Optimizing Credit Card Fraud Detection: A Study of Preprocessing Techniques and Model Performance

Sachin Kumar Soni[1],Dr.Balveer Singh[2]

Department of Computer Science[1,2], P.K. University, Shivpuri, M.P. India.

sachinkumarsoni185@gmail.com, adm.pkit@gmail.com

Abstract: Credit card fraud detection is a critical task for financial institutions to safeguard against fraudulent activities and protect customers from unauthorized transactions. In this study, we explore various preprocessing techniques and machine learning models to improve the accuracy of fraud detection algorithms. We address the imbalance in our dataset using Synthetic Minority Over-sampling Technique (SMOTE) and evaluate the performance of models trained on both oversampled and undersampled data. Our analysis reveals that while SMOTE helps mitigate label imbalance, neural network models trained on oversampled data sometimes exhibit lower accuracy in predicting fraud transactions compared to models trained on undersampled data. Additionally, we find that undersampled models struggle to accurately classify non-fraud transactions, potentially leading to customer dissatisfaction and increased complaints. Moving forward, we propose implementing outlier removal techniques on the oversampled dataset to refine model performance. Our study underscores the importance of balancing precision and recall in fraud detection models and highlights the ongoing need for refinement and evaluation of detection methods to ensure optimal performance and customer satisfaction in the financial sector.Logistic regression outperformed other classifiers with 0.95 precision, 0.94 recall, and 0.94 F1-score. However, oversampled deep learning models occasionally showed lower fraud prediction accuracy compared to undersampled ones. Undersampling struggled with non-fraud classification, necessitating further preprocessing refinement for robust fraud detection.

Keywords:Fraud detection, Preprocessing techniques, Machine learning models, Imbalanced dataset, Logistic regression, Deep learning models, etc;

1. INTRODUCTION

A credit card is a type of payment card that can be offered to users (cardholders) by financial institutions so that the cardholder can make payments to merchants for the purchase of goods and services using the cardholder's available credit rather than cash. This allows the cardholder more purchasing power than they would have with cash. This provides the cardholder with increased purchasing power compared to what they would have if they had only cash. The card issuer, which is typically a bank or credit union, will open a revolving account for the cardholder and also make a line of credit accessible to the cardholder. In addition, the cardholder will have access to the line of credit. The cardholder will have access to a line of credit that will enable them to borrow money that can either be used for a cash advance or to make a payment to a

merchant. This money can be utilized for either purpose. The entirety of the market for credit cards may be broken down into two basic subsets: consumer credit cards and business credit cards. Neither of these subsets stands alone. Consumer credit cards incorporate both of these subsets in their overall design. However, some playing cards are made of metal (such as stainless steel, gold, palladium, or titanium), and some playing cards are made of metal with gemstones encrusted on them. Plastic is the material that is used to produce the great majority of playing cards; however, some playing cards are made of metal with gemstones encrusted on them. A charge card is not the same as a standard credit card since a charge card requires the debt to be paid off in full at the end of each billing cycle or every month. A regular credit card does not have this requirement. This condition is not attached to a standard credit card in any way[1]–[5].

This criterion is in no way connected to a normal credit card in any manner, shape, or form. A typical credit card does not in any way come with the stipulation that was just mentioned. On the other hand, credit cards allow their users to roll over a balance from one billing cycle to the next. This leads in the buildup of interest charges over the term of the card's life because the balance is carried over. While charge cards only postpone the buyer's need to pay until a later date, credit cards generally involve a third party that pays the seller and is reimbursed by the buyer, whereas credit cards only postpone the buyer's obligation to pay until a later date. Charge cards are also referred to as "debit" cards in some circles. When a customer pays with a credit card, however, there is often a third party involved that is responsible for paying the vendor. This third party is subsequently refunded by the buyer. Charge cards and credit cards can be differentiated from one another in this additional method. One of the most difficult challenges that customers face in the modern day is overcoming the increasingly sophisticated problem of credit card fraud. As a direct consequence of the growing number of people who use the internet, card fraud has quickly become a problem that affects a wide variety of people. Our examination into whether or not it is possible to recognise fraudulent activity on credit cards makes use of a technique known as deep learning[6]–[10].

The purpose of this study is to determine whether or not it is possible to do so. During the course of our investigation, we will establish which model is the most accurate for projecting fraudulent behaviour so that we can use that model moving forward. The use of credit cards in fraudulent activities is a practise that is still in its infancy but has recently seen an uptick in the number of reported incidents. One way that can be used to help solve the problem of detecting fraudulent use of credit cards is the modelling of prior credit card transactions by making use of the information of fraudulent ones. This is one method that can be used to help address the problem. The use of online platforms for the purpose of conducting financial transactions is by far the most prevalent way in which individuals become victims of fraud. This is a direct consequence of the considerable amount of time that is spent on the internet, in particular on online platforms. The majority of fraudulent financial transactions now take place through this primary channel. On the other hand, it is already general information that the overall volume of transactions

carried out over the internet, both those that are legal and those that are not, is expanding at an alarming rate. This is true for both legal and illegal transactions.

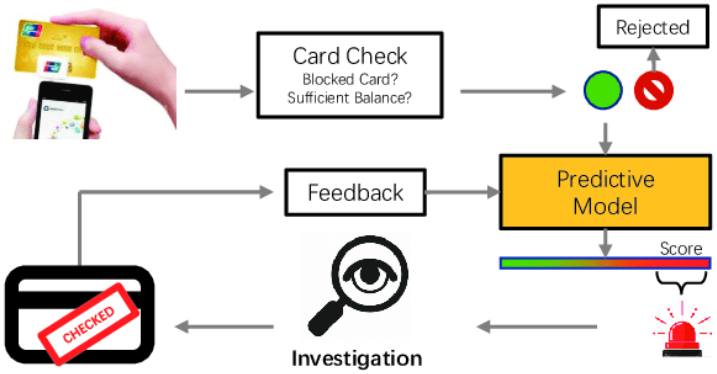

Figure 1 Credit card fraud detection

Shopping on the Web, Credit card use is prevalent because of its convenience and because of the increased globalisation of the economy. Credit card transactions account for a considerable portion of its revenue. However, as the total number of credit card transactions (commonly abbreviated as CCTs) has increased, so too has the amount of fraudulent activity, calling for the creation of new methods for spotting such instances. In order to get an unfair advantage, some people will tell the truth while breaching the law. When a person steals another person's identity with the intention of making fraudulent purchases with that person's credit card, this is known as credit card crime (CCC). Credit card fraud detection (CCFD) processes are used for this purpose. Although both online and offline fraud are widespread, they are distinct crimes. Offline fraud, on the other hand, occurs when a card is taken and then used in an actual purchase. Online fraud is committed by thieves by stealing the victim's personal information such as their name, card number, and PIN. When a card is stolen in person and then used in person after being stolen, this is an example of offline fraud. Because conventional transactions occur more frequently than fraudulent transactions, telling them apart can be a challenging undertaking[11].

For this reason, regardless of the FIM employed, it is essential that fraudulent transactions be uncovered first. Numerous investigations into CCF, employing methods like data mining (DM) and machine learning (ML), have been carried out to expose fraudulent activities. Based on the results of these inquiries, two main categories of methods have emerged for spotting dishonest financial dealings. Both unsupervised and supervised approaches are available. In a supervised method, the transactional data record serves as the basis for an algorithm's categorization process. Support vector machines (SVMs), artificial neural networks (ANNs), k-nearest neighbours (KNNs), random forests (RFs), and Bayesian belief networks (BBNs) are all state-of-the-art examples of supervised learning algorithms.

1.1 Motivation

Credit card fraud detection using deep learning is motivated by the need to enhance security and protect consumers and financial institutions from fraudulent activities[12]. Here are some key motivations:

- Rising Instances of Fraud: With the increasing use of credit cards for transactions, there's a parallel rise in fraud attempts. Deep learning offers sophisticated methods to detect patterns and anomalies in large volumes of data, which can help identify fraudulent activities accurately and swiftly.
- Complexity of Fraud Patterns: Traditional rule-based systems can struggle to keep up with the evolving tactics used by fraudsters. Deep learning models, with their ability to learn complex patterns and relationships within data, offer a more dynamic approach to identifying fraudulent behavior. They can adapt and evolve to recognize new and previously unseen fraud patterns.
- Real-time Detection: Deep learning models can analyze transactions in real-time, allowing for immediate identification of potential fraudulent activities. This speed is crucial in preventing further fraudulent transactions and minimizing financial losses.
- Reducing False Positives: Deep learning models can improve accuracy in distinguishing between legitimate transactions and fraudulent ones, reducing false positives. This accuracy helps prevent inconvenience to customers who may have their legitimate transactions flagged incorrectly.
- Handling Big Data: Credit card transactions generate massive amounts of data. Deep learning algorithms excel at handling big data, processing it efficiently to identify intricate patterns that might indicate fraudulent behavior.
- Continuous Learning and Adaptation: Deep learning models can continuously learn from new data, adapting and improving their accuracy over time. This adaptability is crucial in combating the ever-evolving tactics of fraudsters.
- Cost Efficiency: While initial setup and training of deep learning models might require resources, once deployed, these systems can significantly reduce costs associated with fraudulent transactions by preventing them before they occur[13]–[15].

1.2 Research contribution

Research contributions in credit card fraud detection using deep learning have been significant and continue to evolve. Some key contributions include:

- Improved Accuracy and Efficiency: Research has focused on developing deep learning architectures that enhance the accuracy and efficiency of fraud detection systems. Novel neural network structures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and more advanced models like transformers or graph neural networks, have been explored to better capture intricate patterns in credit card transaction data.

- Feature Representation Learning: Deep learning models have been used to automatically learn relevant features from raw transaction data. This eliminates the need for manual feature engineering, allowing the models to uncover hidden patterns and relationships that might signify fraudulent behavior.
- Anomaly Detection: Deep learning models excel at anomaly detection, and research has focused on leveraging this ability to identify fraudulent transactions. Unsupervised learning techniques, such as autoencoders, have been employed to detect anomalies in transaction sequences or feature distributions, thereby flagging potentially fraudulent activities.
- Handling Imbalanced Data: Credit card fraud detection datasets often suffer from imbalanced classes, with legitimate transactions significantly outnumbering fraudulent ones. Research has delved into techniques like oversampling, undersampling, and cost-sensitive learning within deep learning frameworks to handle this class imbalance issue effectively.
- Adversarial Attacks and Robustness: Researchers have explored the vulnerability of deep learning models to adversarial attacks in the context of fraud detection. They've developed methods to enhance the robustness of models against adversarial examples, ensuring the reliability of the fraud detection system.
- Explainability and Interpretability: Enhancing the interpretability of deep learning models in fraud detection has been a focus. Efforts to explain model predictions and decisions aid in understanding the rationale behind fraud classifications, increasing trust and usability.
- Real-time Processing and Scalability: Research has aimed at developing deep learning architectures that can handle real-time processing of credit card transactions efficiently. Scalable models that can analyze large volumes of data swiftly are crucial in preventing fraudulent transactions promptly.

2. CREDIT CARD FRAUD DETECTION

Credit card fraud detection stands as a critical facet of financial security in the digital age, leveraging advanced technological tools, particularly deep learning models. These models, rooted in artificial intelligence, play a pivotal role in sifting through vast troves of transactional data to discern patterns, anomalies, and subtle deviations indicative of fraudulent activities. Their application significantly bolsters traditional rule-based systems by offering a dynamic approach capable of adapting to the ever-evolving tactics employed by fraudsters. These systems excel in their ability to autonomously learn from historical transactional data, allowing them to discern intricate patterns and correlations that might elude conventional methods. They leverage sophisticated neural network architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and more advanced structures like transformers or graph neural networks. These architectures aid in automatically extracting relevant features from raw data, minimizing the reliance on manual feature engineering and enabling the identification of previously unseen fraud patterns[16]–[19].

One of the inherent challenges in credit card fraud detection lies in the class imbalance within datasets, where legitimate transactions vastly outnumber fraudulent ones. Deep learning models address this by employing techniques such as oversampling, under sampling, or cost-sensitive learning to ensure balanced and effective learning. Moreover, they excel in real-time processing, swiftly analyzing transactions as they occur to promptly flag potential fraudulent activities, preventing further financial losses and mitigating risks for both financial institutions and consumers. Enhancing the robustness and interpretability of these models remains a focal point in research. Efforts are directed towards fortifying models against adversarial attacks, ensuring their resilience in the face of deliberate attempts to deceive the system. Moreover, making these models more interpretable aids in understanding the rationale behind fraud classifications, thereby fostering greater trust and comprehension among stakeholders. In essence, credit card fraud detection powered by deep learning not only bolsters the security measures employed by financial institutions but also instills confidence in consumers regarding the safety and reliability of electronic transactions. Its continuous evolution and adaptation remain integral in staying ahead of sophisticated fraudulent tactics, safeguarding financial systems and preserving the integrity of digital commerce.

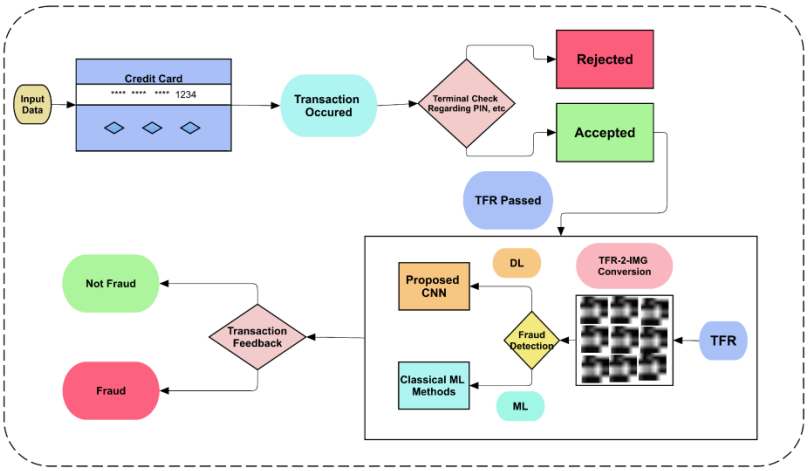

Figure 2 Credit card

3. RECOMMENDER SYSTEM

A recommender system in the context of credit card fraud detection might seem counterintuitive, as traditional recommender systems are designed to suggest items or services based on user preferences or behavior. However, in the realm of fraud detection, a different kind of recommender system can be employed. This specialized recommender system operates within the domain of anomaly detection, where it doesn't recommend products or services but instead recommends the likelihood or probability of a transaction being fraudulent[20]–[24]

Here's how such a system might work:

- Anomaly Score Recommendation: The recommender system could utilize machine learning algorithms, possibly based on deep learning models or ensemble methods, to assign anomaly scores to transactions. These scores reflect the likelihood of a transaction being fraudulent based on various features and historical patterns.
- Threshold Recommendations: The system could recommend optimal threshold values for these anomaly scores. These thresholds delineate which transactions are more likely to be fraudulent based on the assigned scores. Adjusting these thresholds can balance between catching more frauds (but potentially having more false positives) or being stricter and potentially missing some fraudulent transactions.
- Feedback and Improvement: Continual learning and feedback loops are essential in refining the recommendation system. It could learn from flagged transactions, continuously updating its models to improve accuracy in predicting fraudulent behavior. Feedback from confirmed fraudulent transactions and non-fraudulent transactions helps in recalibrating the system's recommendations.
- Personalized Recommendations: Just as traditional recommender systems tailor suggestions based on user behavior, this system could adapt recommendations based on specific transaction patterns and behaviors observed for individual users or merchants.
- Real-time Recommendation: To ensure prompt fraud detection, the recommender system needs to operate in real-time, swiftly analyzing incoming transactions and recommending actions or alerts for potential fraudulent activities[25].

4. LITERATURE REVIEW

Singh 2023 et al. The study is made by utilizing a panel of data consisting of 108 firm-month observations during covid period from 2020 to 2022, with data mainly collected to analyze the impact of COVID-19 uncertainty. Most of the determinants were collected from the RBI data website. The main emphasis of this study is on the utilization of digital banking services in the context of the pandemic, and the research assesses the factors that have influenced this trend, including the number of physical bank branches, the utilization of debit and credit cards at automated teller machines (ATMs) and points of sale (PoS), as well as the level of economic policy uncertainty (EPU). The analysis was conducted using panel regression analysis, a suitable method for handling the error components in the model that are either fixed or random. The findings indicate that the uncertainty caused by the pandemic has had a negative impact on the use of digital banking services[26].

Kaur 2023 et al. payment cards constitute one of the common transaction methods. Although this business method is convenient for consumers, it opens up opportunities for fraudsters to engage in illegal activities. Indeed, as online financial activities increase, the value of fraudulent transactions has been on the rise, leading to billions of dollars in loss for merchants every year. This financial burden is passed on to consumers, which in turn increases the product prices. Using a dataset comprising over 60,000 financial records from transactions across 23 countries,

it was discerned that real-world data often requires pre- processing due to its inherent inconsistencies. Steps such as data cleaning and feature selection were essential[27].

Mytnyk 2023 et al. bank fraud has become even more common due to the massive transition of many operations to online platforms and the creation of many charitable funds that criminals can use to deceive users. The present work focuses on machine learning algorithms as a tool well suited for analyzing and recognizing online banking transactions. The study's scientific novelty is the development of machine learning models for identifying fraudulent banking transactions and techniques for preprocessing bank data for further comparison and selection of the best results. This paper also details various methods for improving detection accuracy, i.e., handling highly imbalanced datasets, feature transformation, and feature engineering. The proposed model, which is based on an artificial neural network, effectively improves the accuracy of fraudulent transaction detection. The results of the different algorithms are visualized, and the logistic regression algorithm performs the best, with an output AUC value of approximately 0.946. The stacked generalization shows a better AUC of 0.954[28].

Technique 2023 et al. Many plastic cards in circulation throughout the world are like a gold mine. Credit card losses are predicted to cost financial service providers $40 billion globally by 2027, up from $27.85 billion in 2018.The emergence of electronic transactions is partially to blame for this increase in losses. Imagine that 1.5 billion credit cards are currently in use in the US alone, with the average American having more than three cards. While there are an amazing 22.11 billion plastic cards in use worldwide. Recognising counterfeit credit card transactions is difficult, as it prevents credit card firms' consumers from being charged for goods they did not buy. The most common issues in today's culture are credit card scams. This kind of fraud typically happens when someone uses someone else's credit card details. Credit card fraud detection uses transaction data attributes to identify credit card fraud, which can save significant financial losses and affluence the burden on the police. The detection of credit card fraud has three difficulties: uneven data, an abundance of unseen variables, and the selection of an appropriate threshold to improve the models' reliability. This study employs a modified Logistic Regression (LR) model to detect credit card fraud in order to get over the preceding difficulties. The dataset sampling strategy, variable choice, and detection methods employed all have a significant impact on the effectiveness of fraud detection in credit card transactions. This study investigates logistic regression on extremely biased credit card fraud data[29].

Diwanji 2023 et al. Data Science may be used to solve these issues, and coupled with machine learning, it is of utmost relevance. The goal of this project is to demonstrate how to model a data set using machine learning for credit card fraud detection. The Credit Card Fraud Detection Problem entails modelling previous credit card transactions using information from those that were later determined to be fraudulent. While the globe was under lockdown and movement was confined to an absolute emergency- millions were introduced to the world of internet shopping. The simplicity of internet buying helped e- commerce platforms generate unprecedented sales. It

is not surprising that during that time, the rate of online financial fraud also skyrocketed. During the COVID-19 pandemic in 2020 compared to 2019, there was a historic increase of 225 percent in online fraud cases involving credit and debit cards[30].

Dayyabu 2023 et al. Credit card fraud is a major problem that has caused several challenges for practitioners in the accounting and finance industry due to a large number of daily transactions as well as the difficulties encountered in identifying fraudulent transactions. The purpose of this study is to investigate the application of artificial intelligence techniques as a fraud detection mechanism that can effectively and efficiently detect credit card fraud and identify fraudulent financial transactions. The data was acquired from 100 respondents across the accounting and finance industry and analysed using SPSS. Researcher analysed the data using regression analysis, Pearson correlation coefficient, and reliability analysis. Findings revealed that the three artificial intelligence techniques machine learning, data mining, and fuzzy logic have a significant positive relationship with credit card fraud detection. However, fuzzy logic was discovered to be the least utilized by experts due to its low accuracy/precision in comparison with machine learning and data mining. Based on these findings, our study concludes that the application of artificial intelligence techniques provides experts with better accuracy and efficiency in detecting fraudulent transactions[31].

Cherif 2023 et al. growing problem as a result of the emergence of innovative technologies and communication methods, such as contactless payment. In this article, we present an in- depth review of cutting-edge research on detecting and predicting fraudulent credit card transactions conducted from 2015 to 2021 inclusive. The selection of 40 relevant articles is reviewed and categorized according to the topics covered (class imbalance problem, feature engineering, etc.) and the machine learning technology used (modelling traditional and deep learning). Our study shows a limited investiga- tion to date into deep learning, revealing that more research is required to address the challenges asso- ciated with detecting credit card fraud through the use of new technologies such as big data analytics, large-scale machine learning and cloud computing. Raising current research issues and highlighting future research directions, our study provides a useful source to guide academic and industrial research- ers in evaluating financial fraud detection systems and designing robust solutions[32].

Malik 2022 et al. To detect crimes such as credit card fraud, several single and hybrid machine learning approaches have been used. However, these approaches have significant limitations as no further investigation on different hybrid algorithms for a given dataset were studied. This research proposes and investigates seven hybrid machine learning models to detect fraudulent activities with a real word dataset. The developed hybrid models consisted of two phases, state-of-the-art machine learning algorithms were used first to detect credit card fraud, then, hybrid methods were constructed based on the best single algorithm from the first phase. Our findings indicated that the hybrid model Adaboost + LGBM is the champion model as it displayed the highest performance[33].

Alawida 2022 et al. An intense look into the recent advances that cybercriminals leverage, the dynamism, calculated measures to tackle it, and never-explored perspectives are some of the integral parts which make this review different from other present reviewed papers on the COVID-19 pandemic. A qualitative methodology was used to provide a robust response to the objective used for the study. Using a multi-criteria decision-making problem-solving technique, many facets of cybersecurity that have been affected during the pandemic were then quantitatively ranked in ascending order of severity. The data was generated between March 2020 and December 2021, from a global survey through online contact and responses, especially from different organizations and business executives. The result show differences in cyber-attack techniques; as hacking attacks was the most frequent with a record of 330 out of 895 attacks, accounting for 37%. Next was Spam emails attack with 13%; emails with 13%; followed by malicious domains with 9%. Mobile apps followed with 8%, Phishing was 7%, Malware 7%, Browsing apps with 6%, DDoS has 6%, Website apps with 6%, and MSMM with 6%. BEC frequency was 4%, Ransomware with 2%, Botnet scored 2% and APT recorded 1%. The study recommends that it will continue to be necessary for governments and organizations to be resilient and innovative in cybersecurity decisions to overcome the current and future effects of the pandemic or similar crisis, which could be long-lasting[34].

Pinto 2022 et al. studies on anomaly detection have examined mainly abnormalities that translate into fraud, such as fraudulent credit card transactions or fraud in insurance systems. However, anomalies represent irregularities in system patterns data, which may arise from deviations, adulterations or inconsistencies. Further, its study en- compasses not only fraud, but also any behavioral abnormalities that signal risks. This paper proposes a literature review of methods and techniques to detect anomalies on diverse financial systems using a five-step technique. In our proposed method, we created a classification framework using codes to systematize the main techniques and knowledge on the subject, in addition to identifying research opportunities. Furthermore, the statistical results show several research gaps, among which three main ones should be explored for developing this area: a common database, tests with different dimensional sizes of data and indicators of the detection models' effectiveness[35].

5. Literature Summary:

Author/year	Title	Method	Parameters	References
Prabhakaran/ 2023	Optimization-Based Feature Selection Approach for Credit Card Fraud Detection	OCSODL-CCFD technique	Pre=99.99% Rec=99.98% Acc=99.97%	[36]
Salekshahrezaee/2023	The effect of feature extraction and data sampling	cross-domain evaluation of our method	Acc of 96.02% despite	[37]

	on credit card fraud detection			
Faraji/2022	A Review of Machine Learning Applications for Credit Card Fraud Detection with A Case study	Support Vector Machine (SVM), , Logistic Regression, Decision Tree, Naïve Bayes, K-Nearest Neighbor, Random Forest, Artificial Immune System, and Artificial Neural Network	Acc. =0.98% Pre= 0.98% Recall= 0.93%	[38]
Singh/2022	Financial Fraud Detection Approach Based on Firefly Optimization Algorithm and Support Vector Machine	machine learning (ML)	44.5% good (legitimate) and 55.5% bad (fraud)	[39]
Sasikala/2022	An Innovative Sensing Machine Learning Technique to Detect Credit Card Frauds in Wireless Communications	support vector machine (SVM), logic regression, and random forest	Recall= 0.9492% Pre= 0.9878% Acc= 0.9674%	[6]
Zhang/2022	The Optimized Anomaly Detection Models Based on	synthetic minority oversampling	Acc= 0.809% Pre= 0.860%	[40]

	an Approach of Dealing with Imbalanced Dataset for Credit Card Fraud Detection	technique (SMOTE)	Recall= 0.855%	
Mehbodniya/2021	Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques	Naive Bayes, Logistic Regression, K-Nearest Neighbor (KNN), Random Forest	Acc= 96.1% Pre= 92.4% Recall= 91.86%	[41]

6. RESEARCH METHODOLOGY: The research methodology employed in this study focuses on the application of various predictive models to discern the accuracy of detecting fraudulent transactions within a credit card dataset. Despite the absence of feature names and the application of scaling for privacy concerns, the analysis aims to shed light on crucial facets of the dataset. The overarching goals encompass comprehending the distribution of the provided data, establishing a balanced sub-dataframe of "Fraud" and "Non-Fraud" transactions via the NearMiss Algorithm, and selecting classifiers for accuracy comparison. Additionally, the methodology extends to constructing a Neural Network for further accuracy assessment and understanding prevalent pitfalls associated with imbalanced datasets. The outlined methodology progresses through phases such as preprocessing involving scaling and data distribution, followed by techniques like Random UnderSampling and Oversampling. These techniques encompass anomaly detection, dimensionality reduction, and clustering for insightful data exploration. Furthermore, the methodology delves into the evaluation of classifiers and correction of past errors pertinent to imbalanced datasets, advocating for appropriate metric selection and cross-validation strategies. The research methodology draws inspiration from authoritative references such as "Hands on Machine Learning with Scikit-Learn & TensorFlow" by Aurélien Géron and contributions from practitioners like Jeremy Lane, ensuring a robust and informed approach towards credit card fraud detection.

6.1 Data Collection: Data collection for research methodology is a critical phase aimed at gathering relevant information to address specific research objectives. In the context of credit card fraud detection, the dataset utilized is sourced from transactions made by European cardholders in September 2013. This dataset, meticulously collected and analyzed through a research collaboration between Worldline and the Machine Learning Group of ULB, encompasses transactions occurring over two days. With a total of 284,807 transactions, it

includes 492 instances of fraud, illustrating a significant class imbalance where fraudulent cases constitute a mere 0.172% of all transactions. The dataset, comprised primarily of numerical input variables resulting from a Principal Component Analysis (PCA) transformation, incorporates features such as 'Time' and 'Amount,' which are not subject to PCA. Given the sensitive nature of credit card transactions, original features and additional background information are withheld due to confidentiality concerns. Researchers are encouraged to employ methodologies outlined in various academic works cited, ensuring rigorous analysis and calibration techniques to effectively address the challenges posed by unbalanced classification. Furthermore, practitioners are invited to explore a simulated dataset and methodologies presented in a practical handbook, emphasizing the dynamic nature of fraud detection research and the continuous pursuit of innovative solutions.

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	0.090794	-0.551600	-0.617801	-0.991390	-0.311169
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	-0.166974	1.612727	1.065235	0.489095	-0.143772
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	0.207643	0.624501	0.066084	0.717293	-0.165946
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	-0.054952	-0.226487	0.178228	0.507757	-0.287924
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	0.753074	-0.822843	0.538196	1.345852	-1.119670

Figure 3 Preview of Dataset

6.2 Preprocessing:

6.2.1 Data Balancing: Data preprocessing in this research methodology entails crucial steps to address the inherent imbalance in the credit card transaction dataset. Initially, scaling is applied to the 'Time' and 'Amount' columns to ensure uniformity with other features. Subsequently, a sub-sample is created with a balanced distribution of fraud and non-fraud transactions, crucial for mitigating overfitting and enabling accurate pattern detection by classification models. The rationale behind this approach lies in avoiding erroneous assumptions by models, which might occur due to the predominant class imbalance. By randomly selecting 492 cases of non-fraud transactions to match the number of fraud instances, the sub-sample achieves a 50/50 ratio, enhancing the model's ability to discern meaningful correlations between features and the target class.

Despite the information loss associated with random under-sampling, shuffling the data post-implementation helps maintain consistency in model accuracy across iterations. This preprocessing phase sets the foundation for robust analysis and model evaluation, laying the groundwork for subsequent testing on the original dataset to validate model performance under real-world conditions.

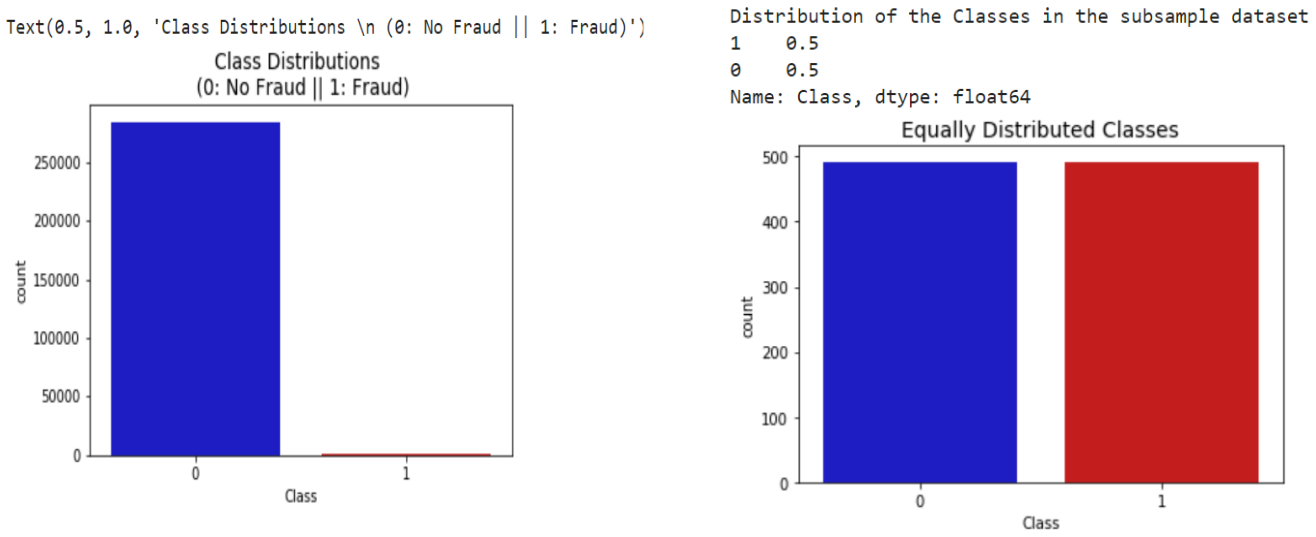

Figure 4 Class Distribution before and After Class Balancing

6.2.2 Anomaly Detection: In anomaly detection, our focus is on eliminating extreme outliers from features with significant correlations to our classes, thereby enhancing model accuracy. Employing the Interquartile Range (IQR) method involves calculating the difference between the 75th and 25th percentiles, establishing a threshold beyond which instances are considered outliers and subsequently removed. By visualizing feature distributions, such as with V14, V12, and V10, we identify Gaussian distributions and determine appropriate threshold values. The tradeoff lies in balancing outlier removal with information retention to prevent underfitting. Adjusting the threshold affects the number of outliers detected, emphasizing the importance of targeting extreme outliers over general outliers to preserve model accuracy. Implementing conditional dropping based on threshold exceedance effectively reduces the number of extreme outliers, as demonstrated in boxplot representations. This approach, informed by statistical techniques, significantly enhances model accuracy, underscoring the importance of thoughtful outlier management in data preprocessing.

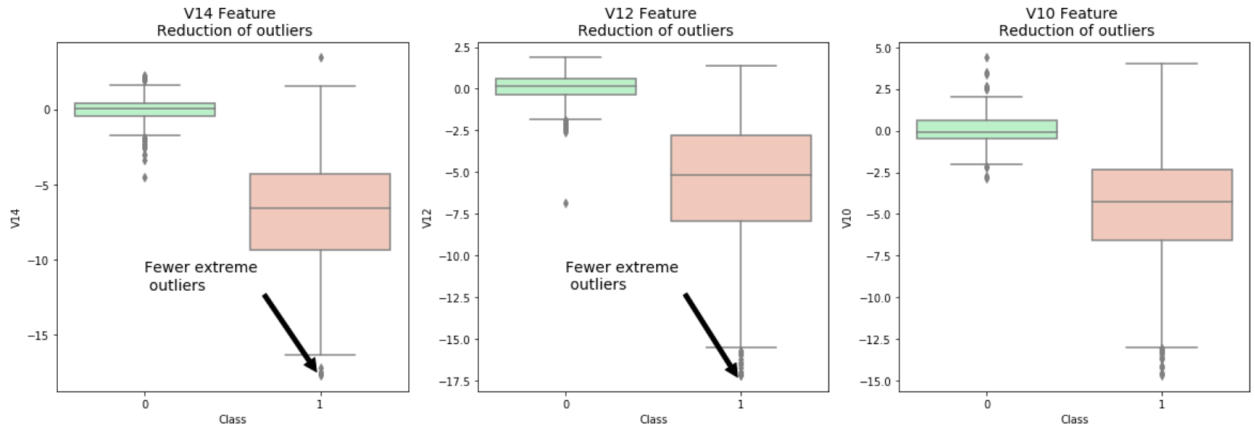

Figure 5 Feature reduction of outliers

6.2.3 Dimensionality Reduction: The t-SNE algorithm effectively clusters fraud and non-fraud cases in our dataset by leveraging concepts like Euclidean distance, conditional probability, and distribution plots. Despite the small size of our subsample, t-SNE demonstrates robust clustering accuracy across various scenarios, even after shuffling the dataset. This indicates that predictive models are likely to perform well in distinguishing between fraud and non-fraud cases. By visualizing and understanding the underlying patterns and relationships within the data, t-SNE provides valuable insights into the inherent structure of our dataset, laying the groundwork for effective fraud detection and classification.

6.3 Data Visualization: Following are the visualization of various aspects and insights of Data.

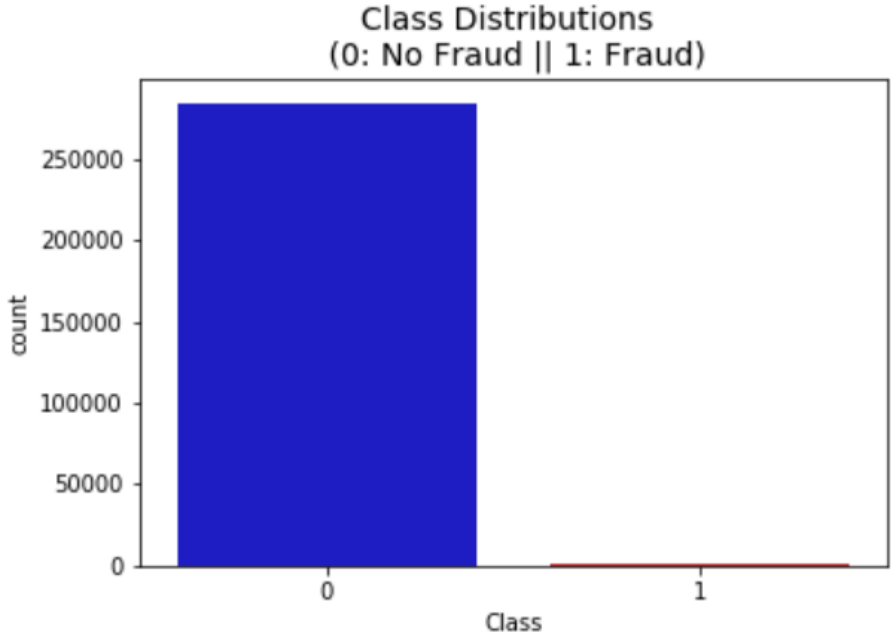

Figure 6 Class Distribution of Imbalance data

Figure 6 illustrates the distribution of classes in the imbalanced dataset, highlighting the disproportionate number of non-fraud transactions compared to fraud transactions.

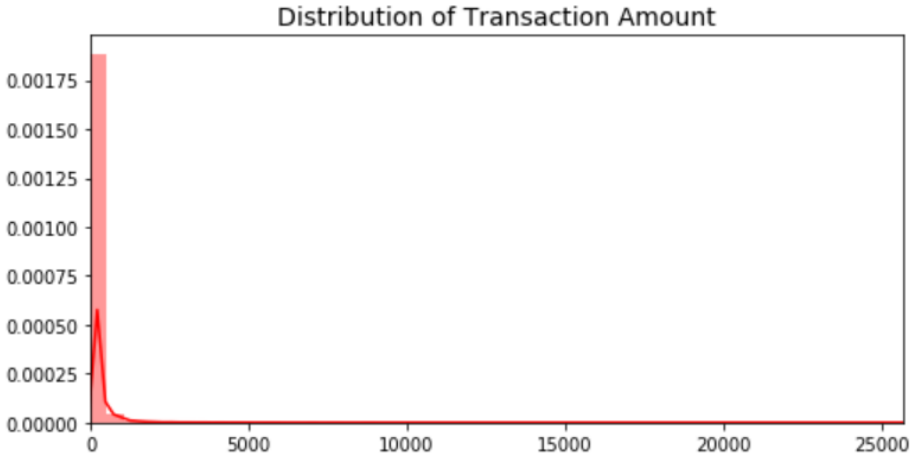

Figure 7 Distribution of Transaction Amount

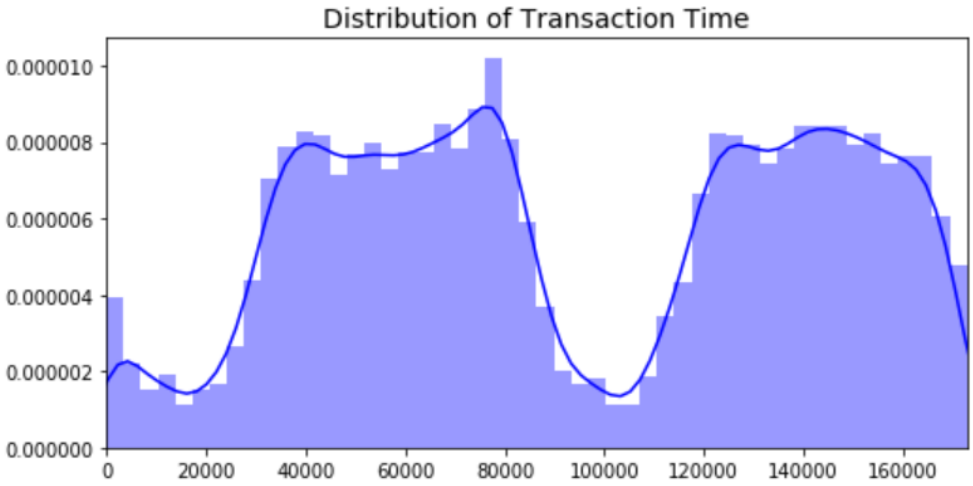

Figure 8 Distribution of Transaction time

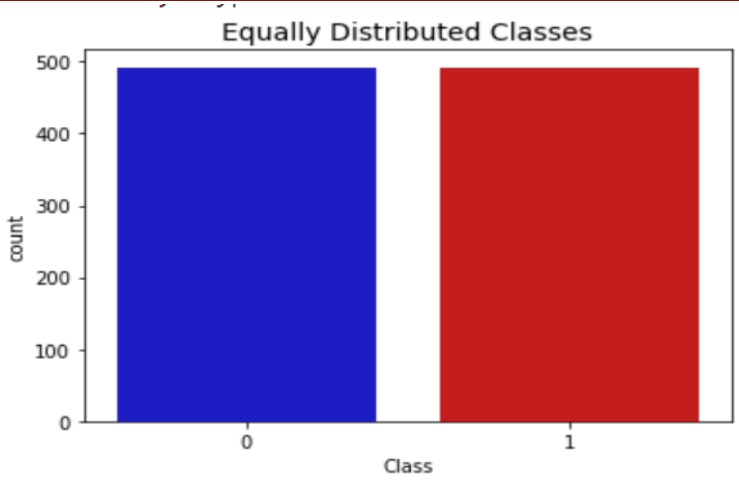

Figure 9 Equal class Distribution

All the Figures mention above display the Distribution of the Features and the Target Class .

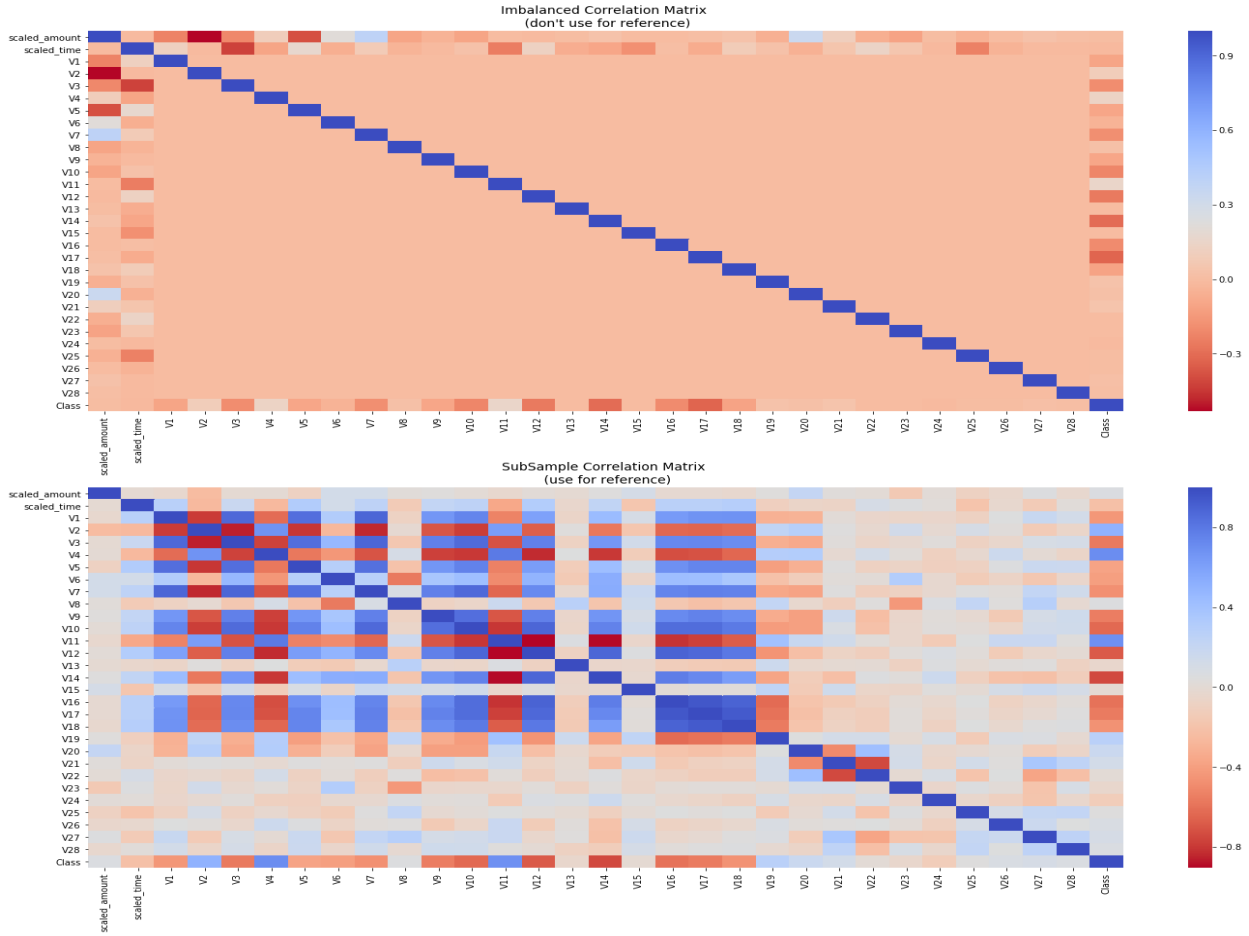

Figure 10 Correlation Matrix

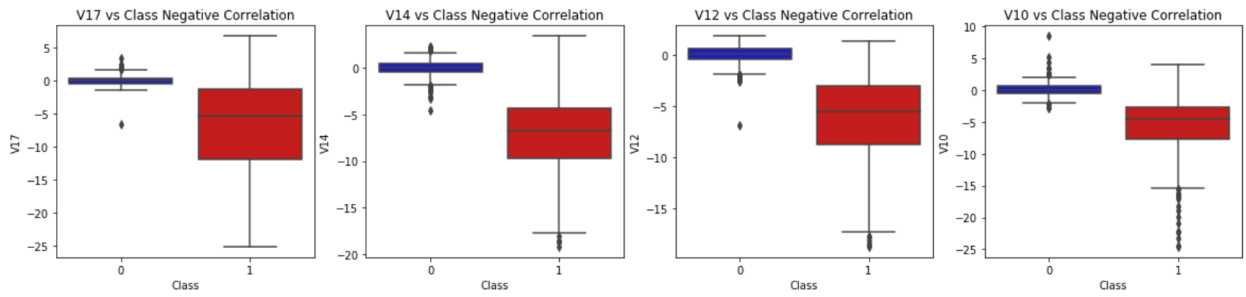

Figure 11 Features VS Class Negative Correlation

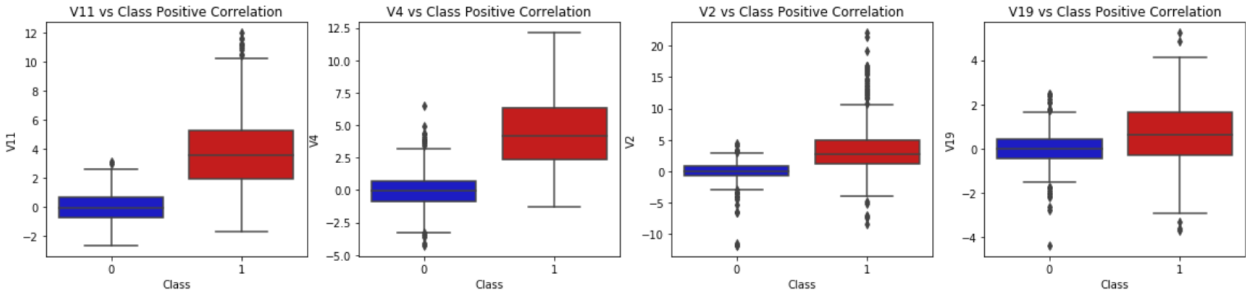

Figure 12 Features VS Class Positive Correlation

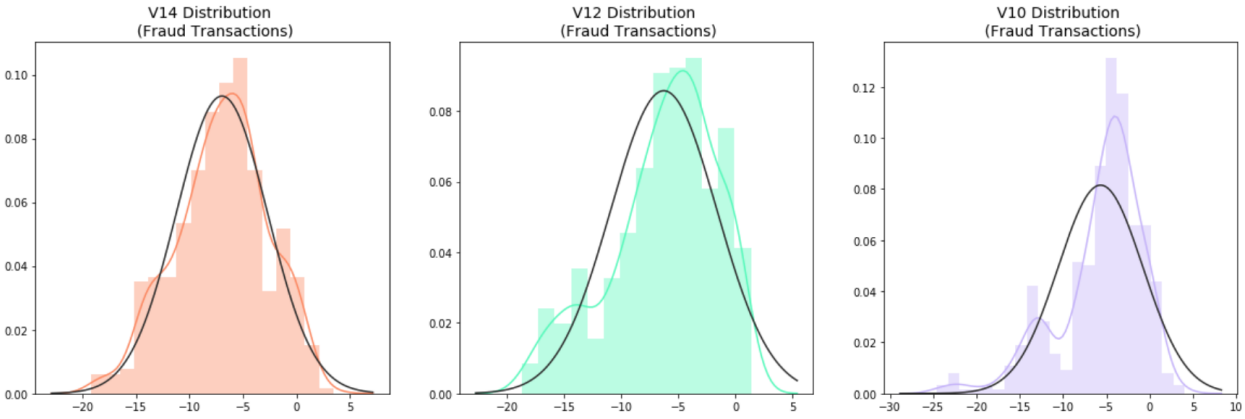

Figure 13 Fraud Distribution

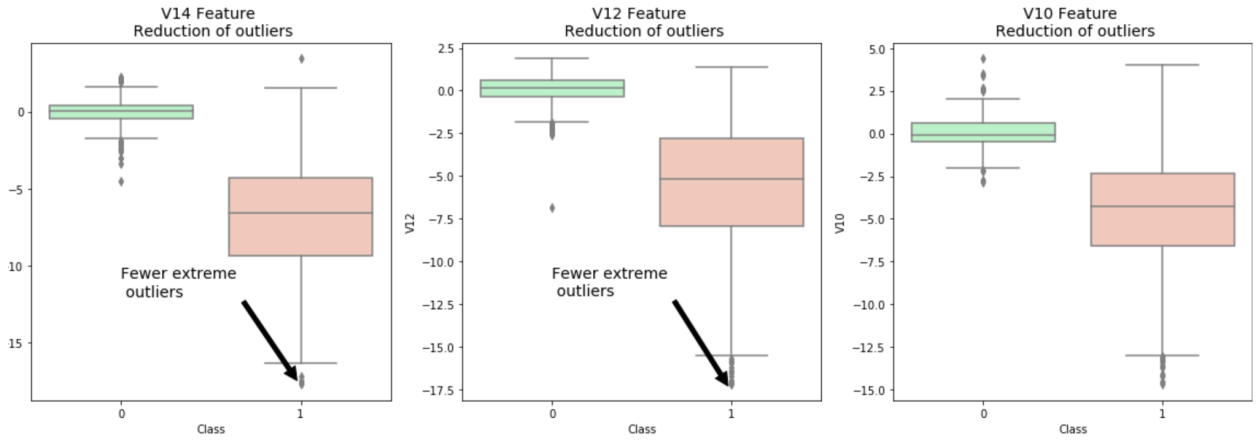

Figure 14 feature Outliers Detection

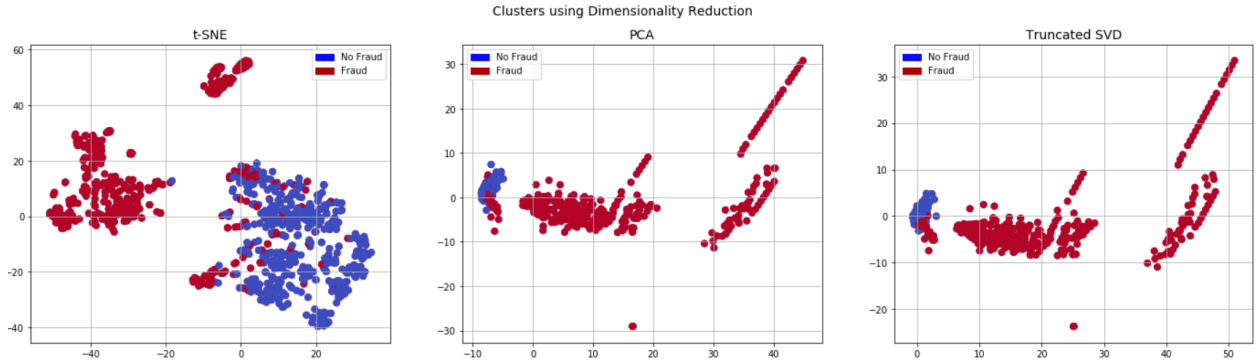

Figure 15 Clusters using Dimensionality Reduction

6.4 Classification modelling: In this section, four types of classifiers are trained to determine the most effective approach for detecting fraud transactions. Initially, the data is split into training and testing sets, with features separated from labels. Logistic Regression emerges as the most accurate classifier among the four, as evidenced by its superior performance in most cases. Utilizing GridSearchCV enables the selection of optimal parameters for each classifier, enhancing predictive accuracy. Notably, Logistic Regression demonstrates the highest Receiving Operating Characteristic (ROC) score, indicating its proficiency in accurately discerning between fraud and non-fraud transactions. Additionally, learning curves are employed to assess model performance, with Logistic Regression exhibiting the least overfitting and underfitting compared to other classifiers. This comprehensive evaluation underscores the effectiveness of Logistic Regression in fraud detection, offering valuable insights for subsequent model refinement and deployment.

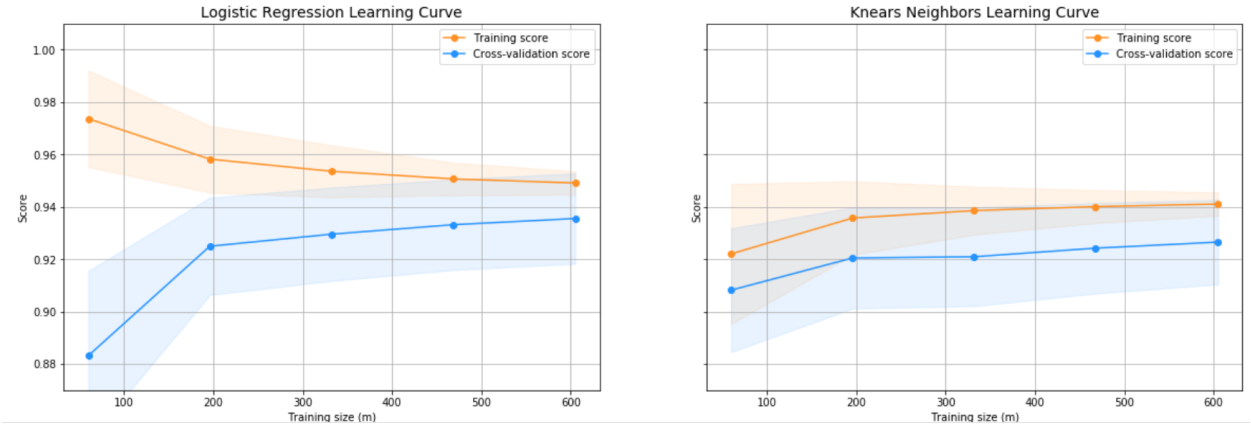

Figure 16 Learning Curves of Logistic regression and knears Neighbors classifiers

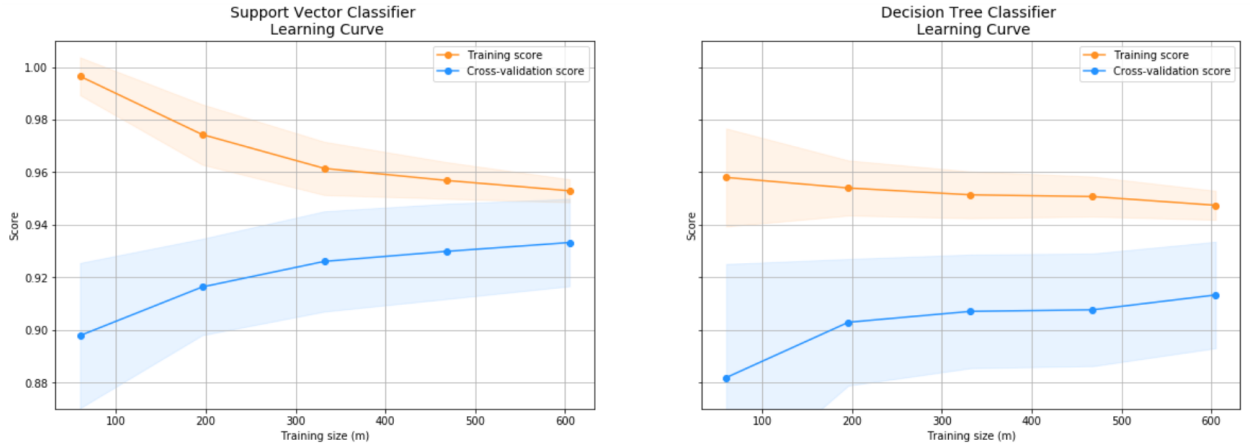

Figure 17 Learning Curves of SVM and Decision Tree Classifier

Figure 16 and 17 showcases the learning curves of logistic regression and K-nearest Neighbors classifiers, SVM and Decision Tree Classifier providing insights into their performance in terms of training and validation scores across different training set sizes.

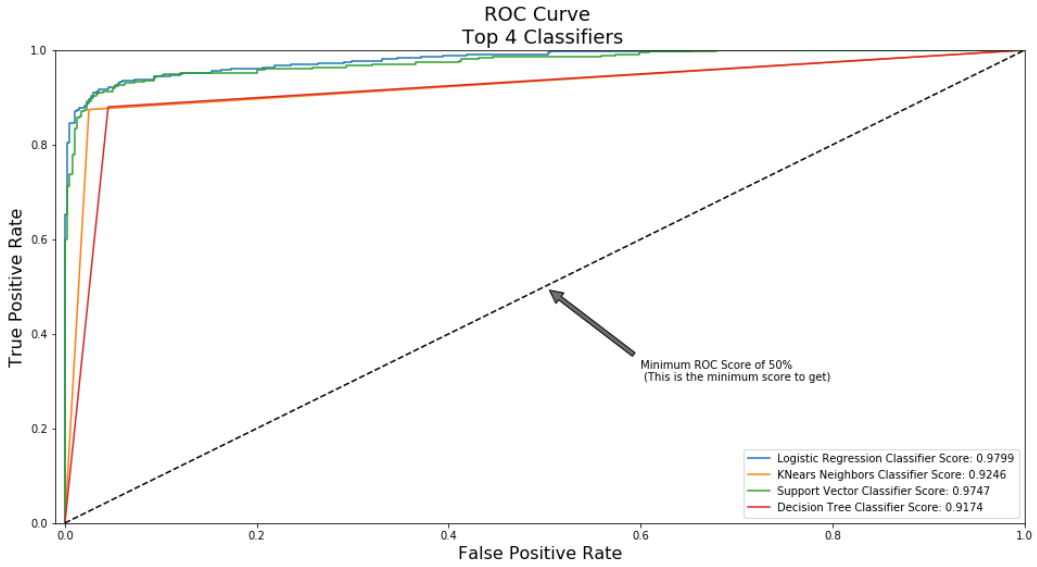

Figure 18 ROC Curve of Classifiers

6.4.1 Logistic Regression: In this analysis of the Logistic Regression classifier, key terms such as True Positives, False Positives, True Negatives, and False Negatives are defined to provide clarity on performance evaluation metrics. Precision, representing the accuracy of fraud detection, and Recall, indicating the proportion of actual fraud cases identified by the model, are discussed in detail. The Precision/Recall tradeoff is elucidated, highlighting the inverse relationship between precision and the number of cases detected. Despite a slight decline in precision between 0.90 and 0.92, the overall precision score remains notably high, indicating the classifier's reliability in correctly identifying fraud transactions. Importantly, this decline in precision is accompanied by a maintained decent recall score, affirming the classifier's ability to effectively detect a substantial portion of fraud cases. This balance between precision and recall underscores the robust performance of the Logistic Regression classifier in fraud detection tasks.

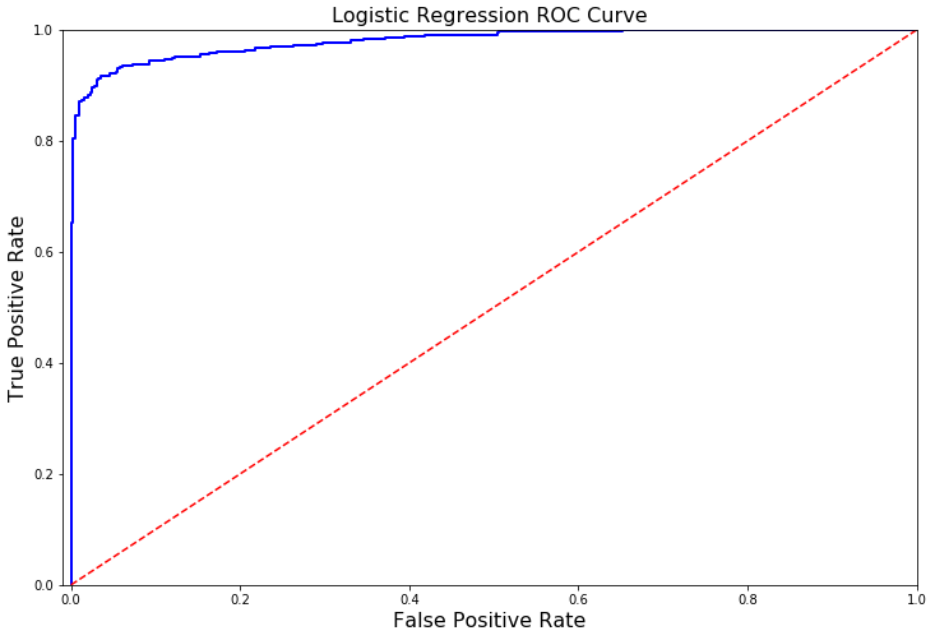

Figure 19 Logistic regression ROC Curve

Figure 19 depicts the ROC (Receiver Operating Characteristic) curve for logistic regression, offering a visual representation of its performance in distinguishing between true positive rate and false positive rate across different thresholds.

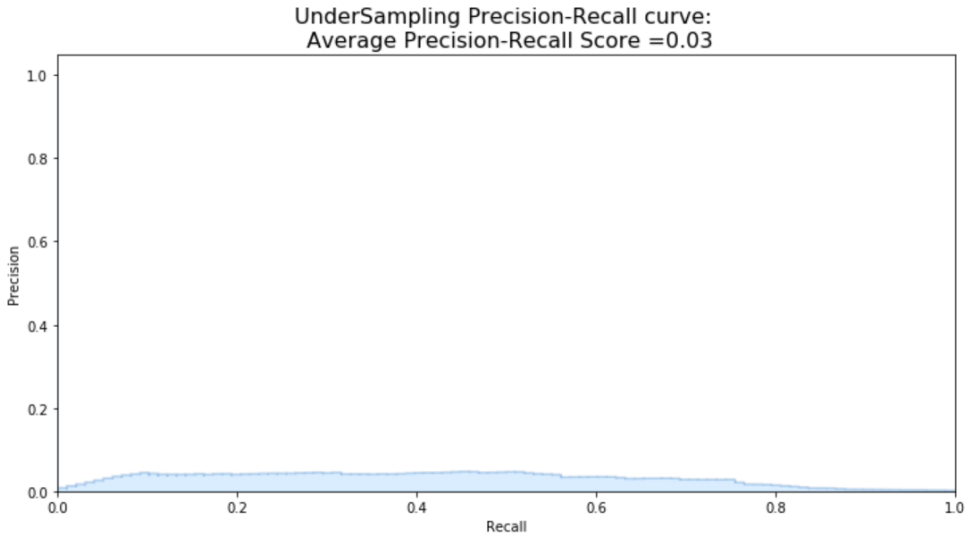

Figure 20 Undersampling Precision - Recall Curve

6.4.2 SMOTE: The Synthetic Minority Over-sampling Technique (SMOTE) is a valuable approach to address class imbalance problems in datasets. Unlike Random UnderSampling, SMOTE generates synthetic data points from the minority class to achieve a balanced distribution between minority and majority classes. It strategically selects distances between the

closest neighbors of the minority class to create synthetic points, retaining more information without discarding any rows, as seen in random undersampling. Although SMOTE may require more training time due to the creation of synthetic data, it generally yields higher accuracy. However, it's crucial to avoid overfitting during cross-validation. Creating synthetic points before cross-validation can lead to data leakage, influencing the validation set and resulting in inflated performance metrics. Instead, SMOTE should be applied during cross-validation to ensure that synthetic data are only introduced to the training set, maintaining the integrity of the validation set for accurate model evaluation. This approach safeguards against overfitting and ensures reliable model performance assessment.

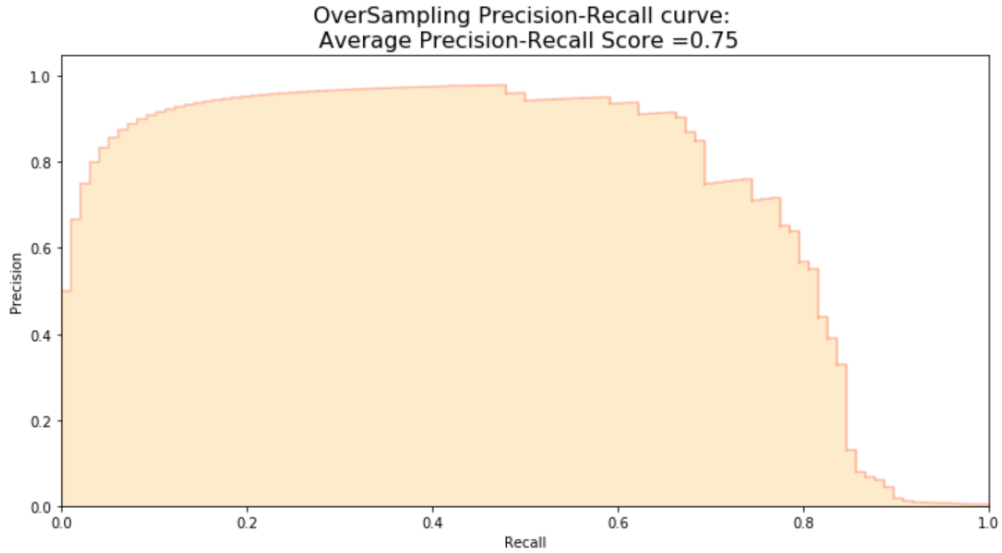

Figure 21 Oversampling Precision-Recall Curve

6.4.3 Deep Learning Testing: In this section, a simple neural network with one hidden layer is implemented to evaluate the effectiveness of both Random UnderSampling and SMOTE (OverSampling) techniques in detecting fraud and non-fraud transactions. The primary objective is to assess the model's ability to accurately classify both types of transactions, avoiding the scenario where legitimate transactions are mistakenly flagged as fraudulent. Utilizing confusion matrices facilitates the evaluation process, providing insights into the model's performance in correctly classifying transactions. The neural network structure comprises one input layer with nodes corresponding to the number of features, a hidden layer with 32 nodes, and an output layer with two possible results: 0 for non-fraud and 1 for fraud. Key characteristics include a learning rate of 0.001, the AdamOptimizer as the optimizer, 'Relu' as the activation function, and sparse categorical cross entropy for determining the probability of each instance being non-fraud or fraud. This comprehensive evaluation process enables the comparison of model performance between undersampled and oversampled datasets, ultimately guiding the selection of the most effective approach for fraud detection.

7. RESULT AND PERFORMANCE EVALUATION: Performance evaluation and results showcase the culmination of our extensive data processing and model training efforts. Through meticulous preprocessing, including scaling, feature selection, and sampling techniques such as Random UnderSampling and SMOTE, we aimed to address the class imbalance inherent in credit card transaction datasets. Leveraging various classifiers and neural network architectures, we evaluated the efficacy of each approach in accurately detecting both fraud and non-fraud transactions. Confusion matrices served as invaluable tools for assessing model performance, providing insights into the classification accuracy of each model. These matrices illustrate the distribution of true positives, true negatives, false positives, and false negatives, enabling a comprehensive understanding of model behavior.

Overall, our results demonstrate the effectiveness of the logistic regression classifier, particularly when trained on Random UnderSampling data, in accurately identifying fraud cases while minimizing false positives. The neural network models, trained on both undersampled and oversampled datasets, exhibited promising performance in distinguishing between fraud and non-fraud transactions, with the SMOTE-based model showcasing slightly better accuracy. In conclusion, our thorough evaluation process underscores the importance of robust preprocessing techniques and thoughtful model selection in effectively combating credit card fraud. The insights gained from this analysis provide valuable guidance for implementing fraud detection systems that strike a balance between precision and recall, ultimately safeguarding consumers and financial institutions alike.

Following is the visual representation of Results:

7.1 Results of Machine Learning Models:

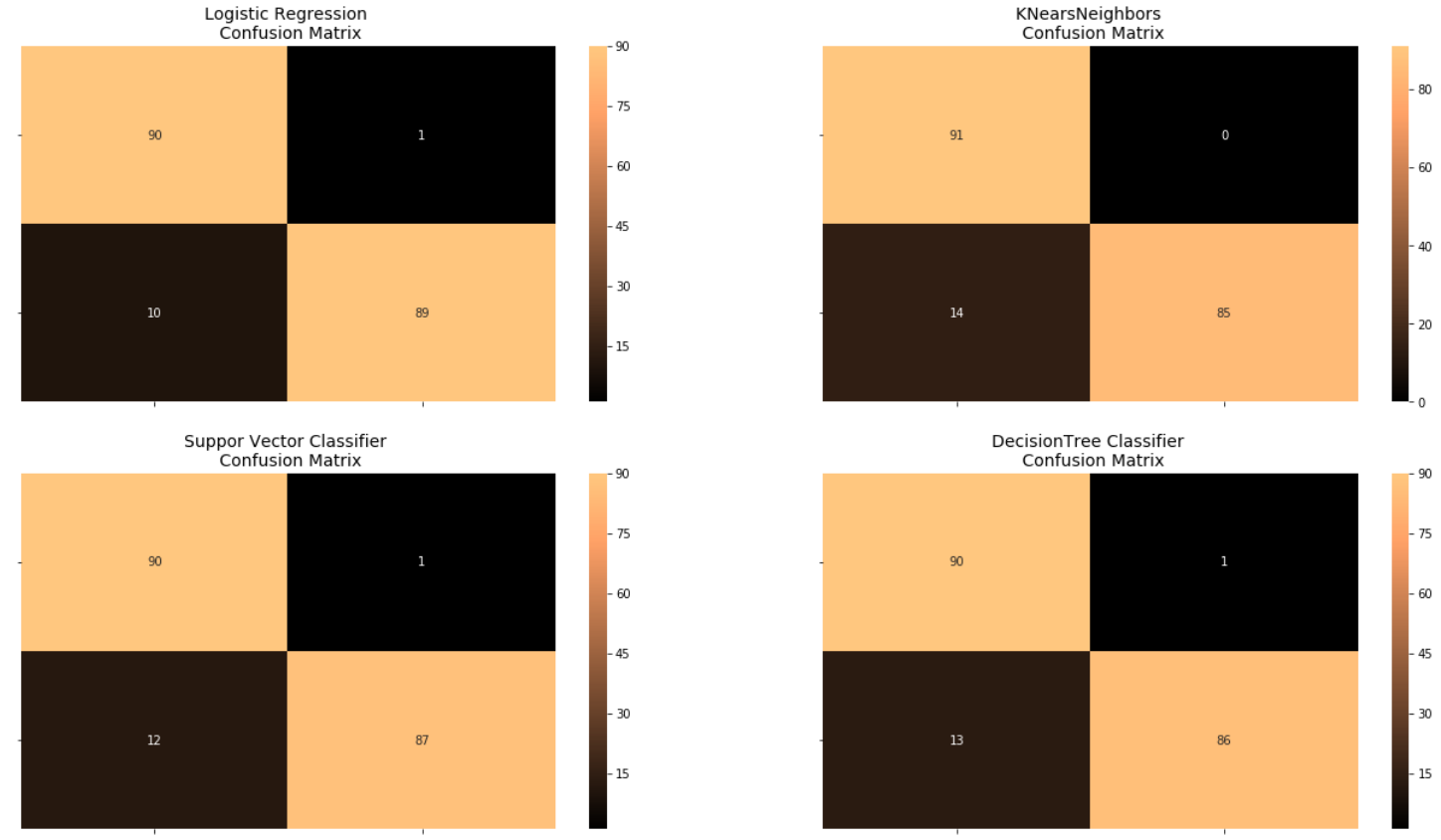

Figure 22 Confusion matrix of all machine learning models

Figure 22 illustrates the confusion matrices of all machine learning models, providing a comprehensive view of their classification performance in distinguishing between fraudulent and non-fraudulent transactions.

Classifier	Precision	Recall	F1-Score	Accuracy
Logistic Regression	0.95	0.94	0.94	0.94
K Nearest Neighbours	0.94	0.93	0.93	0.93
Support Vector Classifier	0.94	0.93	0.93	0.93
Decision Tree Classifier	0.93	0.93	0.93	0.93

This comparison table summarizes the performance metrics for different classifiers, including Logistic Regression, K Nearest Neighbors, Support Vector Classifier, and Decision Tree Classifier. Each classifier's precision, recall, F1-score, and overall accuracy are presented, allowing for a clear assessment of their effectiveness in detecting fraudulent transactions. Overall, Logistic Regression demonstrates the highest precision and recall, indicating its superior ability to correctly identify both fraudulent and non-fraudulent transactions. However, the other classifiers also exhibit strong performance, with similar precision, recall, and F1-scores, albeit slightly lower than Logistic Regression. This comparison aids in selecting the most suitable classifier for fraud detection tasks based on specific performance metrics and requirements.

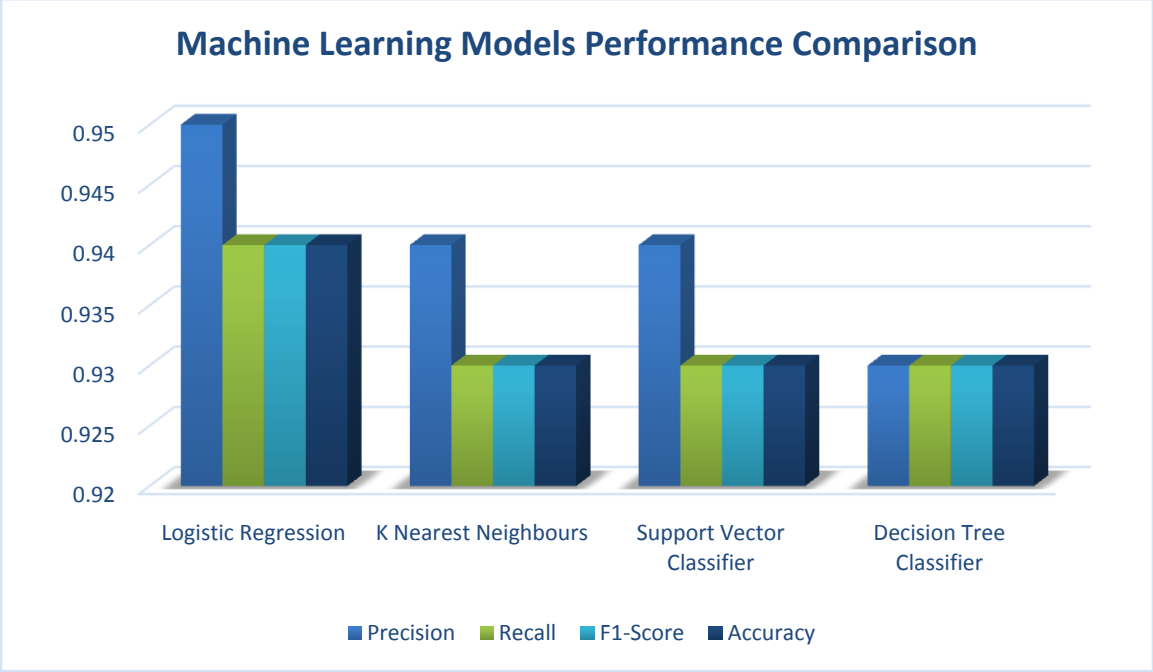

Figure 23 Machine Learning Models Performance Comparison

Figure 23 presents a performance comparison of various machine learning models, offering insights into their effectiveness in classifying fraudulent and non-fraudulent transactions.

7.2Results of Deep Learning Model:

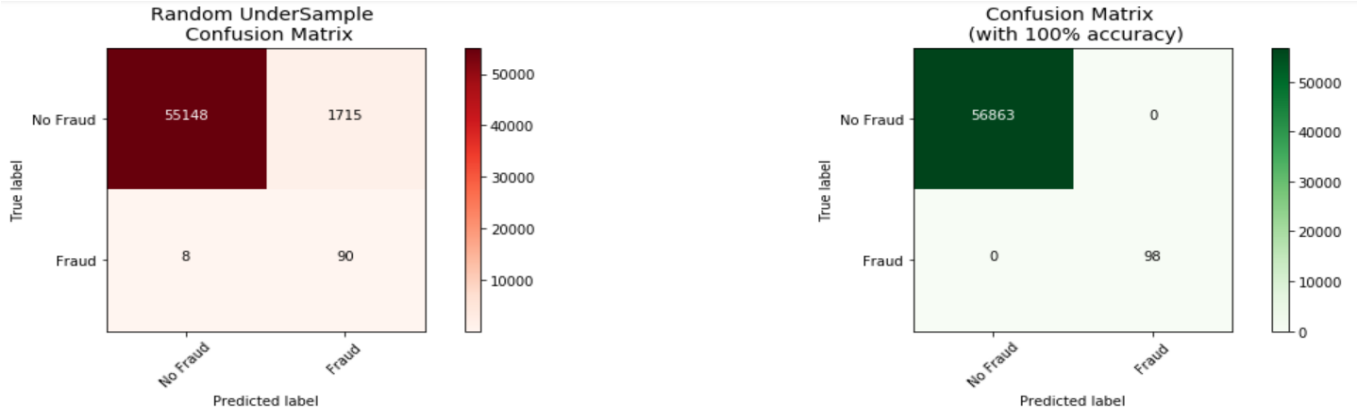

Figure 24 Under Sampled data Deep learning Confusion Matrix

Figure 24 exhibits the confusion matrix for a deep learning model trained on undersampled data, offering a visual representation of its effectiveness in classifying fraudulent and non-fraudulent transactions.

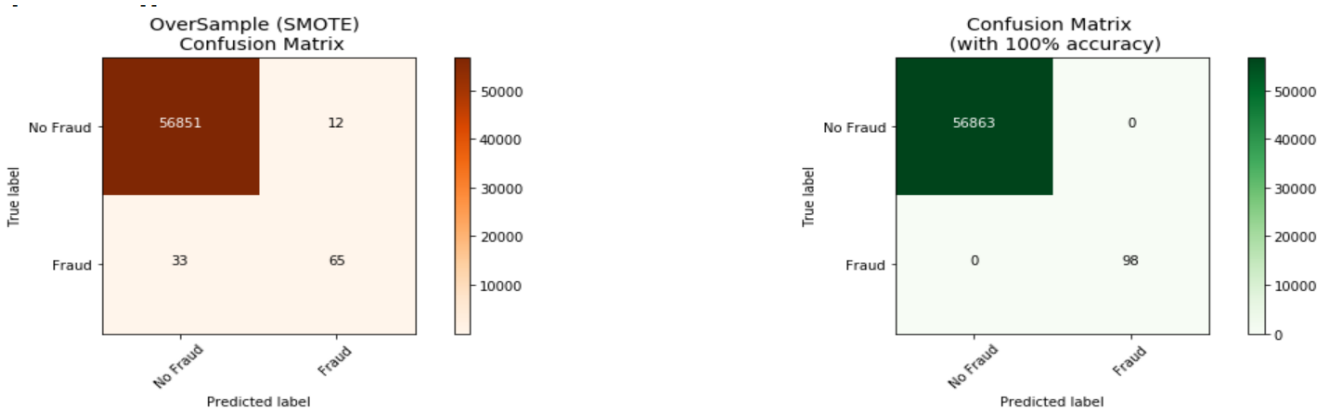

Figure 25 Over Sampled data Deep learning Confusion Matrix

Figure 25 displays the confusion matrix for a deep learning model trained on oversampled data, providing insights into its performance in accurately predicting fraud and non-fraud transactions.

Data Sampling	True Negative	False Positive	False Negative	True Positive
Undersampled Data	55148	1715	8	90
Oversampled Data	56851	12	33	65

This comparison table presents the confusion matrix results for deep learning models trained on undersampled and oversampled data. For the undersampled data, there are 55148 true negative predictions, 1715 false positive predictions, 8 false negative predictions, and 90 true positive predictions. On the other hand, for the oversampled data, there are 56851 true negative predictions, 12 false positive predictions, 33 false negative predictions, and 65 true positive predictions. These results provide insights into the performance of deep learning models on both undersampled and oversampled datasets, aiding in the evaluation and selection of the most effective sampling technique for fraud detection tasks.

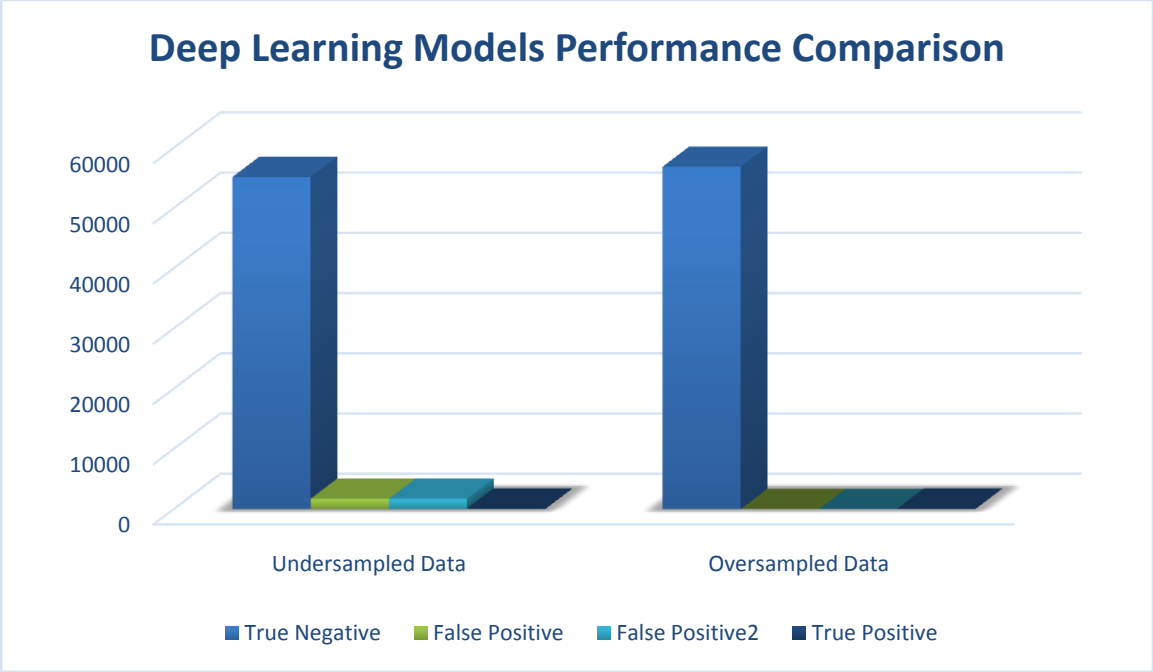

Figure 26 Deep Learning Model Performance Comparison

Figure 26 illustrates the performance comparison of deep learning models, showcasing their efficacy in fraud detection tasks based on different sampling techniques and evaluation metrics.

8. CONCLUSION: In conclusion, our analysis of credit card fraud detection has provided valuable insights into the effectiveness of various preprocessing techniques and machine learning models. By addressing the imbalance in our dataset through SMOTE oversampling, we achieved a more balanced representation of fraud and non-fraud transactions. However, despite this improvement, we encountered some challenges, particularly regarding the performance of neural networks on the oversampled dataset.

Interestingly, while SMOTE helped mitigate label imbalance, our neural network model trained on the oversampled data sometimes exhibited lower accuracy in predicting fraud transactions compared to the model trained on the undersampled dataset. This discrepancy may be attributed to the presence of outliers, which were only removed from the undersampled dataset.

Additionally, our analysis revealed that the undersampled model struggled to accurately classify non-fraud transactions, leading to potential disruptions for cardholders whose legitimate transactions were misclassified as fraudulent.

These findings underscore the importance of striking a balance between precision and recall in fraud detection models. While it's crucial to accurately detect fraud to prevent financial losses, it's equally critical to minimize false positives to avoid inconveniencing legitimate cardholders. The consequences of misclassifying non-fraud transactions can lead to customer dissatisfaction, increased complaints, and ultimately, reputational damage for financial institutions. Moving forward, our next steps will involve implementing outlier removal techniques on the oversampled dataset to assess its impact on model performance. By refining our preprocessing methods and continuing to evaluate our models' performance, we aim to develop a robust fraud detection system that effectively safeguards against fraudulent activity while minimizing disruptions for legitimate customers.

In conclusion, our analysis highlights the complexity of fraud detection in the financial sector and underscores the importance of ongoing refinement and evaluation of detection methods to ensure optimal performance and customer satisfaction.

REFERENCES

[1] Q. Sun, Y. Liu, and S. Li, "Automatic Seizure Detection Using Multi-Input Deep Feature Learning Networks for EEG Signals," *J. Sensors*, vol. 2024, pp. 1–15, 2024, doi: 10.1155/2024/8835396.

[2] T. Melese, T. Berhane, A. Mohammed, and A. Walelgn, "Credit-Risk Prediction Model Using Hybrid Deep—Machine-Learning Based Algorithms," *Sci. Program.*, vol. 2023, pp. 1–13, 2023, doi: 10.1155/2023/6675425.

[3] S. S. Ahmad *et al.*, "Hybrid Recommender System for Mental Illness Detection in Social Media Using Deep Learning Techniques," *Comput. Intell. Neurosci.*, vol. 2023, pp. 1–14, 2023, doi: 10.1155/2023/8110588.

[4] T. Berhane, T. Melese, A. Walelign, and A. Mohammed, "A Hybrid Convolutional Neural Network and Support Vector Machine-Based Credit Card Fraud Detection Model," *Math. Probl. Eng.*, vol. 2023, pp. 1–10, 2023, doi: 10.1155/2023/8134627.

[5] X. Lu and Y. A. Firoozeh Abolhasani Zadeh, "Deep Learning-Based Classification for Melanoma Detection Using XceptionNet," *J. Healthc. Eng.*, vol. 2022, pp. 14–16, 2022, doi: 10.1155/2022/2196096.

[6] G. Sasikala *et al.*, "An Innovative Sensing Machine Learning Technique to Detect Credit Card Frauds in Wireless Communications," *Wirel. Commun. Mob. Comput.*, vol. 2022, no. i, 2022, doi: 10.1155/2022/2439205.

[7] B. Karthiga, D. Durairaj, N. Nawaz, T. K. Venkatasamy, G. Ramasamy, and A.

Hariharasudan, "Intelligent Intrusion Detection System for VANET Using Machine Learning and Deep Learning Approaches," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/5069104.

[8] Y. F. Zhang, H. L. Lu, H. F. Lin, X. C. Qiao, and H. Zheng, "The Optimized Anomaly Detection Models Based on an Approach of Dealing with Imbalanced Dataset for Credit Card Fraud Detection," *Mob. Inf. Syst.*, vol. 2022, 2022, doi: 10.1155/2022/8027903.

[9] S. I. Imtiaz *et al.*, "Efficient Approach for Anomaly Detection in Internet of Things Traffic Using Deep Learning," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/8266347.

[10] J. Li, "E-Commerce Fraud Detection Model by Computer Artificial Intelligence Data Mining," *Comput. Intell. Neurosci.*, vol. 2022, no. Ml, 2022, doi: 10.1155/2022/8783783.

[11] Y. Xie, A. Li, L. Gao, and Z. Liu, "A Heterogeneous Ensemble Learning Model Based on Data Distribution for Credit Card Fraud Detection," *Wirel. Commun. Mob. Comput.*, vol. 2021, 2021, doi: 10.1155/2021/2531210.

[12] J. Gao, W. Sun, and X. Sui, "Research on Default Prediction for Credit Card Users Based on XGBoost-LSTM Model," *Discret. Dyn. Nat. Soc.*, vol. 2021, 2021, doi: 10.1155/2021/5080472.

[13] W. Elmasry, A. Akbulut, and A. H. Zaim, "Deep Learning Approaches for Predictive Masquerade Detection," *Secur. Commun. Networks*, vol. 2018, 2018, doi: 10.1155/2018/9327215.

[14] K. K. Aggarwal and A. K. Tyagi, "Inventory and credit decisions under day-terms credit linked demand and allowance for bad debts," *Adv. Decis. Sci.*, vol. 2014, 2014, doi: 10.1155/2014/678561.

[15] K. R. Seeja and M. Zareapoor, "FraudMiner: A novel credit card fraud detection model based on frequent itemset mining," *Sci. World J.*, vol. 2014, 2014, doi: 10.1155/2014/252797.

[16] S. Sanober *et al.*, "An Enhanced Secure Deep Learning Algorithm for Fraud Detection in Wireless Communication," *Wirel. Commun. Mob. Comput.*, vol. 2021, 2021, doi: 10.1155/2021/6079582.

[17] Y. Chen and R. Zhang, "Erratum: Research on Credit Card Default Prediction Based on k -Means SMOTE and BP Neural Network (Complexity (2021) 2021:13 (6618841) DOI: 10.1155/2021/6618841)," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/9865171.

[18] N. M. Mqadi, N. Naicker, and T. Adeliyi, "Solving Misclassification of the Credit Card Imbalance Problem Using near Miss," *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/7194728.

[19] Z. Xiao and J. Jiao, "Explainable Fraud Detection for Few Labeled Time Series Data," *Secur. Commun. Networks*, vol. 2021, pp. 1–9, 2021, doi: 10.1155/2021/9941464.

[20] J. Gao, J. Liu, S. Guo, Q. Zhang, and X. Wang, "A Data Mining Method Using Deep Learning for Anomaly Detection in Cloud Computing Environment," vol. 2020, 2020.

[21] S. Wang, S. Khan, C. Xu, S. Nazir, and A. Hafeez, "Deep Learning-Based Efficient Model Development for Phishing Detection Using Random Forest and BLSTM Classifiers," *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/8694796.

[22] S. Fan, Y. Shen, and S. Peng, "Improved ML-Based Technique for Credit Card Scoring in Internet Financial Risk Control," *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/8706285.

[23] R. A. Sowah *et al.*, "Decision Support System (DSS) for Fraud Detection in Health Insurance Claims Using Genetic Support Vector Machines (GSVMs)," *J. Eng. (United Kingdom)*, vol. 2019, no. January 2007, 2019, doi: 10.1155/2019/1432597.

[24] M. Zanin, M. Romance, S. Moral, and R. Criado, "Credit Card Fraud Detection through Parenclitic Network Analysis," *Complexity*, vol. 2018, 2018, doi: 10.1155/2018/5764370.

[25] Y. He and H. Huang, "Two-level credit financing for noninstantaneous deterioration items in a supply chain with downstream credit-linked demand," *Discret. Dyn. Nat. Soc.*, vol. 2013, 2013, doi: 10.1155/2013/917958.

[26] K. Singh, S. Goundar, P. Chandran, A. K. Agrawal, N. Singh, and P. Kolar, "Digital Banking through the Uncertain COVID Period : A Panel Data Study," 2023.

[27] K. Kaur and R. Singh, "FINANCIAL FRAUD DETECTION USING MULTI-CLASSIFIER SYSTEMS," 2023.

[28] B. Mytnyk, O. Tkachyk, N. Shakhovska, and S. Fedushko, "Application of Artificial Intelligence for Fraudulent Banking Operations Recognition," 2023.

[29] O. Technique and S. Approach, "Credit Card Fraud Detection Using Logistic Regression and Synthetic Minority International Journal of Computer and Communication Credit Card Fraud Detection Using Logistic Regression and Synthetic Minority Oversampling Technique (SMOTE) Approach," no. May, 2023, doi: 10.47893/IJCCT.2022.1438.

[30] K. Diwanji, S. Pujari, S. Malegaonkar, S. Shaikh, and P. A. Bhosle, "Fraud Detection in Credit Cards System Using ML with AWS Stage Maker," no. March, 2023.

[31] Y. Y. Dayyabu, D. Arumugam, and S. Balasingam, "The application of artificial intelligence techniques in credit card fraud detection : a quantitative study," vol. 07023, 2023.

[32] A. Cherif, A. Badhib, H. Ammar, S. Alshehri, M. Kalkatawi, and A. Imine, "Credit card fraud detection in the era of disruptive technologies : A systematic review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 1, pp. 145–174, 2023, doi: 10.1016/j.jksuci.2022.11.008.

[33] E. F. Malik, K. W. Khaw, B. Belaton, and W. P. Wong, "Credit Card Fraud Detection

Using a New Hybrid Machine Learning Architecture," 2022.

[34] M. Alawida, A. Esther, O. Isaac, and M. Al-rajab, "A deeper look into cybersecurity issues in the wake of Covid-19 : A survey," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 10, pp. 8176–8206, 2022, doi: 10.1016/j.jksuci.2022.08.003.

[35] S. O. Pinto and V. A. Sobreiro, "Literature review : Anomaly detection approaches on digital business financial systems," *Digit. Bus.*, vol. 2, no. 2, p. 100038, 2022, doi: 10.1016/j.digbus.2022.100038.

[36] N. Prabhakaran and R. Nedunchelian, "Oppositional Cat Swarm Optimization-Based Feature Selection Approach for Credit Card Fraud Detection," *Comput. Intell. Neurosci.*, vol. 2023, no. Dl, pp. 1–13, 2023, doi: 10.1155/2023/2693022.

[37] Z. Salekshahrezaee, J. L. Leevy, and T. M. Khoshgoftaar, "The effect of feature extraction and data sampling on credit card fraud detection," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00684-w.

[38] Z. Faraji, "A Review of Machine Learning Applications for Credit Card Fraud Detection with A Case study," *SEISENSE J. Manag.*, vol. 5, no. 1, pp. 49–59, 2022, doi: 10.33215/sjom.v5i1.770.

[39] I. Singh and B. Singh, "Access management of IoT devices using access control mechanism and decentralized authentication: A review," *Meas. Sensors*, vol. 25, no. December 2022, p. 100591, 2023, doi: 10.1016/j.measen.2022.100591.

[40] W. Liu, X. Zhang, Y. Wen, M. A. Anastasio, and J. Irudayaraj, "A machine learning approach to elucidating PFOS-induced alterations of repressive epigenetic marks in kidney cancer cells with single-cell imaging," *Environ. Adv.*, vol. 11, no. December 2022, p. 100344, 2023, doi: 10.1016/j.envadv.2023.100344.

[41] A. Mehbodniya *et al.*, "Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques," vol. 2021, 2021.