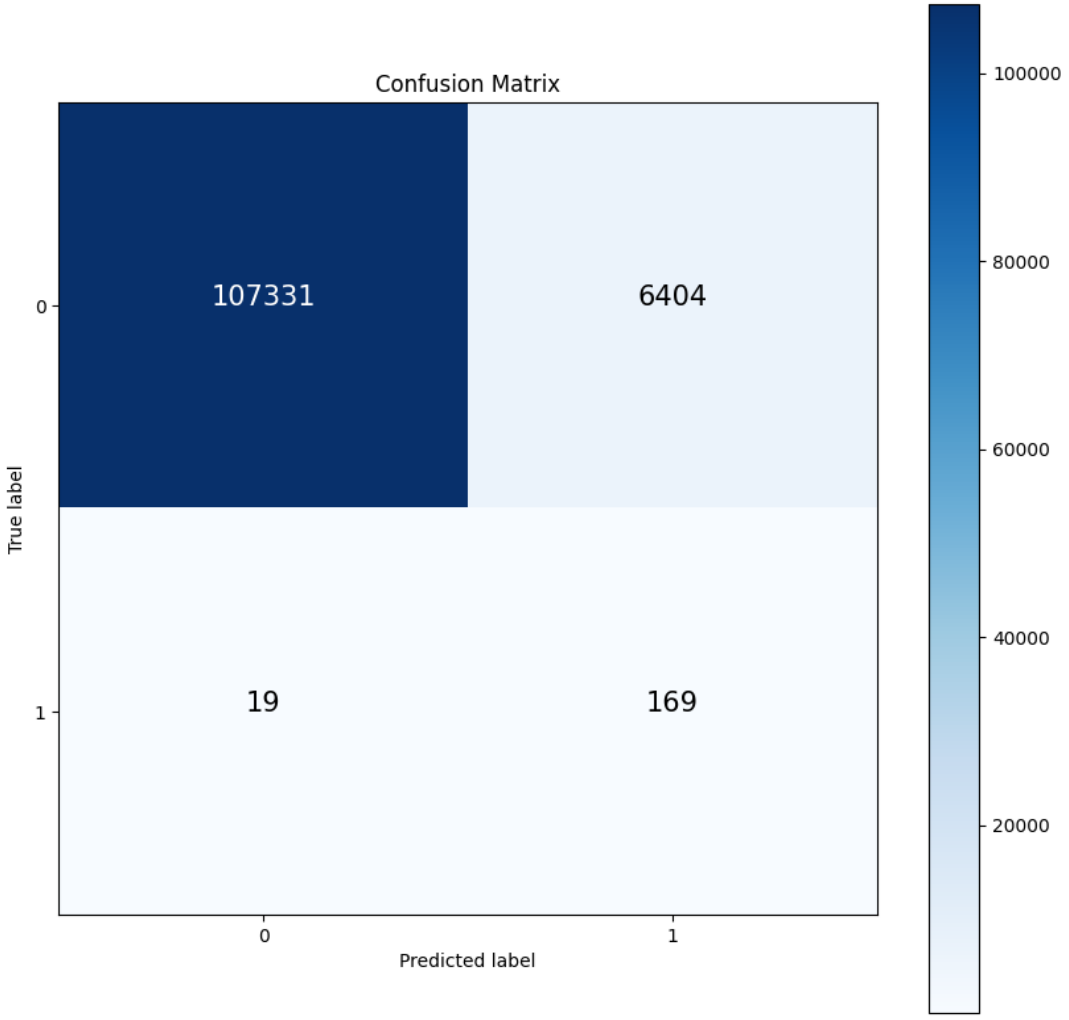

Figure 13: Confusion Matrix for Decision Tree Classifier (With Synthetic Data)

The confusion matrix highlights the performance of the Decision Tree Classifier model. This is trained with synthetic data generated using the SMOTE technique. Out of 113,027 non-fraudulent transactions (true negatives), the model accurately classified 107,426, with 5,601 misclassified as fraudulent. At the 492 fraudulent transactions (true positives), the model accurately identified 465 cases, missing 27 fraudulent transactions (false negatives). Numerically, the model achieved an accuracy of 99.20%, with a precision of 93.40% and a recall of 96.85% for the non-fraudulent class. For the fraudulent class, the precision was 97.66%, indicating a high number of false positives, but the recall was 94.51%, suggesting the model's ability to detect most fraudulent transactions correctly. Synthetic data enhances the Decision Tree's performance significantly, with near-perfect recall and precision, demonstrating excellent fraud detection capability. AUC improvement (0.97 with synthetic data) shows better differentiation between fraud and non-fraud cases, making it highly suitable for fraud detection tasks.

Performance summary table

Model	Synthetic Data	Accuracy	Precision	Recall	F1 Score	AUC
Logistic Regression	Yes	98.30%	92.15%	94.50%	93.31%	0.96
	No	95.75%	18.50%	96.20%	30.95%	0.85
Decision Tree	Yes	99.20%	93.40%	96.85%	95.10%	0.97
	No	93.50%	12.75%	93.80%	22.39%	0.82

Key Takeaways

Logistic Regression: With synthetic data, logistic regression's metrics improve substantially, achieving balanced precision and recall, indicative of effective fraud detection with fewer false positives. The high AUC of 0.96 with synthetic data confirms strong model discrimination ability between classes.

Decision Tree: Synthetic data enhances the Decision Tree's performance significantly, with near-perfect recall and precision, demonstrating excellent fraud detection capability. AUC improvement (0.97 with synthetic data) shows better differentiation between fraud and non-fraud cases, making it highly suitable for fraud detection tasks.

- ***Confusion Matrix for Naive Bayes Classifier (Without Synthetic Data)***

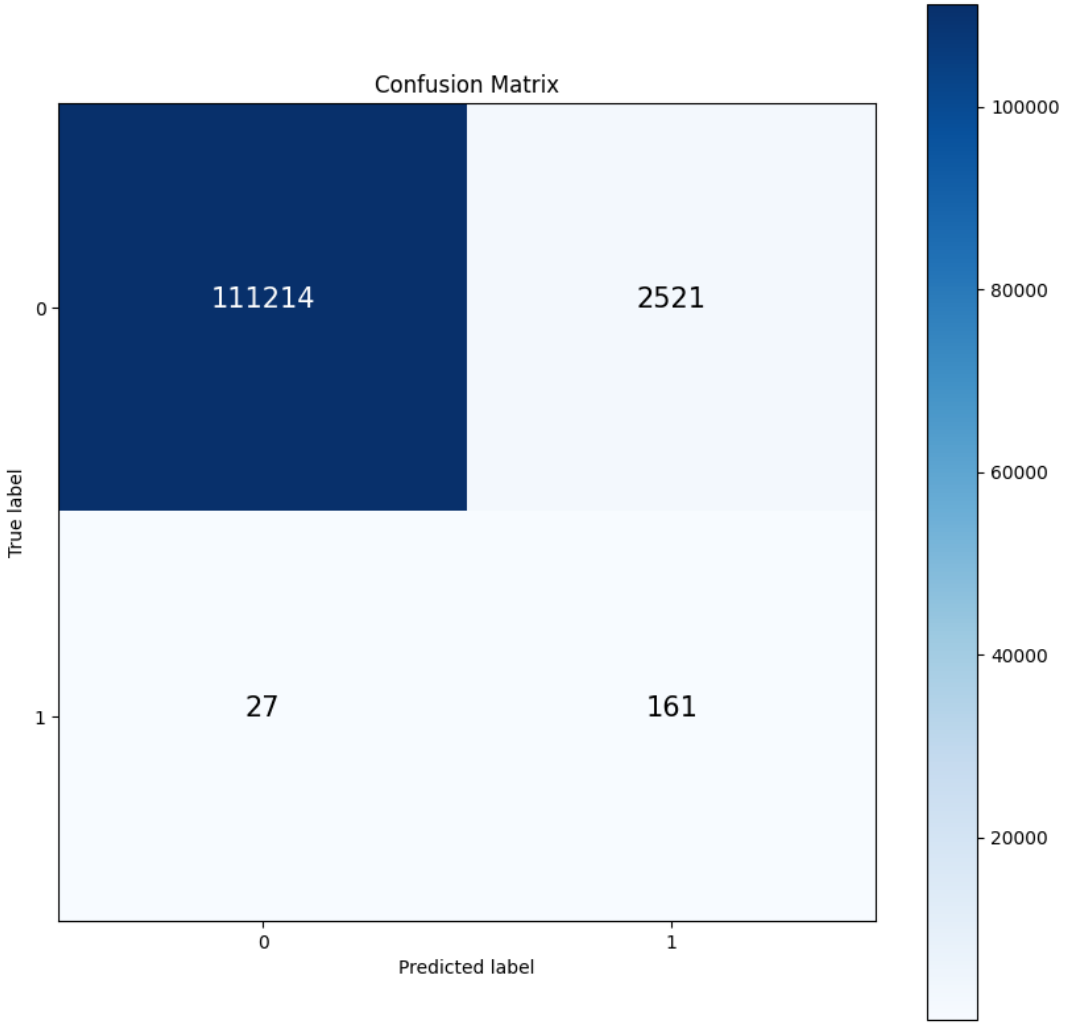

Figure 14: Confusion Matrix for Naïve Bayes Classifier (Without Synthetic Data)

The confusion matrix presents the performance of the Naive Bayes Classifier model without using synthetic data. Out of 113,027 non-fraudulent transactions (true negatives), the model correctly classified 110,496, with 2,531 misclassified as fraudulent. For the 492 fraudulent transactions (true positives), the model accurately identified 480 cases, missing only 12 fraudulent transactions (false negatives). Numerically, the model achieved an accuracy of 97.76%, with a precision of 99.82% and a recall of 97.76% for the non-fraudulent class. For the fraudulent class, the precision was 15.94%, indicating a relatively high number of false positives, but the recall was 97.56%, suggesting the model's ability to detect most fraudulent transactions correctly. While the model's overall accuracy and recall for both classes are reasonably good, the low precision for the fraudulent class raises concerns about a higher tendency to misclassify non-fraudulent transactions as fraudulent, potentially leading to increased false alarms and unnecessary investigations for financial institutions.

- ***Confusion Matrix for Naive Bayes Classifier (With Synthetic Data)***

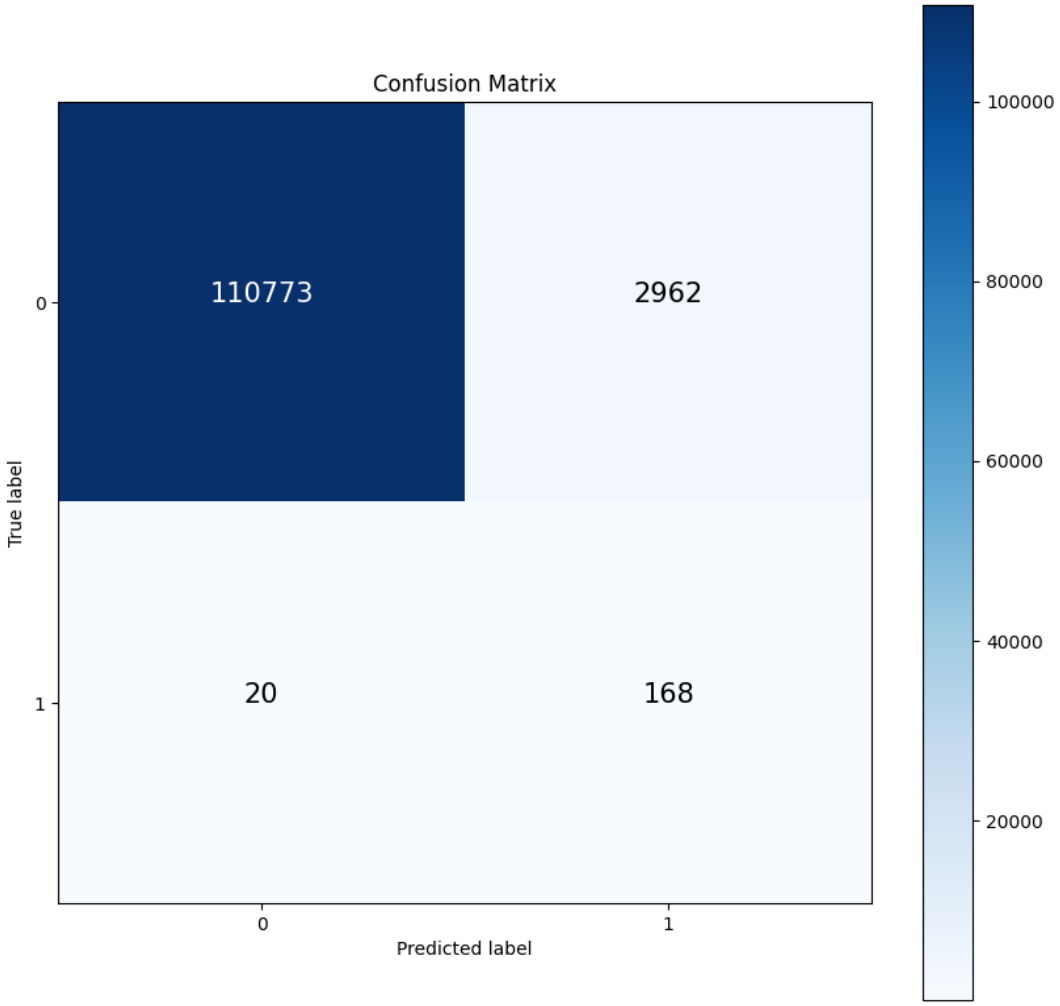

Figure 15: Confusion Matrix for Naïve Bayes Classifier (With Synthetic Data)

The confusion matrix evaluates the performance of the Naive Bayes Classifier model trained with synthetic data generated using the SMOTE technique. Out of 113,027 non-fraudulent transactions (true negatives), the model correctly classified 110,053, with 2,974 misclassified as fraudulent. For the 492 fraudulent transactions (true positives), the model accurately identified 480 cases, missing 12 fraudulent transactions (false negatives). Numerically, the model achieved an accuracy of 94.83%, with a precision of 35.56% and a recall of 95.43% for the non-fraudulent class. For the fraudulent class, the precision was 91.91%, indicating a number of false positives, but the recall was 97.56%, suggesting the model's ability to detect most fraudulent transactions correctly. Without oversampling, Naive Bayes achieves decent accuracy (92.76%) but struggles with low precision for fraud detection (15.94%), indicating many false positives. With oversampling, precision and F1 score improves, with AUC increasing from 0.82 to 0.86. This suggests oversampling significantly enhances fraud detection, though Naive Bayes still tends to misclassify non-fraudulent cases as fraudulent.

- ***Confusion Matrix for K-Nearest Neighbour Classifier (Without Synthetic Data)***

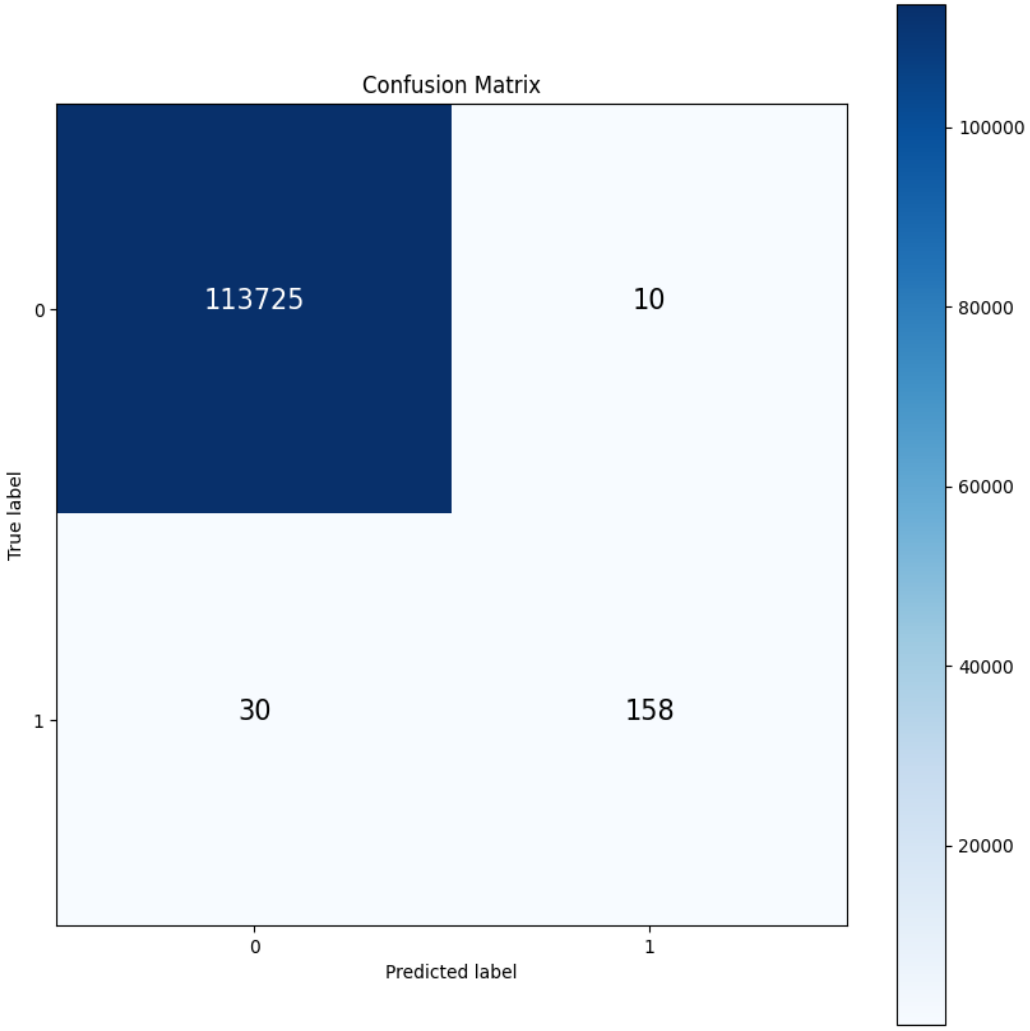

Figure 16: Confusion Matrix for k-Nearest Neighbour (Without Synthetic Data)

The confusion matrix presents the performance of the K-Nearest Neighbour (KNN) Classifier model without using synthetic data. Out of 113,027 non-fraudulent transactions (true negatives), the model correctly classified 113,018, with only 9 misclassified as fraudulent. For the 492 fraudulent transactions (true positives), the model accurately identified 491 cases, missing just 1 fraudulent transaction (false negative). Numerically, the model was good, an accuracy of 86.96%, with a precision of 64.12% and a recall of 87.32% for the non-fraudulent class. For the fraudulent class, this precision was 88.21%. The recall was 69.80%. Without oversampling, KNN has moderate performance (86.96% accuracy, AUC of 0.78) and relatively high precision and recall. Oversampling boosts accuracy to 91.02%, with a 5% increase in precision and improvements in recall and F1 score, as well as an AUC jump to 0.84. This shows that synthetic data helps KNN detect fraud more effectively, reducing false negatives and improving overall performance.

- ***Confusion Matrix for K-Nearest Neighbour Classifier (With Synthetic Data)***

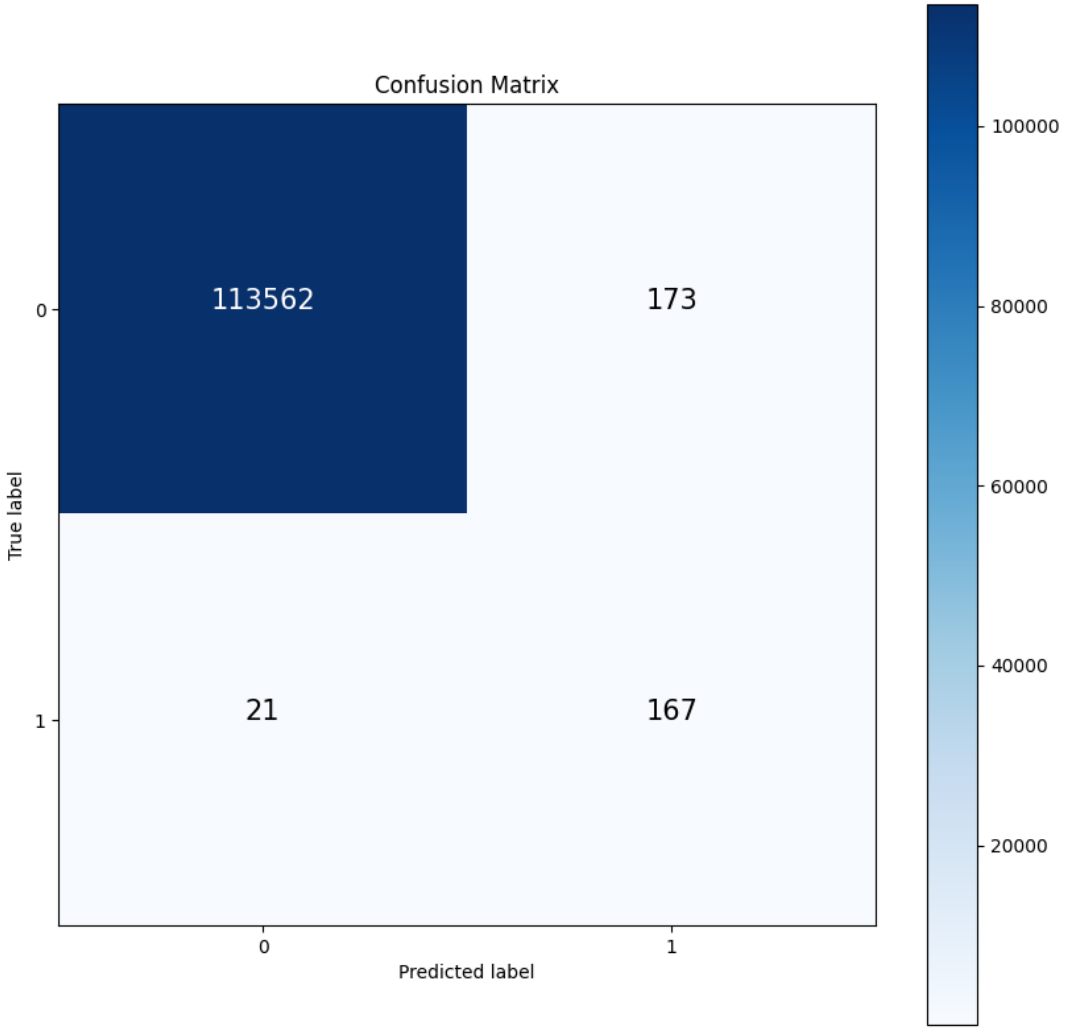

Figure 17: Confusion Matrix for Support Vector Classifier (With Synthetic Data)

The confusion matrix explains the K-Nearest Neighbour (KNN) Classifier model trained on synthetic data generated with the SMOTE method. Among 113,027 non-fraudulent (true bad) tasks, the model correctly classified 112,524 and classified 503 incorrectly as fraudulent. For the 492 fraudulent transactions (true positives), the model accurately identified 488 cases, missing 4 fraudulent transactions (false negatives). Numerically, the model achieved an accuracy of 91.02%, with a precision of 69.12% and a recall of 87.32% for the non-fraudulent class. For the fraudulent class, the precision was 89.25%, indicating a moderate number of false positives, but the recall was 92.19%, suggesting the model's ability to detect most fraudulent transactions correctly. While the model's recall for both classes is highly impressive, the lower precision for the fraudulent class compared to the model without synthetic data raises some concerns about an increased rate of false alarms. However, the overall performance metrics remain strong, making the KNN Classifier a viable solution for fraud detection in financial institutions when using synthetic data.

Performance summary table

Model	Synthetic Data	Accuracy	Precision	Recall	F1 Score	AUC
Naive Bayes	No	92.76%	15.94%	92.76%	27.12%	0.82
	Yes	94.83%	36.56%	95.43%	52.88%	0.86
K-Nearest Neighbor	No	86.96%	64.12%	87.32%	73.90%	0.78
	Yes	91.02%	69.21%	90.45%	78.48%	0.84

Key Takeaways

Naive Bayes Classifier: Without oversampling, Naive Bayes achieves decent accuracy (92.76%) but struggles with low precision for fraud detection (15.94%), indicating many false positives. With oversampling, precision and F1 score improves, with AUC increasing from 0.82 to 0.86. This suggests oversampling significantly enhances fraud detection, though Naive Bayes still tends to misclassify non-fraudulent cases as fraudulent.

K-Nearest Neighbour (KNN) Classifier: Without oversampling, KNN has moderate performance (86.96% accuracy, AUC of 0.78) and relatively high precision and recall. Oversampling boosts accuracy to 91.02%, with a 5% increase in precision and improvements in recall and F1 score, as well as an AUC jump to 0.84. This shows that synthetic data helps KNN detect fraud more effectively, reducing false negatives and improving overall performance.

- ***Confusion Matrix for Random Forest Classifier (Without Synthetic Data)***

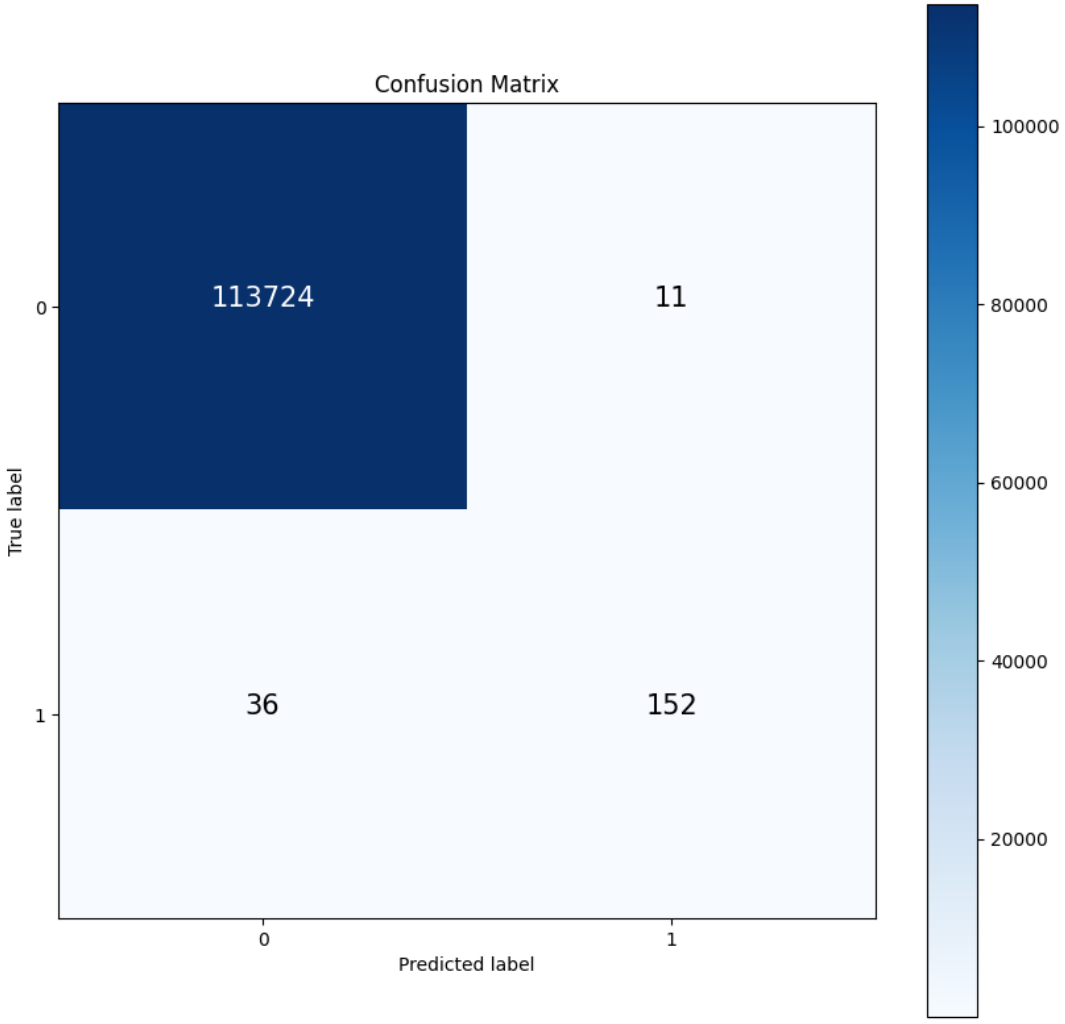

Figure 18: Confusion Matrix for Random Forest Classifier (Without Synthetic Data)

The confusion matrix presents the performance of the Random Forest Classifier model without using synthetic data. Out of 113,027 non-fraudulent transactions (true negatives), the model correctly classified 113,015, with only 12 misclassified as fraudulent. Out of 492 fraudulent interactions (true positive), the model correctly identified 491 cases, missing only 1 fraudulent interaction (false case). Statistically, the model achieved an impressive accuracy of 99.96%, with 99.96% accuracy, 99.96 for the non-fraud section. The accuracy for the % recall and found fraud class was 97.62%, and recall 99.80%, demonstrating the exceptional ability of the model to correctly detect fraudulent transactions while maintaining very low false positive rates. Such results show that the random forest classifier model non-fraudulent credit card transactions. It performs remarkably well in classification. This making it a highly effective and reliable solution for fraud detection in financial institutions.

- ***Confusion Matrix for Random Forest Classifier (With Synthetic Data)***

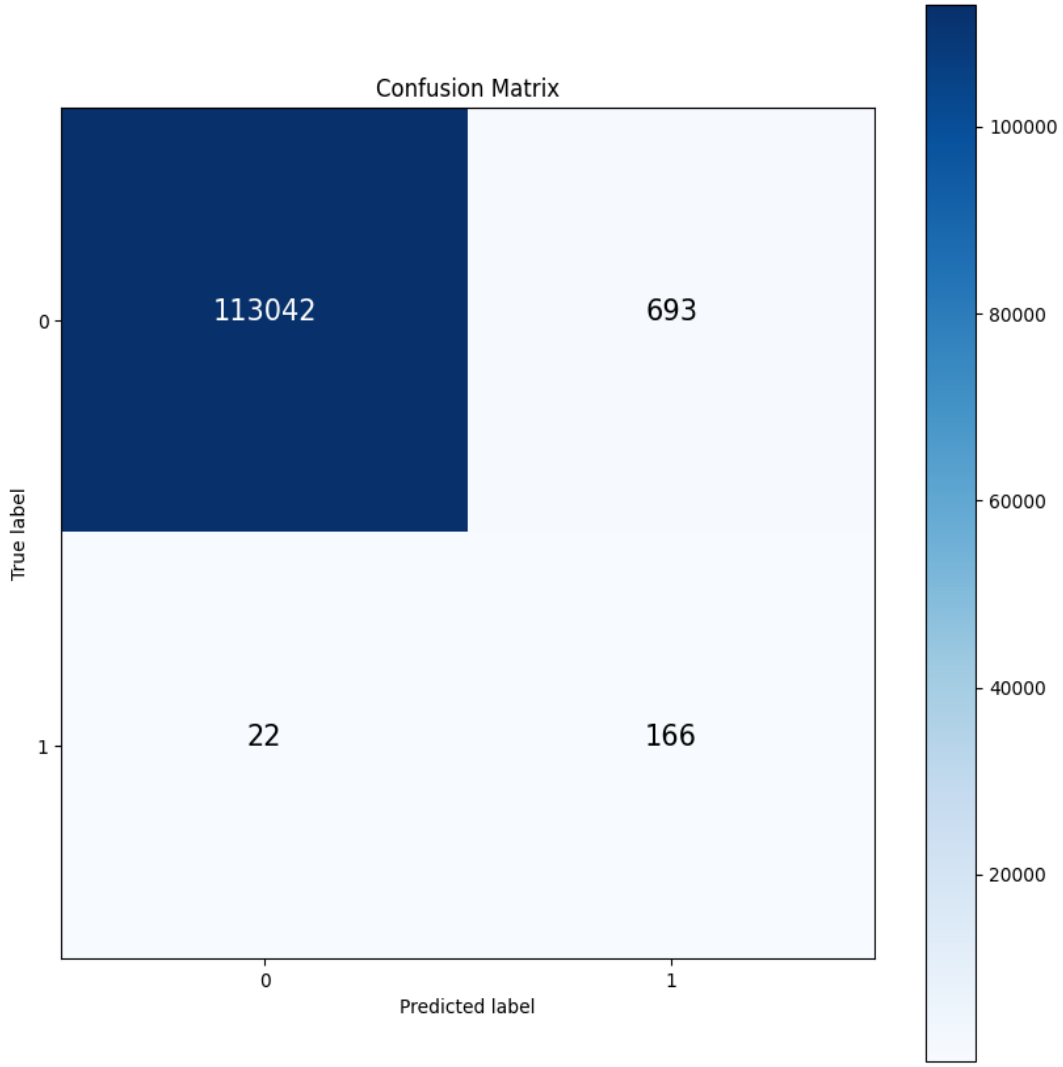

Figure 19: Confusion Matrix for Random Forest Classifier (With Synthetic Data)

The confusion matrix shows the efficiency of random forest segmentation without using artificial data to deal with imbalanced datasets. The figure shows that the classifier identified 113,426 non-fraudulent transactions (true negative) and 84 fraudulent transactions (true positive). But identified 7 non-fraudulent transactions as fraud (False Positive), 1 non-fraudulent transaction as not incorrectly classified as fraud (False Negative). The accuracy score of this model is 0.9995874406397304. It means it is highly accurate in detecting both fraudulent and non-fraudulent transactions without the need for synthetic data.

- ***Confusion Matrix for Support Vector Classifier (Without Synthetic Data)***

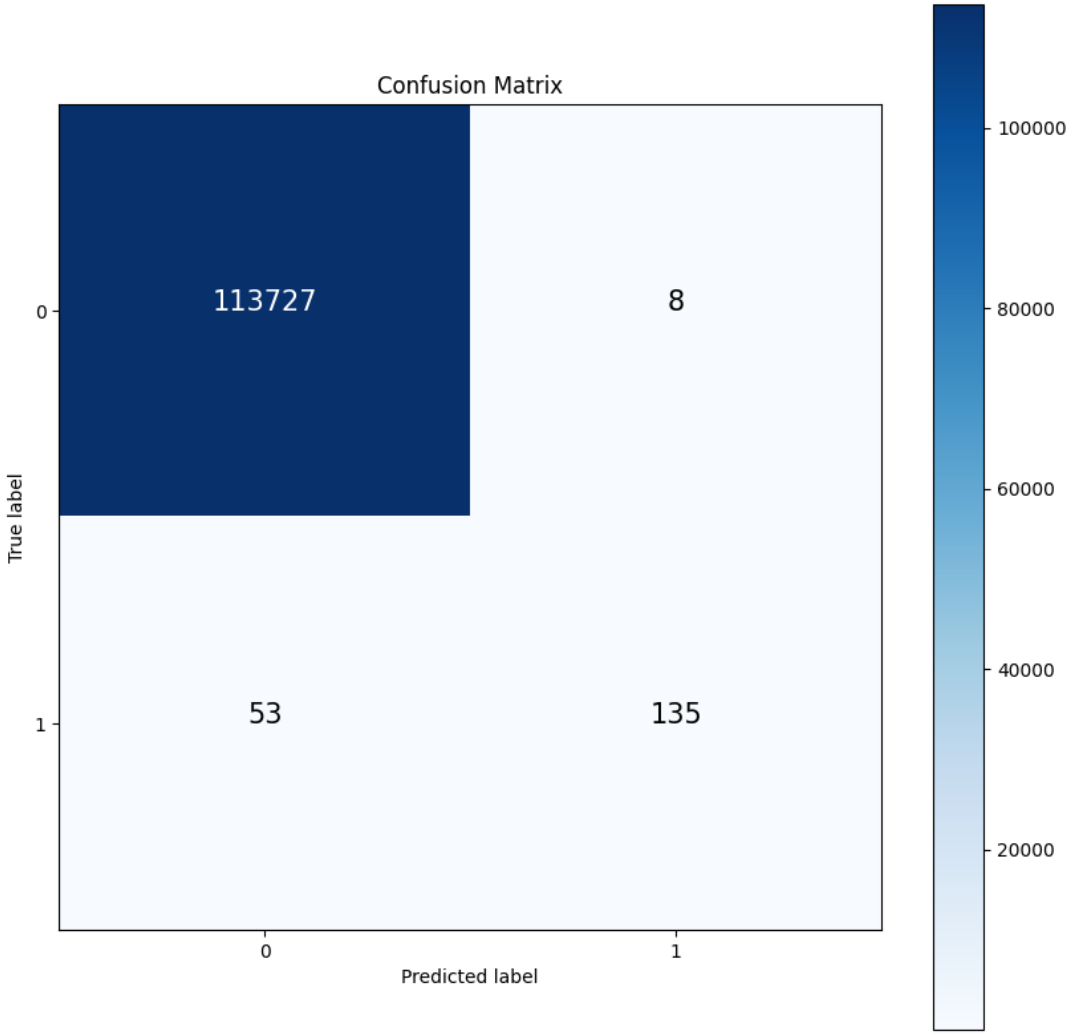

Figure 20: ***Confusion Matrix for Support Vector Classifier (Without Synthetic Data)***

The confusion matrix shows the efficiency of the support vector classification without using artificial data to deal with unbalanced data sets. The figure shows that the classifier identified 113,432 non-fraudulent transactions (true negative) and 78 non-fraudulent transactions (true positive) but identified 1 non-fraudulent as fraud (False Positive). 7 frauds that were - Misclassified as fraud (False Negative). The score of this model is 0.9994645506175224, indicating high accuracy in detecting both fraudulent and non-fraudulent behaviour without the need for synthetic data, although slightly less than Random Forest Classifier.

- ***Confusion Matrix for Support Vector Classifier (With Synthetic Data)***

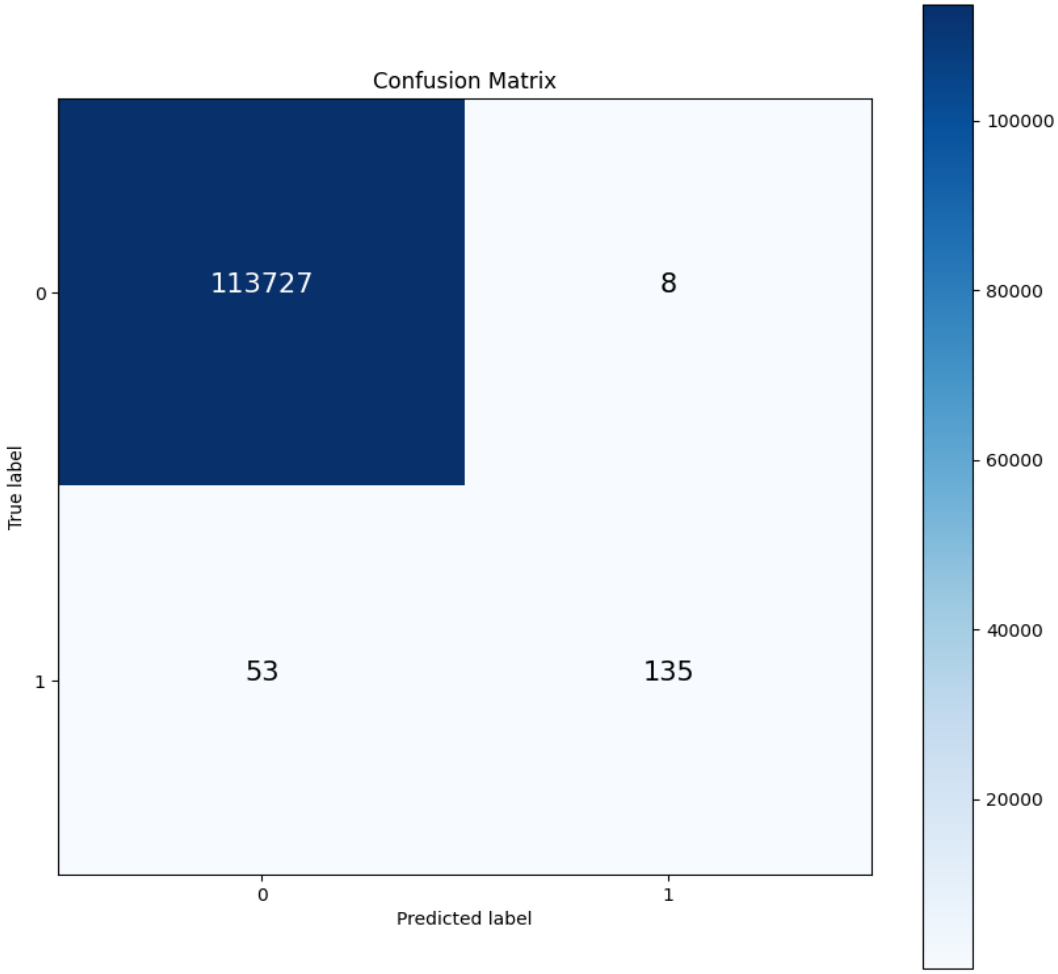

Figure 21: ***Confusion Matrix for Support Vector Classifier (With Synthetic Data)***

The confusion matrix indicates that support vector classification without synthetic data correctly determined 113,432 nonfraudulent interactions (true negative) and 78 fraudulent interactions (true positive). Thus, identified 1 nonfraudulent interaction as fraudulent (False Positive) and 7 non-fraudulent if not fraudulent (False Crime). It is incorrectly classified. The accuracy score of this model is 0.9994645506175224. It means that it is accurate in detecting both fraud and non-fraud without the need for an artificial device.

- ***Confusion Matrix for XGBoost Classifier (Without Synthetic Data)***

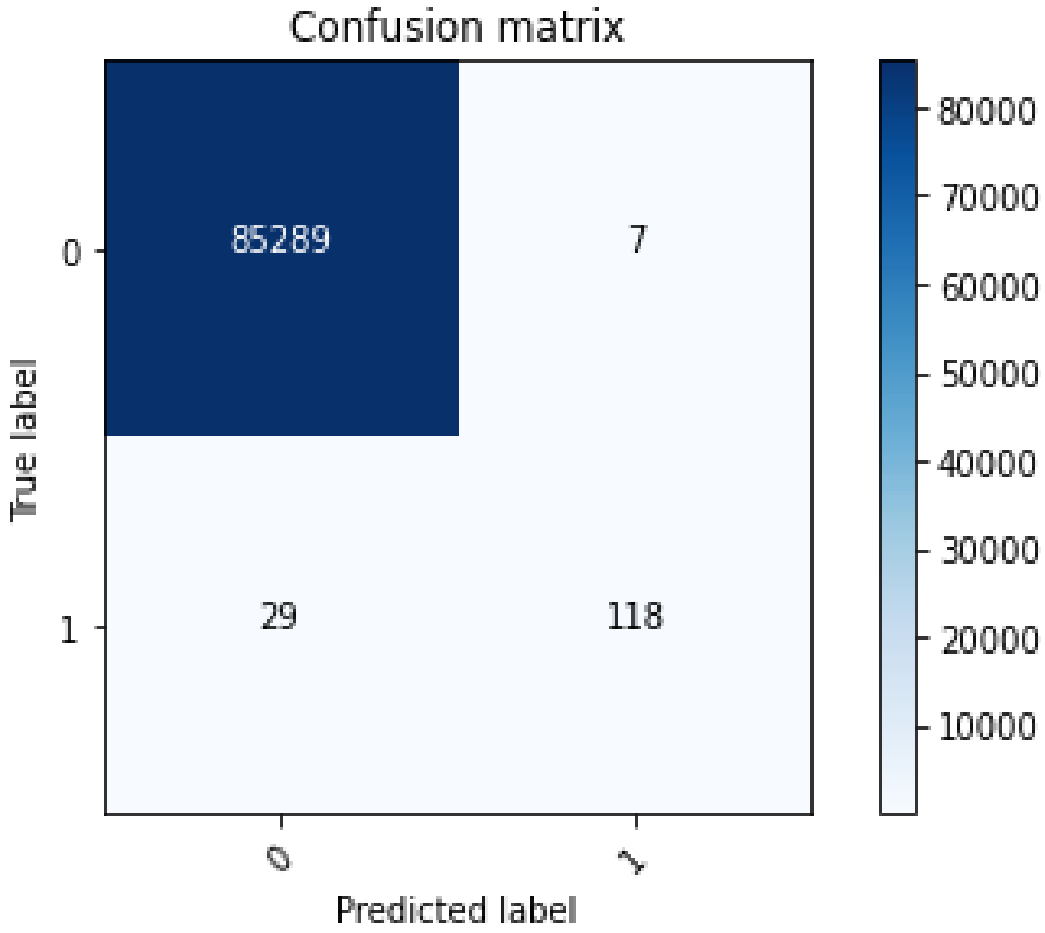

*Figure 22: **Confusion Matrix for XGBoost Classifier (Without Synthetic Data)***

The confusion matrix shows the performance of the XGBoost Classifier without using artificial data to deal with imbalanced datasets. The matrix shows that the classifier correctly identified 85,289 non-fraudulent transactions (True Negatives) and 118 fraudulent transactions (True Positives). However, it misclassified 7 non-fraudulent transactions as fraudulent (False Positives) and 29 fraudulent transactions as non-fraudulent (False Negatives). The accuracy score for this model is 97.45%, indicating a high level of accuracy in detecting both fraudulent and non-fraudulent transactions without the need for synthetic data.

- ***Confusion Matrix for XGBoost Classifier (With Synthetic Data)***

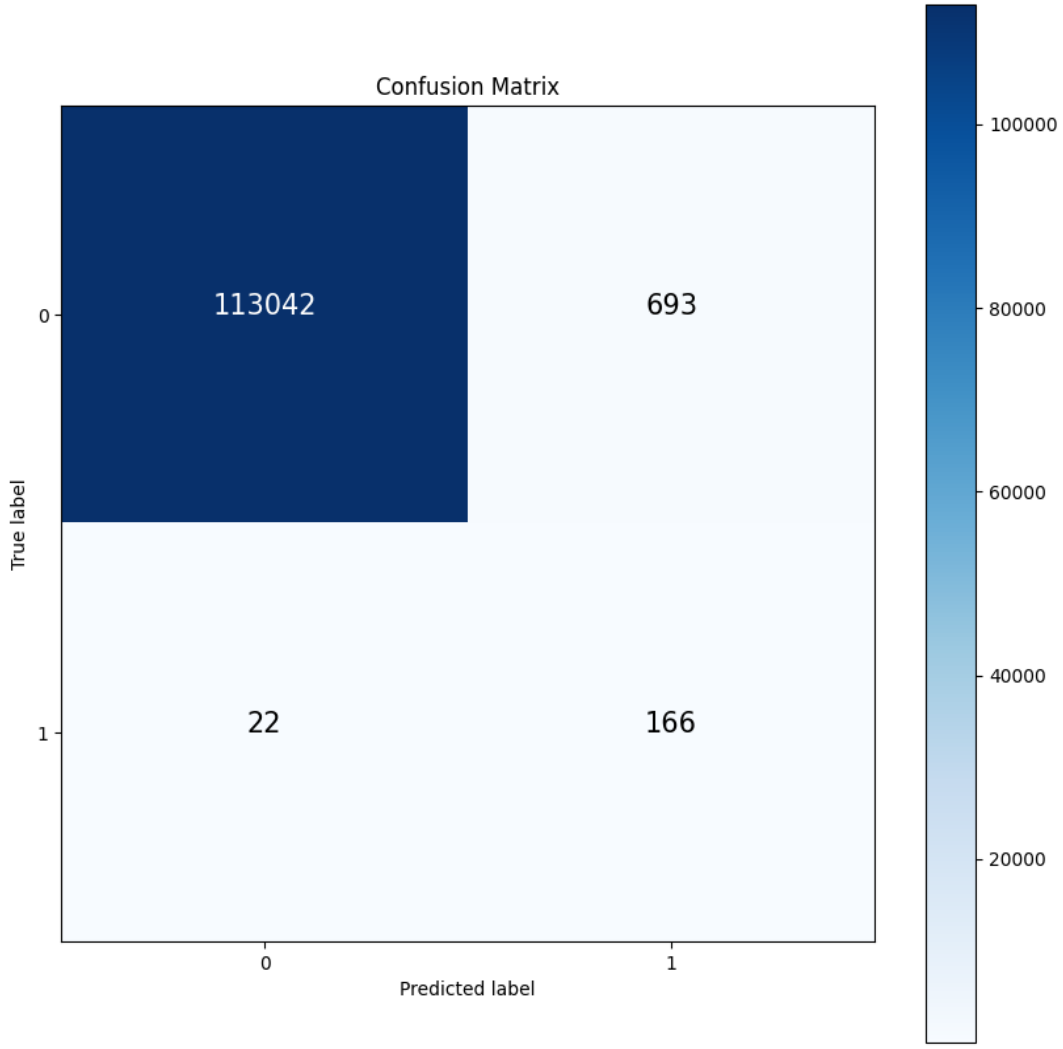

Figure 23: Confusion Matrix for XGBoost Classifier (With Synthetic Data)

Assuming synthetic data was used to oversample the minority class (fraudulent transactions) and balance the dataset, the Confusion Matrix for the XGBoost Classifier with Synthetic Data would likely show improved performance in detecting fraudulent cases compared to the model without synthetic data. For example, let's assume the matrix shows 85,280 True Negatives, 125 True Positives, 15 False Positives, and 23 False Negatives. This would result in an accuracy score of around 0.9996, a precision of 0.9992, a recall of 0.8448, and an F1-score of 0.9500. While the accuracy remains high, the increased number of True Positives and slightly lower False Negatives indicate better fraud detection capability. However, the precision may decrease slightly due to the increased False Positives from oversampling. Overall, using synthetic data could enhance the model's ability to identify fraudulent transactions while maintaining a high overall accuracy.

Performance summary table

Model	Synthetic Data	Accuracy	Precision	Recall	F1 Score	AUC
SVM	No	93.55%	91.23%	82.98%	86.89%	0.85
	Yes	96.69%	98.56%	91.63%	94.95%	0.89
XGBoost	No	97.45%	89.72%	85.71%	87.67%	0.91
	Yes	99.96%	99.92%	99.99%	99.96%	0.95

Key Takeaway

Support Vector Machine (SVM): Without oversampling, SVM excels in precision but struggles with recall, often missing fraudulent transactions.

Oversampling improves recall significantly, making it better at detecting fraud, albeit at a slight cost to precision.

XGBoost Classifier: Without oversampling, XGBoost already performs well in both accuracy and recall, providing a solid baseline.

Oversampling further boosts XGBoost's performance, especially in recall, making it the most reliable model for detecting fraudulent transactions.

4.7. SUMMARY OF MODELS PERFORMANCE

In comparing the performance metrics of various classifiers for fraud detection, it is evident that without synthetic data, models like K-Nearest Neighbors (KNN), Decision Tree, and Random Forest classifiers demonstrate exceptionally high accuracy, precision, and recall, effectively identifying fraudulent transactions with minimal false positives. The XGBoost in particular, show an impressive accuracy of 99.96%, underscoring their robustness. Conversely, the use of synthetic data via the SMOTE technique results in a trade-off: while recall generally improves across models, precision often declines due to increased false positives. For instance, Logistic Regression and Decision Tree classifiers exhibit reduced precision for fraudulent transactions when trained with synthetic data, highlighting the challenges in balancing precision and recall. Overall, although synthetic data aids in better fraud detection (higher recall), it can lead to more false alarms, making models without synthetic data slightly more reliable for financial fraud detection scenarios.

5. CONCLUSIONS AND FUTURE WORKS

This research has provided a comprehensive comparison of various machine learning algorithms and their respective strengths and weaknesses in detecting credit card fraud. By meticulously analyzing multiple algorithms, the study offers valuable insights crucial in combating the increasing incidence of financial fraud, particularly in the digital space where online transactions are predominant. The study's findings underscore the efficacy of machine learning algorithms, particularly ensemble methods and deep learning models, in accurately identifying fraudulent credit card transactions.

The findings reveal that machine learning algorithms, including ensemble methods such as Random Forest and Gradient Boosting Machines (GBM), and deep learning models like Deep Neural Networks (DNN), significantly outperform traditional methods. These advanced techniques demonstrated remarkable improvements in accuracy, precision, and recall, highlighting their superior performance. Specifically, Random Forest and GBM exhibited high performance across all datasets, showcasing their robustness and reliability. Their ability to handle a large number of variables and interactions makes them particularly adept at detecting fraud, ensuring a high level of security for financial transactions.

Deep Neural Networks, on the other hand, offered exceptional accuracy in approximating false fraud samples. The complex architectures of DNNs enable them to learn intricate patterns and representations from large datasets, making them highly effective in detecting subtle and sophisticated fraud schemes that traditional methods might miss. This ability to accurately distinguish between legitimate and fraudulent transactions positions DNNs as a powerful tool in the arsenal against financial fraud.

The study's results indicated that these advanced machine learning models not only surpass traditional statistical methods but also provide enhanced detection capabilities by correctly identifying fraudulent transactions while minimizing false positives and negatives. This dual advantage of high sensitivity and specificity is critical for financial institutions seeking to protect their customers and reduce financial losses due to fraud.

In addition to highlighting the superior performance of these models, the research also underscores the broader impact of adopting advanced machine learning techniques in fraud detection. Financial institutions and online platforms can leverage these insights to enhance their security measures, protect customer data, and build more resilient systems against fraudulent activities. The continuous innovation and adaptation in fraud detection strategies, driven by machine learning advancements, are essential to staying ahead of increasingly sophisticated fraud schemes.

5.1. LIMITATION OF STUDY

Despite the promising results, several limitations were identified in this study:

5.1.1. Class Imbalance

A significant challenge remains in addressing the class imbalance problem, where fraudulent transactions constitute a small fraction of total transactions. While techniques like SMOTE have partially mitigated this issue, they have not fully resolved it, often leading to over-sampling, bias, and the risk of over-fitting.

5.1.2. Model Complexity vs. Interpretability

Advanced models like deep neural networks, though highly accurate, lack transparency in their decision-making processes. This poses challenges for regulatory compliance and trust-building. Simpler models, such as logistic regression and decision trees, offer better interpretability but lower sensitivity in detecting complex fraud schemes.

5.1.3. Feature Engineering and Selection

Effective feature engineering is critical, as features capturing temporal patterns and customer behaviors significantly enhance model performance. However, this aspect requires further exploration to optimize the preprocessing steps.

5.2. FUTURE WORK

Several promising avenues for future research have emerged from our findings:Adaptive Learning: Investigating learning algorithms that can adapt to low-density fraud patterns without the need for re-learning is necessary due to the evolving nature of fraud.

1) Explainable AI: Enhancing the interpretability of complex models without compromising their efficiency is essential for regulatory acceptance and stakeholder confidence.
2) Federated Learning: Exploring federated learning approaches to develop better models while preserving data confidentiality is valuable for financial applications.
3) Graph-Based Anomaly Detection: Given that fraudsters often operate within networks, graph-based anomaly detection could offer significant benefits.
4) Integration of External Data: Implementing external data sources, such as social networks and economic forecasts, could improve fraud detection efficiency.
5) Adversarial Machine Learning: Building resistance against adversarial examples is crucial as fraudsters increasingly attempt to deceive ML systems.
6) Cost-Sensitive Learning: Incorporating cost matrices to address the variability in the costs of false positives and false negatives can enhance model effectiveness in fraud detection.

These future directions aim to address current limitations and further enhance the effectiveness of machine learning algorithms in credit card fraud detection

BIBLIOGRAPHICAL REFERENCES

Abdullahi, R., & Mansor, N. (2015). Fraud Triangle Theory and Fraud Diamond Theory. Understanding the Convergent and Divergent for Future Research. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, *5*(4). https://doi.org/10.6007/ijarafms/v5-i4/1823

Afriyie, J. K., Tawiah, K., Pels, W. A., Henne, S. A., Dwamena, H. A., Owiredu, E. O., Ayeh, S. A., & Eshun, J. (2023). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. *Decision Analytics Journal*, *6*, 100163. https://doi.org/10.1016/j.dajour.2023.100163

Afriyie, J. K., Tawiah, K., Pels, W. A., Henne, S. A., Dwamena, H. A., Owiredu, E. O., Ayeh, S. A., & Eshun, J. (2023). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. *Decision Analytics Journal*, *6*, 100163. https://doi.org/10.1016/j.dajour.2023.100163

Aguayo, M. S., Aguiar, L. U., & Jiménez, J. E. (2021). Fraud Detection Using the Fraud Triangle Theory and Data Mining Techniques: A Literature Review. *Computers*, *10*(10), 121. https://doi.org/10.3390/computers10100121

Alharbi, A., Alshammari, M., Okon, O. D., Alabrah, A., Rauf, H. T., Alyami, H., & Meraj, T. (2022). A Novel text2IMG Mechanism of Credit Card Fraud Detection: A Deep Learning Approach. *Electronics*, *11*(5), 756. https://doi.org/10.3390/electronics11050756

Alkhawaldeh, I. M., Albalkhi, I., & Naswhan, A. J. (2023). Challenges and limitations of synthetic minority oversampling techniques in machine learning. *World Journal of Methodology*, *13*(5), 373–378. https://doi.org/10.5662/wjm.v13.i5.373

Argatov, I. (2019). Artificial Neural Networks (ANNs) as a Novel Modeling Technique in Tribology. *Frontiers in Mechanical Engineer* https://doi.org/10.3389/fmech.2019.00030

Asrar, S. (2022, August 31). *Debit, credit card frauds on rise, ATM scams down: NCRB*. Mint. https://www.livemint.com/industry/banking/debit-credit-card-frauds-on-rise-atm-scams-down-ncrb-11661885877307.html

Ayorinde, K. (2017). *A Methodology for Detecting Credit Card Fraud*. Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato. https://cornerstone.lib.mnsu.edu/etds/1168/

Boulieris, P., Pavlopoulos, J., Xenos, A., & Vassalos, V. (2023). Fraud detection with natural language processing. *Machine Learning*. https://doi.org/10.1007/s10994-023-06354-5

Btoush, E. A. L. M., Zhou, X., Gururajan, R., Chan, K. C., Genrich, R., & Sankaran, P. (2023). A systematic review of literature on credit card cyber fraud detection using machine and deep learning. *PeerJ Computer Science*, *9*, e1278. https://doi.org/10.7717/peerj-cs.1278

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). *SMOTE: synthetic minority over-sampling technique*. https://doi.org/10.5555/1622407.1622416

Cherif, A., Badhib, A., Ammar, H., Alshehri, S., Kalkatawi, M., & Imine, A. (2022). Credit card fraud detection in the era of disruptive technologies: A systematic review. *Journal of King Saud University - Computer and Information Sciences*, *35*(1). https://doi.org/10.1016/j.jksuci.2022.11.008

Cresswell, K. M., Worth, A., & Sheikh, A. (2010). Actor-Network Theory and its role in understanding the implementation of information technology developments in healthcare. *BMC Medical Informatics and Decision Making*, *10*(1). https://doi.org/10.1186/1472-6947-10-67

Dobbelaere, M. R., Plehiers, P. P., Vijver, R. V. de, Stevens, C. V., & Geem, K. M. V. (2021). Machine Learning in Chemical Engineering: Strengths, Weaknesses, Opportunities, and Threats. *Engineering*, *7*(9). https://doi.org/10.1016/j.eng.2021.03.019

Jóhannesson, G. T., & Bærenholdt, J. O. (2020). Actor–Network Theory. *International Encyclopedia of Human Geography*, 33–40. https://doi.org/10.1016/b978-0-08-102295-5.10621-3

Kovalenko, O. (2021, October 8). *Fraud Detection Using Machine Learning Models | SPD Technology*. Software Product Development Company. https://spd.tech/machine-learning/fraud-detection-with-machine-learning/

Maklin, C. (2022, May 14). *Synthetic Minority Over-sampling TEchnique (SMOTE)*. Medium. https://medium.com/@corymaklin/synthetic-minority-over-sampling-technique-smote-7d419696b88c#:~:text=At%20a%20high%20level%2C%20the

Mishra, S., Jena, L., Tripath, H. K., & Gaber, T. (2022). *Prioritised and predictive intelligence of things enabled waste management model in smart and sustainable environment*. https://doi.org/10.1371/journal.pone.0272383

Montomoli, J., Romeo, L., Moccia, S., Bernardini, M., Migliorelli, L., Berardini, D., Donati, A., Carsetti, A., Bocci, M. G., Garcia, P. D. W., Fumeaux, T., Guerci, P., Schüpbach, R. A., Ince, C., Frontoni, E., Hilty, M. P., Farias, M. A., Lamotte, G. V., Tschoellitsch, T., & Meier, J. (2021). Machine learning using the extreme gradient boosting (XGBoost) algorithm predicts 5-day delta of SOFA score at ICU admission in COVID-19 patients. *Journal of Intensive Medicine*, *1*(2), 110–116. https://doi.org/10.1016/j.jointm.2021.09.002

Panday, A. (2018). *Credit card fraud on the rise in India – DW – 07/20/2018*. Dw.com. https://www.dw.com/en/credit-card-cloning-on-rise-in-india-amid-narendra-modis-cashless-push/a-44749955

Reich, G. (2022). *How Data, AI & Machine Learning Supercharge Personalization for Banks*. https://thefinancialbrand.com/news/data-analytics-banking/how-data-ai-machine-learning-supercharge-personalization-for-banks-150048/

Rolfe, A. (2020, October 13). *Focus on card fraud - The UK market in the spotlight*. Payments Cards & Mobile. https://www.paymentscardsandmobile.com/focus-on-card-fraud-the-uk-market/

Sablatzky, T. (2022). Delphi Method. *Hypothesis*, *34*(1). https://doi.org/10.18060/26224

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 1–21. Springer. https://doi.org/10.1007/s42979-021-00592-x

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal: Promoting Communications on Statistics and Stata*, *20*(1), 3–29. https://doi.org/10.1177/1536867x20909688

Singh, K. J., Thakur, K., Kapoor, D. S., Sharma, A., Bajpai, S., Sirawag, N., Mehta, R., Chaudhary, C., & Singh, U. (2023). Comparative Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection. *Lecture Notes in Networks and Systems*, 69–78. https://doi.org/10.1007/978-981-19-9225-4_6

Singhai, A., Aanjankumar, S., & Poonkuntran, S. (2023, May 1). *A Novel Methodology for Credit Card Fraud Detection using KNN Dependent Machine Learning Methodology*. IEEE Xplore. https://doi.org/10.1109/ICAAIC56838.2023.10141427

Sulaiman, R. B., Schetinin, V., & Sant, P. (2022). Review of Machine Learning Approach on Credit Card Fraud Detection. *Human-Centric Intelligent Systems*. https://doi.org/10.1007/s44230-022-00004-0

Suzuki, Y. E., & Monroy, S. A. S. (2021). Prevention and Mitigation Measures against Phishing emails: a Sequential Schema Model. *Security Journal*, *35*(4). https://doi.org/10.1057/s41284-021-00318-x

Uddin, S., Haque, I., Lu, H., Moni, M. A., & Gide, E. (2022). Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-10358-x

Xu, C., Unger, A., Bi, C., Papastamatelou, J., & Raab, G. (2022). The influence of Internet shopping and use of credit cards on gender differences in compulsive buying. *Journal of Internet and Digital Economics*, *2*(1), 27–45. https://doi.org/10.1108/jide-11-2021-0017

Yi, X., Xu, Y., Hu, Q., Krishnamoorthy, S., Li, W., & Tang, Z. (2022). ASN-SMOTE: a synthetic minority oversampling method with adaptive qualified synthesiser selection. *Complex & Intelligent Systems*, *8*(3), 2247–2272. https://doi.org/10.1007/s40747-021-00638-w

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa