credit card fraud detection using ensemble learning with boosting technique

[1]Mahmud Mustapha Gana, [2]Mustapha Ismail, [3]Audu Musa Mabu

[1]BSc, [2,3]PhD
[1]Department of Computer Science,
[1]Gombe State University, Gombe, Nigeria

***Abstract*- This research paper proposes a novel approach for credit card fraud detection in the banking sector. The study utilizes ensemble learning with boosting techniques, combining the Random Forest(RF), Support Vector Machine(SVM), and Extreme Gradient Boosting(XGBoost) algorithms to create a powerful ensemble classifier. The approach is evaluated using an extensive dataset of credit card transactions. The results demonstrate exceptional recall, accuracy, precision, and F-score values with result of 1.0 for each evaluation metrics. In this study ensemble model developed outperforms previous studies by incorporating multiple evaluation measures and effectively leveraging the strengths of each base classifier. The research highlights the importance of considering a range of evaluation metrics and suggests avenues for further research in improving fraud detection systems. By addressing the limitations of earlier studies and using resampling techniques to handle imbalanced data, the proposed ensemble model offers significant potential for enhancing fraud detection and security protocols in the financial sector. The findings are considered trustworthy and have important implications for the industry, as they improve the realism and generalizability of credit card fraud detection through the use of the Kaggle.com dataset and ensemble learning techniques.**

***Keywords*: Credit Card Fraud, Ensemble Learning, Imbalanced Dataset, Boosting Technique & Machine learning.**

INTRODUCTION

Credit cards are increasingly popular in both developed and developing countries due to the wide range of uses for which they are used, including online transactions, bill payment, and shopping. However, as more people use credit cards, there is a higher chance of credit card fraud. Credit card fraud costs billions of dollars annually and occurs in a variety of methods, including lost or stolen cards, fake or counterfeit cards, website cloning, altering the data on the magnetic strip, phishing, skimming, and data theft from businesses [1].

It is exceedingly challenging to detect credit card fraud since there are so few fraudulent transactions compared to all other transactions [2]. This makes it difficult to correctly and swiftly identify fraud. Additionally, because fraudulent methods are continually getting better, it is essential to develop effective models for spotting and detecting credit card theft at its earliest stages [3].

Due to the reluctance of many firms to divulge information about fraudulent transactions, there is a dearth of labeled data available for training and testing these models, which presents another significant difficulty. Furthermore, numerous fraud detection models were built using conventional machine learning methods, which might be sensitive to unbalanced data and may not be able to keep up with new and emerging fraud patterns [4].

In order to get around these issues, researchers need to look at the use of ensemble learning techniques employing effective algorithms like Extreme Gradient Boosting (XGBoost), Random Forest (RF), and Support Vector Machine (SVM). There is also growing interest in using other types of data, such as network data and demographic data, to enhance the accuracy and robustness of fraud detection systems.

The fight against credit card fraud is undoubtedly a never-ending battle that demands ongoing innovation and adaptation. Algorithms for fraud detection must be able to handle skewed data and respond to evolving fraud trends.

This research aimed at the developing a machine learning model for credit card fraud detention system to prevent fraudulent transactions from occurring and to protect both the credit card issuer and the cardholder from financial loss with the following objectives: To study the user transaction pattern, to propose a model of detecting fraudulent credit card transactions, to easily track and report fraudulent transactions to relevant agency and to test and validate the model developed.

Significance of this study is to prevent the financial interests of the credit card issuer, safeguard the cardholder's financial interests, maintaining the trust of cardholders, reducing the possibility of identity theft and improving the security of the financial system. The scope of this study is to assess how well credit card fraud may be detected using machine learning algorithms and ensemble learning approaches. Other preventative measures against identity theft are not covered by this study.

OVERVIEW OF CREDIT CARD FRAUD DETECTION

Credit card fraud is an unlawful use of a credit card or credit card information to make purchases or apply for loans [5]. It can happen in a number of ways, including when the card is physically taken, when card information is stolen online, or when imitation cards are made using stolen card information. Credit card fraud can lead to large financial losses for both the cardholder and the financial institution that issued the card [6].

Credit card fraud detection is the process of identifying suspicious credit card transactions and taking action to prevent fraudulent activity [7]. It is an important issue for financial institutions and merchants, as credit card fraud can result in significant financial

losses and damage to the reputation of the business. There are several techniques that can be used to detect or prevent credit card fraud, financial institutions and merchants use a variety of techniques, including: Fraud monitoring, Fraud detection algorithms, Verification procedures and Fraud alerts.

By using these and other techniques, it is possible to significantly reduce the incidence of credit card fraud and protect both the financial institution and the cardholder from financial losses.

Machine Learning

Machine learning is a technique of training computers to learn from data, without being explicitly programmed [8]. It is a subset of artificial intelligence that focuses on the development of algorithms and statistical models that allow computers to learn from and make decisions based on data inputs [9]. There are several different types of machine learning, including: Supervised learning, Unsupervised learning, Semi-supervised learning and Reinforcement learning

Ensemble Learning

Ensemble learning is a method of training a model that combines the predictions of multiple smaller models to make a final prediction. The idea behind ensemble learning is to train several models independently, then combine their predictions in a way that is more accurate than any individual model [10].

There are several different methods for combining the predictions of multiple models out of which boosting technique was chosen for this study. Due to its ability to integrate the capabilities of numerous models to yield a more accurate final prediction, ensemble learning has been a potent technique for improving the performance of a machine learning model.

Extreme Gradient Boosting (XGBoost)

XGBoost, which stands for "Extreme Gradient Boosting", is a popular machine learning library that can be used for a variety of tasks [11], including credit card fraud detection. The library provides an implementation of gradient boosting, which is a type of ensemble learning that can be used to improve the accuracy of predictive models. In the context of credit card fraud detection, XGBoost can be used to train a model that can predict whether a given credit card transaction is fraudulent or not [12].

Random Forest (RF)

A machine learning system called Random Forest (RF) can be used to find credit card fraud. According to [13], it is a form of ensemble learning method that integrates several decision trees to produce a prediction. Because it can handle enormous volumes of data and spot intricate patterns in the data that may point to fraud, this approach is helpful in the identification of fraud. As there are many more valid transactions than fraudulent ones, imbalanced datasets are a typical issue in fraud detection and can be handled using RF [14]. The most crucial elements in the dataset that are influencing the prediction of fraud can be found using the feature importance data that RF can also offer [14].

Support Vector Machines (SVMs)

Popular machine learning technique Support Vector Machine (SVM) can be used to identify credit card fraud [15]. The algorithm is a supervised learning method that may be applied to both regression and classification applications. The algorithm is trained on a dataset of previous credit card transactions, both fraudulent and legitimate, in the instance of credit card fraud detection. The objective is to find data patterns that can be utilized to distinguish between fraudulent and legitimate transactions. The algorithm can be used to categorize new transactions as either fraudulent or lawful after it has been trained.

Material and methods

Machine learning is a type of quantitative research that uses statistical and mathematical models to detect patterns and predict outcomes. Credit card fraud detection can be done by training a machine learning model on historical credit card transaction data to detect fraudulent transactions. Ensemble learning and boosting techniques are also powerful methods that can be used to improve the performance and robustness of a credit card fraud detection model. By combining multiple models using ensemble techniques, or iteratively training models with boosting, the system can become more robust to noise and variations in the data, potentially leading to a more accurate and reliable detection of fraudulent transactions.

Data Collection

Data collection is a crucial task that influences every kind of research, and it's even more important when building a model for credit card fraud detection [16]. However, collecting data for credit card fraud detection can be difficult as the data is sensitive and private. Financial institutions may be hesitant to share data due to privacy and security concerns. Additionally, fraud data is unbalanced and rare, which can lead to difficulties in obtaining a representative sample. Furthermore, credit card fraud detection data is dynamic and changing, which means that the model should be updated frequently to be able to detect the latest types of frauds. In this research, a secondary source(https://www.kaggle.com) was used as a source of data collection, and the data preprocessing, cleaning and feature selection has been used with the help of python programming language in google colab to ensure the quality of the dataset.

Understanding the Dataset

The study utilized a comprehensive European dataset from (https://www.kaggle.com) for credit card transactions analysis. This dataset provides valuable insights into consumer spending habits, credit risk, and fraud detection. It also enabled us to build robust predictive models to improve financial institutions' decision-making processes.

The dataset consists of credit card transactions done by European cardholders over a two-day period in September 2013 (kaggle.com). 492 fraudulent transactions out of 284,807 total transactions were made. With the positive class (fraud) accounting for 0.172% of all transactions, this dataset is very lopsided. In order to ensure confidentiality, the dataset has also been adjusted

using Principal Component Analysis (PCA). All other features (V1, V2, V3, etc.) other than "Time" and "Amount" are the Principal Components produced using PCA. The seconds that passed between the dataset's first transaction and subsequent transactions are contained in the feature "Time." The transaction amount is the "Amount" feature. The feature "Class" represents class labeling and has a value of "1" in fraud cases and "0" in all other cases [17].

Gap Identified

This study is required to enhance the results the detecting credit card fraud. It has been demonstrated that combining the XGBoost, Random Forest, and SVM algorithms with ensemble learning, specifically the boosting technique, improved the prediction result for credit card fraud. This is because these algorithms have been successful in other classification tasks, and the ensemble method has been demonstrated to improve performance in a variety of applications. However, more investigation is required to support this claim utilizing a particular dataset for credit card fraud detection. This would more appropriately represent the dearth of prior research on this particular ensemble method and algorithm combination for detecting credit card fraud and would also highlight the need for future study.

PROPOSED CREDIT CARD FRAUD DETECTION TECHNIQUE

➢ **Ensemble Learning (Boosting Technique)**

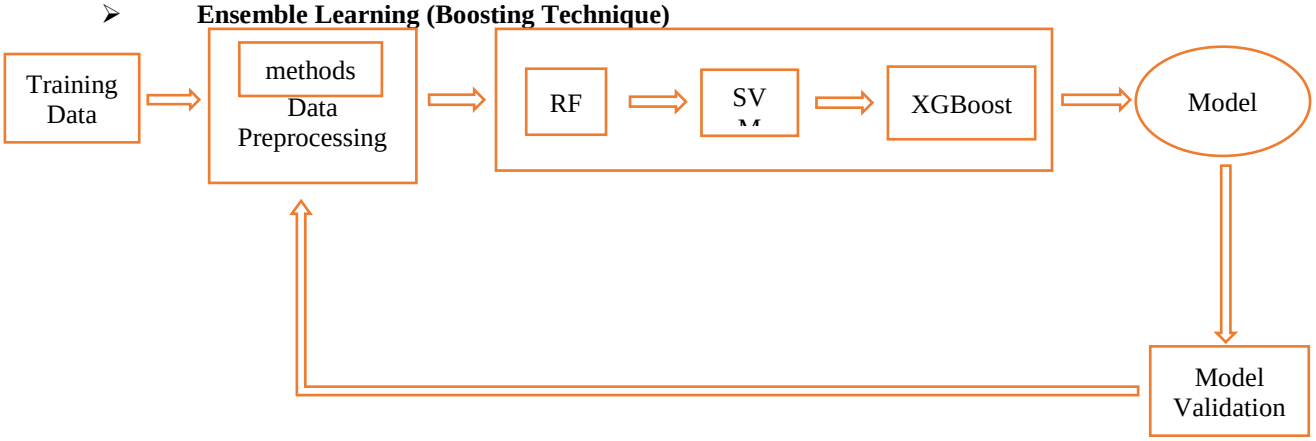

Fig 1: Training Phase

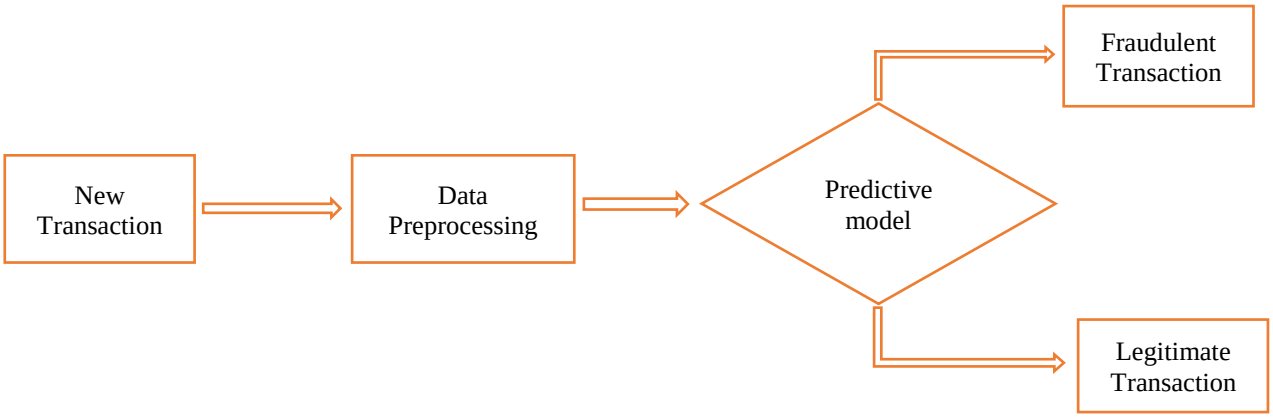

Fig 2: Prediction Phase

RESULTS AND DISCUSSION

Researchers have always been interested in the identification of credit card fraud, and they will continue to be so in the future. The continuous changes in fraud patterns are largely to blame for this. By applying best-fitting algorithms to identify distinct fraudulent transaction patterns and by addressing the related problems discovered by earlier researchers in credit card fraud detection, a novel credit-card fraud detection system is offered in this study. Detecting credit card theft is made easier using predictive analytics. Several literature reviews were observed to select the best algorithms to combat fraud. The sampling technique was also carefully considered to remedy the skewed distribution of the data. As a result, it was determined that using resampling techniques to significantly improve the performance of the classifiers has a substantial impact.

The outcomes of the experiment are presented and discussed in this section. This research work also described the step-by-step procedure to implement preventative measures. In order to determine the outcome of all three models using ensemble learning, the study first analyzes and comprehends the 284,807 data set. This study also develops and assesses various machine learning algorithms, including Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Machine (SVM). Python was used as a programming language and Jupyter notebook in the Google Collaboratory to write, execute, and visualize the python code in order to do all of this. For data analysis, Numpy, Panda, and Sklearn libraries; for model creation.

DATA VISUALIZATION

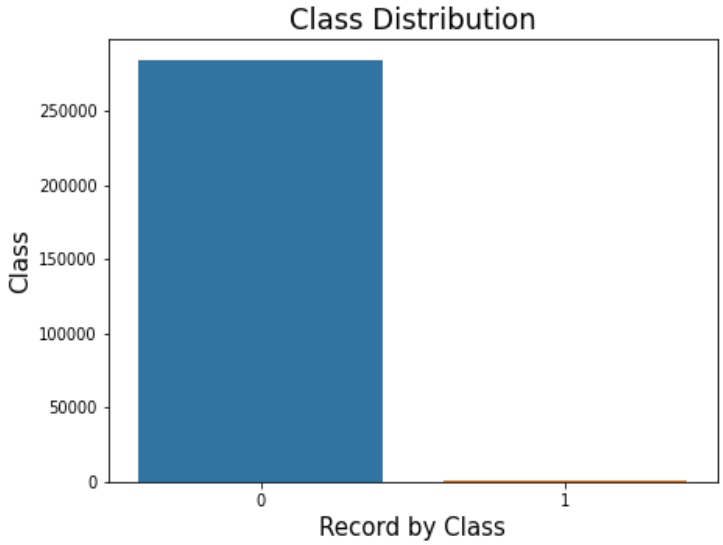

Fig 3: presents a bar chart that illustrates the distribution of classes within the dataset.

The chart visually depicts the proportion or frequency of each class category. In this case, the classes represent fraudulent transactions (labeled as '1') and legitimate transactions (labeled as '0'). Upon analysis, it is observed that out of the total transactions considered, which represents 100%, only a mere 0.173% corresponds to legitimate transactions. This indicates a significant class imbalance, with fraudulent transactions being the dominant class.

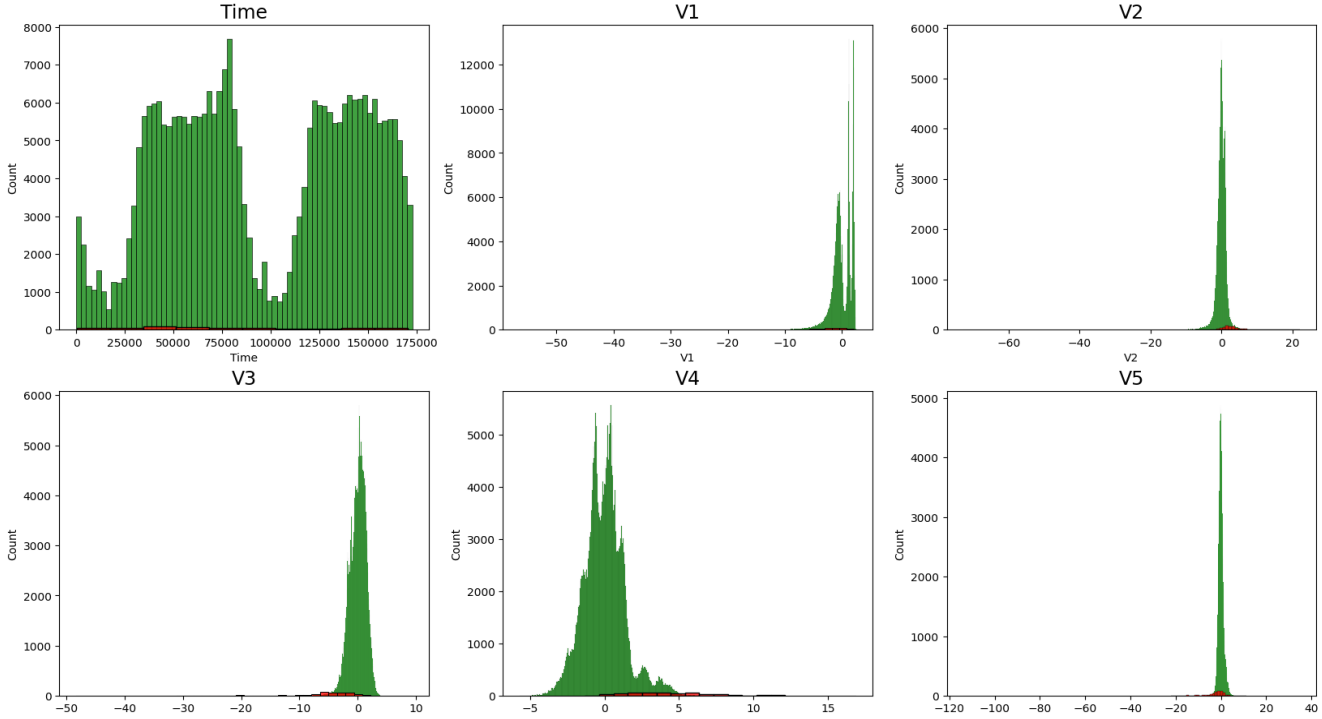

Fig 4: presents a histogram plot that illustrates the skewness of the data for the first 6 features.

A histogram is a graphical representation that displays the distribution of data by dividing it into intervals or bins along the x-axis and showing the frequency or count of observations falling into each bin on the y-axis. The histogram provides insights into the skewness of the data distribution for each feature. Skewness refers to the measure of asymmetry in the data distribution. A perfectly symmetrical distribution has a skewness of 0, while positive and negative skewness indicate a right-skewed (tail on the right) and left-skewed (tail on the left) distribution, respectively. By analyzing the histogram, we can assess the skewness of the data for the first 6 features. If the distribution of a feature appears to be skewed, it suggests that the data is not evenly distributed and may exhibit a bias towards certain values.

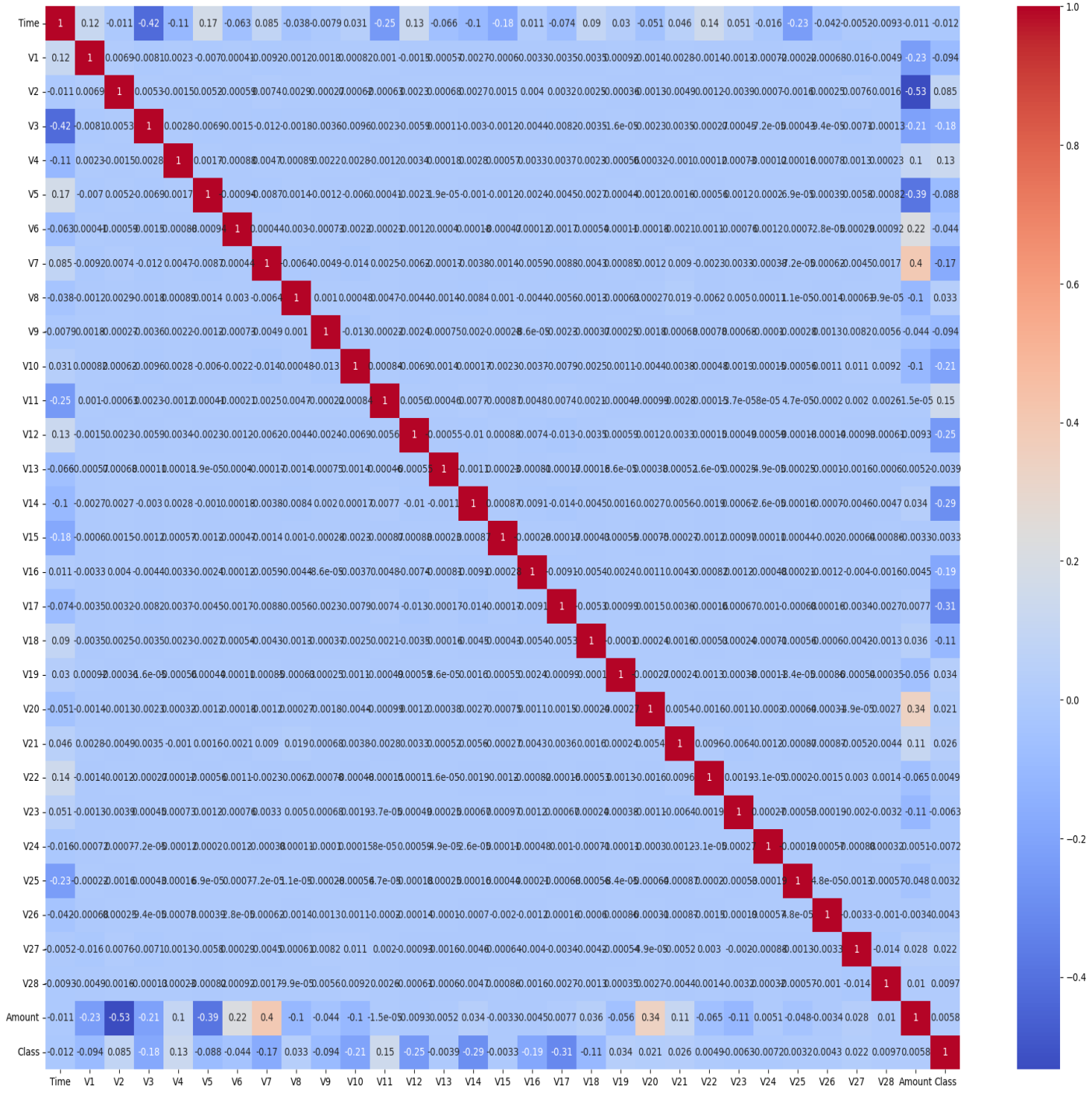

Fig 5 Correlation map: plot of heat-map to show the correlation between feature variables such as strength, direction, pattern identification, and outliers & unusual correlations. Where dark blue indicate highly correlation and light blue indicate low or no correlation between features.

After evaluating the model's performance using mean squared error (MSE) on the training and test sets and determines whether the model is overfitting, underfitting, or achieving a good fit. The target variable was first predicted using the ensemble model on the corresponding input data and calculated the mean squared error (MSE) between the true target values and the predicted values using the mean_squared_error function. The MSE values on both the training set and the test set were printed to check for overfitting, underfitting, or a good fit by comparing the MSE values on the training and test sets. If mse_train is less than mse_test, it indicates overfitting. If mse_train is greater than mse_test, it indicates underfitting, otherwise good fit. The result of the check is printed as "Good fit" in this case, as the MSE on both the training and test sets is 0.0, implying that the model performs equally well on both sets. Finally, a learning curve is plotted with the MSE values on the y-axis and the training and testing errors labeled on the legend. Learning curve was visualized below to show the trend of the training and testing errors.

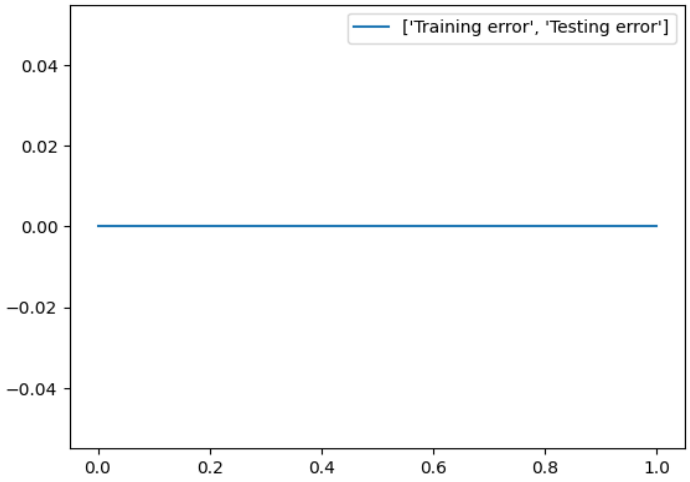

Fig 6: plot of learning curve for ensemble learning.

The following table presents the performance metrics of individual classifiers, namely Random Forest, Extreme Gradient Boosting, and Support Vector Machine, in terms of accuracy, precision, recall, and F1-score.

Table 1: The result of three individual model before ensemble.

Classifier	**Accuracy**	**Precision**	**Recall**	**F1** score
Random Forest	1.0	1.0	1.0	1.0
Extreme Gradient Boosting	1.0	1.0	1.0	1.0
Support Vector Machine	0.99	1.0	1.0	1.0

The results of **Table 1** indicate that both Random Forest and Extreme Gradient Boosting achieved perfect scores (1.0) across all metrics, demonstrating excellent performance. The Support Vector Machine classifier achieved a slightly lower accuracy of 0.99, while maintaining perfect precision, recall, and F1 score.

Table 2: Result of Ensemble Learning (Boosting)

Classifier	**Accuracy**	**Precision**	**Recall**	**F1** score
Ensemble Learning(Boosting)	1.0	1.0	1.0	1.0

As shown in **Table 2**, the ensemble learning approach using Boosting yielded excellent results, with all three classifiers achieving perfect scores (1.0) across all metrics. This reinforces the effectiveness of the ensemble learning technique in improving classification performance.

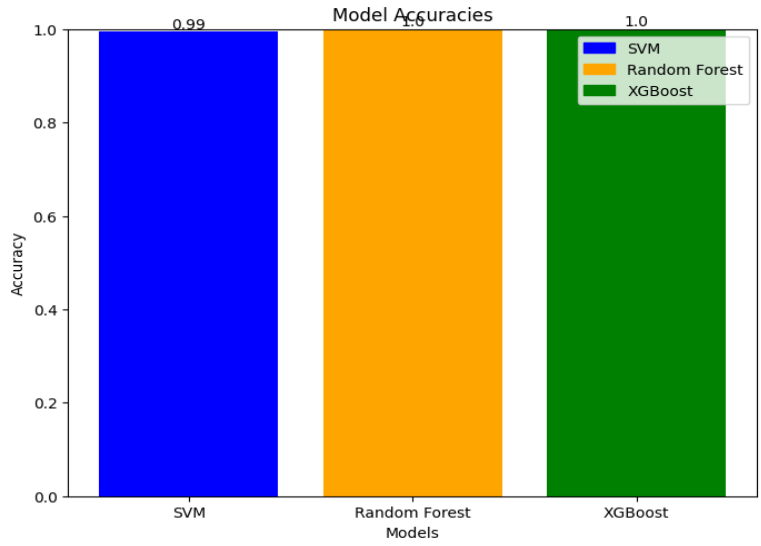

Fig 7(a) plot of Accuracies for three individual algorithms before ensemble in bar chart

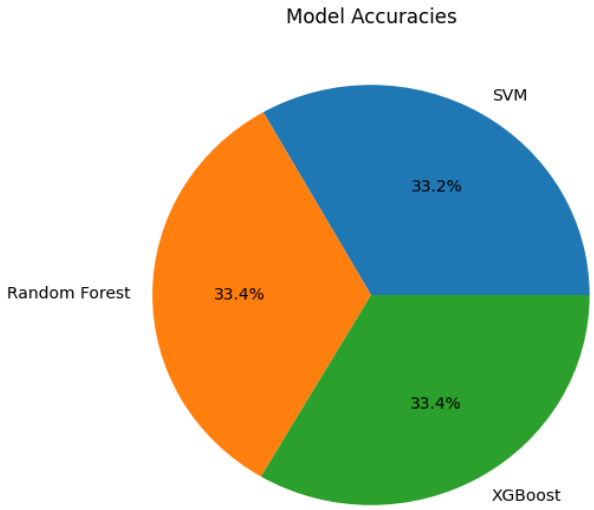

Fig 7(b) plot of Accuracies for individual algorithm before ensemble in pie chart.

Confusion Metrics

F-measure was used to calculate the accuracy, recall, and precision of machine learning algorithms. A Confusion matrix can be used to calculate each of the metrics given. According to these metrics, a model performance was assessed. Resampling is crucial, as evidenced by the results of model testing on both original and resampled data. Since the test set represents 30% of the entire dataset, there are 85118 total samples out of which 134 are fraud transactions. The labels of actual vs. predicted were summarized using confusion metrics, where the X-axis represents the predicted label and the Y-axis represents the actual label:

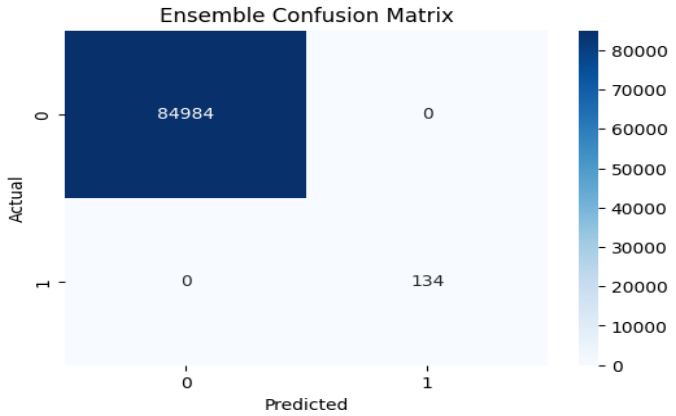

Fig 8: plot of confusion matrix for the ensemble learning

Table 3: The result of confusion matrix

True Class		**Actual Class**		
		Legit (0)	Fraud (1)	
Predicted Class	Legit (0)	84984	0	**84984**
	Fraud (1)	0	134	**134**
		84984	**134**	**85118**

The performance outcome of a binary classification model is shown in **Table 3** above. It contrasts the model-predicted class with the genuine class (actual class) of the data. via way of the matrix. The number of occurrences for which the model correctly predicted Legit (0) is shown in the top-left cell (84984). The top-right cell (0) shows that the model did not predict any occurrences as Fraud (1) in error. There are no cases that the model mistakenly predicted as Legit (0), as indicated by the bottom-left cell (0). The number of instances that the model properly identified as Fraud (1) is shown in the bottom-right cell (134).

Performance Evaluation Metrics

In this section the performance evaluation metrics of the classifier are reported. The metrics include accuracy, precision, recall, and F1 score. The calculations for each metric are as follows:

Accuracy = (84984 + 0) / 84984 = **1.0**
Precision = 84984 / (84984 + 0) = **1.0**
Recall = 84984 / (84984 + 0) = **1.0**
F1 Score = (2 * 1 * 1) / (1 + 1) = **1.0**

These performance evaluation metrics indicate excellent results, with all metrics achieving a perfect score of 1.0. It demonstrates the classifier's high accuracy, precision, recall, and F1 score in correctly classifying instances.

Receiver Operating Characteristics (ROC) Curve

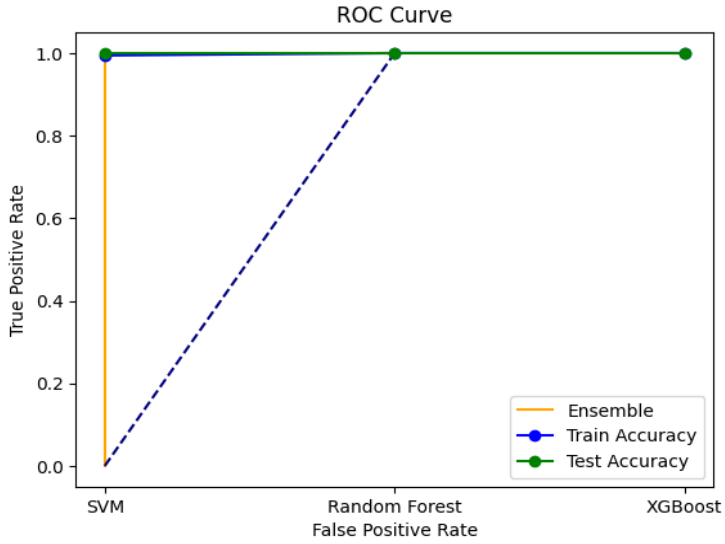

Fig 9(a) plot of Area Under Curve (AUC) for all three algorithms.

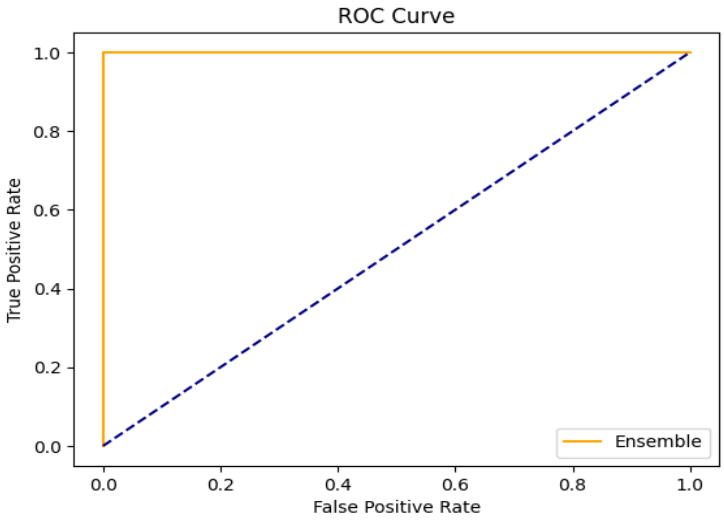

Fig 9(b) plot of Area Under Curve (AUC) for the ensemble learning algorithms

The true positive rate (TPR), also known as recall, represents the proportion of actual positive instances that are correctly identified by the model. Conversely, the false positive rate (FPR) quantifies the probability of a false alarm, which indicates the proportion of negative instances incorrectly classified as positive. The Receiver Operating Characteristic (ROC) curve plots the true positive rate against the false positive rate, considering different decision thresholds for classifying positive instances. The decision threshold (c) is a constant that determines the classification boundary between positive and negative predictions.

The performance of the model is assessed by calculating the area under the ROC curve (AUC). The AUC provides a single metric that summarizes the overall performance of the model in distinguishing between positive and negative instances. A perfect AUC score of 1.0 indicates that the model achieves a true positive rate of 100% without any false alarms (FPR = 0). Having a 100% AUC indicates that the model performs exceptionally well, achieving perfect discrimination between positive and negative instances. This implies that the model correctly classifies all positive instances without incorrectly labeling any negative instances as positive. Such a high AUC score reflects the model's strong predictive capability and robust performance.

Table 4: The performance grade of AUC value range

S/N	AUC value range	Performance grade
1	AUC ≥ 0.9	Excellent performance
2	0.8 ≤ AUC < 0.9	Very good performance
3	0.7 ≤ AUC < 0.8	Good performance
4	0.6 ≤ AUC < 0.7	Fair performance
5	AUC < 0.6	Poor performance

The above **Table 4** of AUC value range is used to assess the model's performance in distinguishing between positive and negative instances. The performance grades are assigned based on the AUC value falling within a specific range. If the AUC value is greater than or equal to 0.9, it is considered to have excellent performance. AUC values between 0.8 and less than 0.9 indicate very good performance. AUC values ranging from 0.7 to less than 0.8 are classified as good performance. AUC values falling between 0.6 and less than 0.7 represent fair performance. Lastly, AUC values below 0.6 are classified as poor performance. By referring to this table, researchers can easily determine the performance grade of their model based on the calculated AUC value, providing a clear assessment of its discrimination power in classifying positive and negative instances.

Table 5: Execution Time of the models

Classifier	Running Time
Random Forest	12 min.
Extreme Gradient Boosting	7 min.
Support Vector Machine	1 min.

The execution time of the models was recorded and summarized in Table 5. The Random Forest model took approximately 12 minutes to complete its execution. The Extreme Gradient Boosting model required around 7 minutes for execution and the Support Vector Machine model executed within just 1 minute. The successful execution of each model within their respective time frames demonstrates the efficiency of the implemented algorithms. Researchers can utilize this information to assess the computational requirements and efficiency of the models, aiding in the selection of appropriate algorithms for future experiments or real-time applications.

CONCLUSION

This study introduces an innovative method for identifying credit card fraud utilizing ensemble learning, more specifically boosting approaches. The suggested model performs exceptionally well because it incorporates the Random Forest (RF), Support Vector Machine (SVM), and XGBoost algorithms. A perfect receiver operating characteristic (ROC) score of 1.0 indicates accurate identification of fraudulent and legitimate transactions with zero False Positive Rate (FPR) and False Negative Rate (FNR) in an evaluation using a comprehensive dataset of credit card transactions made by European cardholders over a two-day period in September 2013 (sourced from Kaggle.com). By taking into account numerous evaluation criteria and utilizing the advantages of each base classifier through ensemble learning, the research tackles the shortcomings of earlier studies. This method performs better than earlier studies, which mainly concentrated on accuracy. The performance of the model has significantly improved as a result of using resampling techniques like the Synthetic Minority Oversampling Technique (SMOTE) and Random Undersampling Technique (RUS) to address the unbalanced nature of the data. Given that they offer a solid and trustworthy solution for credit card fraud detection, these findings have significance for the financial sector. The use of the Kaggle.com dataset and ensemble learning techniques improves the results' realism and generalizability.

Future Scope

Future research directions may explore additional ensemble methods and large-scale datasets to advance credit card fraud detection systems and bolster security measures in the financial sector.

REFERENCES:

[1] Jain, Y., Tiwari, N., Dubey, S., & Jain, S. (2019). A comparative analysis of various credit card fraud detection techniques. *Int J Recent Technol Eng*, *7*(5S2), 402-407.

[2] Nguyen, T. T., Tahir, H., Abdelrazek, M., & Babar, A. (2020). Deep learning methods for credit card fraud detection. *arXiv preprint arXiv:2012.03754*.

[3] Hilal, W., Gadsden, S. A., & Yawney, J. (2021). A review of anomaly detection techniques and applications in financial fraud. *Expert Systems with Applications*, 116429.

[4] Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, *479*, 448-455.

[5] Chen, J. I. Z., & Lai, K. L. (2021). Deep convolution neural network model for credit-card fraud detection and alert. *Journal of Artificial Intelligence*, *3*(02), 101-112.

[6] Nami, S., & Shajari, M. (2018). Cost-sensitive payment card fraud detection based on dynamic random forest and k-nearest neighbors. *Expert Systems with Applications*, *110*, 381-392.

[7] Yee, O. S., Sagadevan, S., & Malim, N. H. A. H. (2018). Credit card fraud detection using machine learning as data mining technique. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *10*(1-4), 23-27.

[8] Sharma, D., & Kumar, N. (2017). A review on machine learning algorithms, tasks and applications. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *6*(10), 2278-1323.

[9] Bur, A. M., Shew, M., & New, J. (2019). Artificial intelligence for the otolaryngologist: a state of the art review. *Otolaryngology–Head and Neck Surgery*, *160*(4), 603-611.

[10] Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1249.

[11] Murauer, B., & Specht, G. (2018, April). Detecting music genre using extreme gradient boosting. In *Companion proceedings of the the web conference 2018* (pp. 1923-1927).

[12] Meng, C., Zhou, L., & Liu, B. (2020, August). A case study in credit fraud detection with SMOTE and XGboost. In *Journal of Physics: Conference Series* (Vol. 1601, No. 5, p. 052016). IOP Publishing.

[13] Dhankhad, S., Mohammed, E., & Far, B. (2018, July). Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study. In *2018 IEEE international conference on information reuse and integration (IRI)* (pp. 122-125). IEEE.

[14] Mbona, I., & Eloff, J. H. (2022). Feature selection using Benford's law to support detection of malicious social media bots. *Information Sciences*, *582*, 369-381.

[15] Gyamfi, N. K., & Abdulai, J. D. (2018, November). Bank fraud detection using support vector machine. In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 37-41). IEEE.

[16] Bao, Y., Hilary, G., & Ke, B. (2022). Artificial intelligence and fraud detection. In *Innovative technology at the interface of finance and operations* (pp. 223-247). Springer, Cham.

[17] KJ GamTech Studio(2021, December 18). *Data Science: Credit Card Fraud Detection Project*[Video].https://www.youtube.com/watch?v=pohp5AZrp3I&t=447s