SHOULD THE GOVERNMENT PROMOTE OR CONTROL DEVELOPMENTS IN

MACHINE LEARNING AND AI IN CREDIT CARD FRAUD?

AGBOOLA ISAAC OLUWATOMIWA

31th MARCH 2024

ABSTRACT

This topic shows how the government should promote, control and develop machine learning and Artificial intelligence in fraud by focusing on the historical evolution of machine learning and AI context, various regulatory approaches, and their respective strengths and weaknesses.

Ultimately, this assessment explains the government's involvement in machine learning and AI development, advocating for a nuanced approach that acknowledges both the potential benefits and risks associated with these transformative technologies. As society stands at the precipice of an AI-driven future, understanding the role of government in steering this course is of paramount importance.

This will be done by making use of a dataset from Kaggle and using Jupyter Notebook and Python to code by working on credit card fraud transactions by grouping them into different categories. The training test data will be used for the total entities, statistics, accuracy, precision, recall, Auc-Roc, F1_Score and confusion matrix of different AI models of the fraudster and non-fraudster. The result and discussion represent the visualization of data by using graphs to describe the machine learning and AI algorithm of fraud and non-fraud credit card fraud transactions.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND

The world is becoming digitalized every day and is changing the future of the government, with the "use of machine learning and Artificial intelligence (AI) enabling us to reduce our work daily in companies, industries, banking, agriculture, and finance".

Artificial Intelligence (AI) is the subset of machine learning that allows the machine to interact with each other from data (Vrnika Jain & Palki, 2021). It is a concept of using "computer science to stimulate human thinking by using computers, programming, and software developers (Omar, 2022), to solve complex problems" (Aarvik, 2019), this can be done by applying tools such as Machine learning, Natural language processing, Computer Vision, Deep learning, and Neural Networks.

According to (Longley, 2022), fraud is an illegal act and criminal, this is while the government can use machine learning to know the minimum and maximum amount of money being taken, the time when the fraud started and ended, to solve the problem by providing the necessary facility, having a network device to detect frauded with the use of face detection, NIN, passport, KYC, and human recognition (Neil Shah, et al., 2021). This can control and promote the development of machine learning and Artificial intelligence in the credit card in an organization in the ecosystem of the data ecology by using machine learning to detect to prevent fraud and cybersecurity attacks (Shah, 2021).

Implementing fraud to monitor the development of fraudulent activities in credit card machine learning, algorithms, training data, performance models, and minimum, and maximum value is very crucial for the government to detect.

The applications can reduce scammer, banking, e-commerce, and telecommunication such as corruption, public funds, fake text messages, phone calls, links, and Facebook, that are legit to defraud everyone in society (Pradheepan & Neamat , 2019). “This can be done by ensuring that the government should stop the fraudsters by using machine learning” and employing more officials and securities agencies to solve the problem to develop a society that can impact data ecology in the ecosystem.

1.2 THESIS STATEMENT

When working with the data of fraud on the procedures of machine learning and AI it takes a lot of time manually and the use of code is more faster. The governments should actively promote innovation and ensure ethical guidelines, others contend that this can seriously control the potential risk. There are lots of questions that arise from this research including the following:

1. What should be the role of the government in regulating machine learning and artificial intelligence development?
2. How do fraud, machine learning and artificial come into existence?
3. How can the government control, promote and develop machine learning and artificial intelligence in fraud when there is a large volume of data?

This essay will navigate these opposing viewpoints, critically evaluating the merits and offering insights into the implications for data science, data ecology, and society. As we focus on this through the complex landscape of machine learning, artificial intelligence, and governance, it is crucial to keep in mind the broader implications of our findings.

CHAPTER TWO

LITERATURE REVIEW

2.1 HISTORICAL OVERVIEW AND EVOLUTION OF MACHINE LEARNING AND AI?

Fraud detection affects many lives and properties, especially in social devices, bank, government sectors, and phone devices especially in the area of credit cards, making calls, sending messages on SMS, and most problems on mobile communication this has been in existence long ago (P. Barson, et al., n.d.). Machine learning used to detect credit card fraud can be divided into two categories supervised machine learning and unsupervised machine learning (Musibau Adekunle Ibrahim & Patric Ozoh , 2023).

Supervised machine learning is the “building of statistical model” that contains large amounts of data (Bart van Liebergen, n.d.) that give the accurate result of “the fraudulent transaction which has input and the corresponding output variable is known”. It can be used for identifying fraudulent transactions. As the technology is increasing in the banking sector fraudsters always find other means for committing fraud. Due to the small percentage of fraudulent transactions, these models are built on highly imbalanced datasets, with methods such as Support Vector Machines, Logistic Regression and Decision Trees used for modelling.

Unsupervised machine learning has only input data and no corresponding output variable, that shows pattern, and structure about the “data through analysis of multiple features, it can be done by mapping new cases”, investigating fraud in a particular location (Mahendra Prasad Nath, et al., 2018). This can take the features about the transaction and history of a customer to classify them into one of the clusters. K-means clustering and K-neighbor classifiers are examples of unsupervised machine learning. It can be used to detect the amount spent, and time of the transactions this Non-Principle Component Analysis builds up a “model for a legitimate account which can be used to identify anomalies with the new record” (Mahendra Prasad Nath, et al., 2018).

The advantages and disadvantages of supervised machine learning are, that they can be used for labelled datasets, and it takes a lot of time when needed to predict data. For Unsupervised machine learning the advantages and disadvantages are that the technique is used for jobs, and when a dataset is not mapped it can give problems.

2.2 THE MILESTONES AND TIMELINE OF MACHINE LEARNING

In 1950, Alang Turning a computer scientist and mathematician created a Turning Test to predict if intelligence is real in computers to figure out if a machine can behave like a human and the responses they act such as speaking and feeling (Alexander, 2020). He published an article in his paper answering the question, “Can machines think?”.

Arthur Samuel was an American pioneer in the field of computer science, a programmer and he wrote the first computer learning program in 1952, the research was done on computer-based gaming and recognition. He defined machine learning, and created the phrase “Machine Learning”. "The field of study that gives computers the ability to learn without being explicitly programmed," (Dr, et al., 2023) . He made a various contribution to computer science including computer graphics, artificial intelligence, and computer chess.

American psychologist and computer scientist Frank Rosenblatt, conducted research at Cornell’s Department of Neurophysiology, where he explored neurobiology and the functioning of the brain also designed the first neural network for computers that formed the ideas of the works of the human nervous system (Alexander, 2020), artificial intelligence and neural network systems. Computational models were developed that deals with the behaviour of neurons and also led to the creation of perception which simulated the thought processes of the human brain and machine learning model in 1957 and also demonstrated the potential practical applications of neural networks (Dr, et al., 2023).

Bernard Widrow and Madeline are from the United States of America studying as a professor of electrical engineering at the University of Stanford, the research was done to develop signal processing in the field of Geophysics, adaptive antennas, and adaptive filtering. They created two neural network models named Adaline in the year 1959, that could detect binary patterns and input signals having values of +1 and -1 output so that they can be read a bit (0 or 1) from phone lines and other network devices (Bernard & Marcian, n.d.), it was done to illustrate adaptive behaviour and artificial learning by ensuring how to switch circuit worked (Bernard & Marcian, n.d.).

In 1967, the “Nearest neighbour” algorithm was the first algorithm that was used to map a route for travelling salesmen by starting at a random city by ensuring all cities are visited. It is a non-parametric, supervised learning classifier that can be used to solve problems especially large datasets by finding the solution of two or more data sets (Dr, et al., 2023).

CHAPTER THREE

METHODOLOGY

3.1 APPROACHES TO MACHINE LEARNING AND REVOLUTION

The use of “banking services, mobile banking applications and payments are done through credit cards are increasing rapidly in the world” and this can result in fraud due to their challenges, advantages and disadvantages. Machine learning mostly uses the model to detect fraud in credit cards and produces a noticeable result (Rao, et al., 2021)

There is 284807 total number of credit card fraud transactions, having 492 fraud transactions, and 284315 not fraud transactions. The percentage of fraud and not fraud is 0.2% and 99.8%. From this transaction, the value that has the highest number of fraudulent transactions is not fraud and it has a percentage of 99.8%. There are two integers (binary) values to represent both of them which are 0 and 1.

According to (MindStream, 2024), the machine learning revolution is developing and increasing every day in healthcare, Education, Agriculture, Government, online (Facebook, WhatsApp, Telegram), and social media. It has made work easy and has a significant impact on society. Some techniques can be used for detecting credit card systems for fraud, the different machine learning algorithms consist of Logistic Regression and Random Forest to determine the best way to identify fraudulent transactions (Lakshmi S V & Selvani, 2018)

DESCRIPTION	FRAUD	NON-FRAUD
TRANSACTION	492	284315
PERCENTAGE (%)	0.2	99.8

Table 1: The Number of Fraudulent Transactions.

3.2 REGULATORY FRAMEWORK

A. Logistic regression is a supervised learning and it is useful for fraud detection that is used to predict binary values in a given set of independent variables (Faraji, 2022). It is also used to predict features that a transaction is fraudulent.

B. Random Forest is mostly used for credit card detection and is a supervised machine learning algorithm used in machine learning techniques that acts as a binary classifier that consists of 0 or 1 (K.Ratna, et al., n.d.). It is used for online and offline credit card transactions.

C. Support Vector Machine: SVM is another supervised method that is applicable for classification and regression problems. SVM can be applied to detect different types of fraud in banking industries and create the decision boundary to segregate the classes.

D. Decision Trees: It is used for classification problems. It is tree-likes that build a structure that divides the dataset depending on the feature value (Faraji, 2022). It can be used for building several trees.

E. K- Neighbor Classifier: This can be used to find the unknown feature of testing data and several "anomaly detection techniques and it is mostly used for credit card fraud detection". It is part of supervised machine learning (M. Ummul Safa & R. M. Ganga, 2019).

3.3 MACHINE LEARNING ALGORITHM

1. Get the dataset from the Kaggle.

2. Import the libraries of the dataset.

3. Calculate the statistics of the dataset.

4. Divide the dataset into two parts i.e., Train dataset and Test dataset

5. Calculate performance metrics and confusion matrix of each variable.

6. Plot the graph for the AUC-ROC curve and recall-precision curve.

3.4 DATA SET

A credit card fraud dataset can be obtained from Kaggle that contains a combination of fraud and non-fraud transactions. CSV files are mostly used formats for machine learning data. The total number of transactions for credit card detection is 284807 and the number of frauds is 492 and non-fraud is 284315. The dataset is highly imbalanced; fraud has a percentage of 0.2% and non-fraud has a percentage of 99.8%. The dataset features are transformed using the principal component analysis (PCA) such as V1, V2, V3 ……..V28 are PCA features and Time, Amount and Class are non-principal component analysis features (Vaishnavi & Geetha, 2019).

S/N	FEATURE	DESCRIPTION
1	Time	Time in seconds to specify the elapses between the current transaction and the first transaction
2	Amount	It is the transaction Amount
3	Class	1 = It indicates Fraud 0 = It indicates No Fraud

Table 2: The Dataset of Non-Principal Analysis.

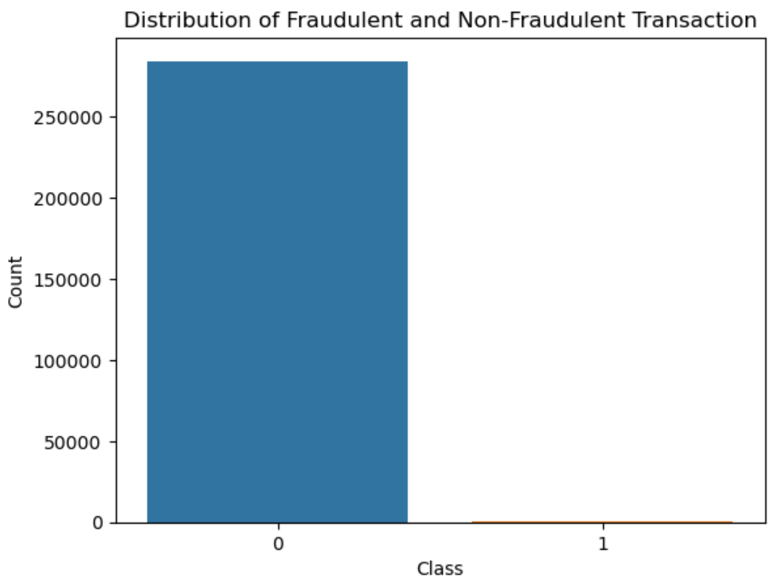

Figure 1: The bar chart shows the highest number of non-fraud and the lowest number of fraud.

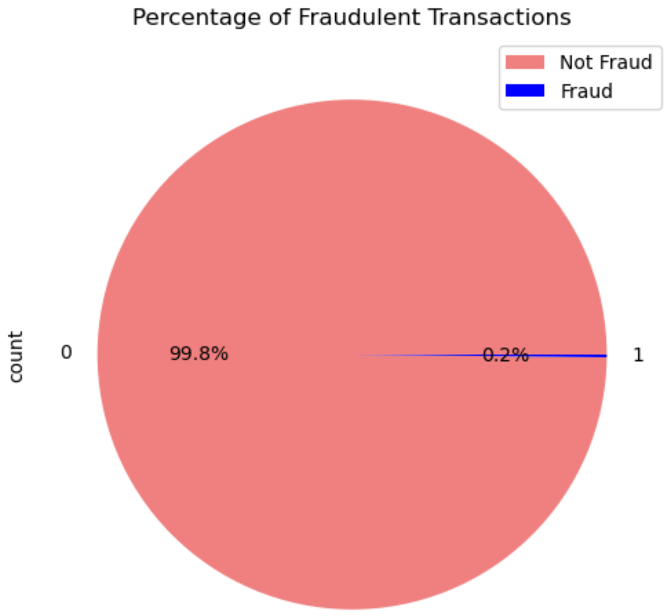

Figure 2: The pie chart shows the percentage of fraud and not fraud transactions.

3.5 PERFORMANCE METRICS

The confusion matrix is a part of performance metrics that consists of two by two matrix that have four quadrants that are produced by the binary classifier. Accuracy, precision, recall, and F1_Score are obtained from the confusion matrix.

1. Accuracy: It is the sum of True positive and True Negative divided by the total sum of True Positive, False Positive, True Negative and False Negative

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \qquad \text{equation 1}$$

2. Precision: It is the ratio of the true positives divided by the sum of the true positives and false positives.

$$Precision = \frac{TP}{TP+FP} \qquad \text{equation 2}$$

3. Recall: It is the ratio of true positives divided by the sum of true positives and false negatives.

$$Recall = \frac{TP}{TP+FP} \qquad \text{equation 3}$$

7. F1_Score: It is a weighted harmonic mean of precision and recall such that the best score is 1.0 and the worst value is 0.0, F1 scores are considered lower than the accuracy measure.

$$F1 = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \qquad \text{equation 4}$$

	Fraud (Positive)	Not Fraud (Negative)
Fraud (Positive)	True Positive	False Negative
Not Fraud (Negative)	False Positive	True Negative

Table 3: Confusion Matrix.

TRUE POSITIVE (TP): This shows the actual number of fraudulent transactions predicted as fraud.

TRUE NEGATIVE (TN): This shows that the actual number of fraudulent transactions is predicted as not fraud.

FALSE POSITIVE (FP): The number of non-fraud transactions predicted as fraud**.**

FALSE NEGATIVE (FN): The number of fraud transactions predicted as not fraud.

3.6 STRENGTHS AND WEAKNESS

The government identified that working with credit card detection in the banking system by using machine learning algorithms and models is a big challenge and can lead to problems. This can be solved by further practices, research and investigation to know which models of machine learning to use in Data science.

CHAPTER FOUR

THE IMPACT AI ON DATA ECOLOGY AND SOCIETY

Fraud has been the greatest concern for financial institutions, banks, e-commerce websites and payment processing platforms that cause financial and affect millions of people globally (Osmond, et al., 2023). The growth of e-commerce and the threat of card fraud have increased significantly, affecting the growth of the economy and leading to great losses. The government needs to take risks and proactive measures to control and measure the financial cost so that the agencies can conduct their business effectively (Yolanda Aji Abei, 2021). The impact of credit card fraud, can cause problems and lead to identity theft, loss of ATM card, and damage to credit score. To be able to solve this problem it takes a longer process of “recovering their identities and finances”.

In the United States of America (USA) fraud has been increasing every year (Fumiko Hayashi, n.d.). The rate of identity theft increased significantly from 2019 to 2023, growing from 2.7 million to 4.2 million reports. In 2023, there were 53% more reported cases of credit card fraud than that in 2019.

In Nigeria, the Central Bank of Nigeria (CBN) reported that the rate of fraud has increased, banks recorded 78,584 cases in a year. Second half of 2022 and the second half of 2023, there were 10,098 fraud cases involving N1.95 billion recorded on POS channels.

According to “Forgeries in Nigeria Banks” in the first quarter of 2023, electronic transactions grew from 44.84 per cent to N126.73 trillion starting in 2022.

The positive impacts are Strengthening regulations and policies, Collaboration and information sharing, Awareness and Education, and using personal identification numbers (PINs) to prevent stolen cards (Fumiko Hayashi, n.d.). The negative impact is Financial loss, identity theft legal issues, decreased consumer confidence, business, and economic, distress.

CHAPTER FIVE

CONCLUSION AND HYPOTHESIS FOR FUTURE RESEARCH

5.1 RESULT AND DISCUSSION

A. THE DATASET OF MACHINE LEARNING

Table 4 shows the minimum amount maximum amount of money spent on credit cards for fraudulent transactions.

	Time	**Amount**	**Class**
Count	284807.000000	284807.000000	284807.000000
Mean	94813.859575	88.349619	0.001727
Std	47488.145955	250.120109	0.041527
Min	0.000000	0.000000	0.000000
25%	54201.500000	5.600000	0.000000
50%	84692.000000	22.000000	0.000000
75%	139320.500000	77.165000	0.000000
Max	172792.000000	25691.160000	1.000000

Table 4: The Descriptive analysis of fraudulent transactions.

B. CONFUSION MATRIX

Figure 3 shows the confusion matrix of the Decision Tree Classifier. In Fraudulent transactions it has a precision of 0.73, recall of 0.84 and f1_score of 0.78, this indicates that the model performed well in detecting fraudulent transactions. The precision for not fraudulent transactions is 1.00 indicating a high proportion of correctly predicted non-fraudulent transactions.

```
                    Decision TreeClassifier:
                       [[56833    31]
                        [   16    82]]
                          precision    recall  f1-score   support


   Not fraud               1.00      1.00      1.00     56864
     Fraud                 0.73      0.84      0.78        98


    accuracy                                   1.00     56962
   macro avg             0.86      0.92      0.89     56962
weighted avg               1.00      1.00      1.00     56962
```

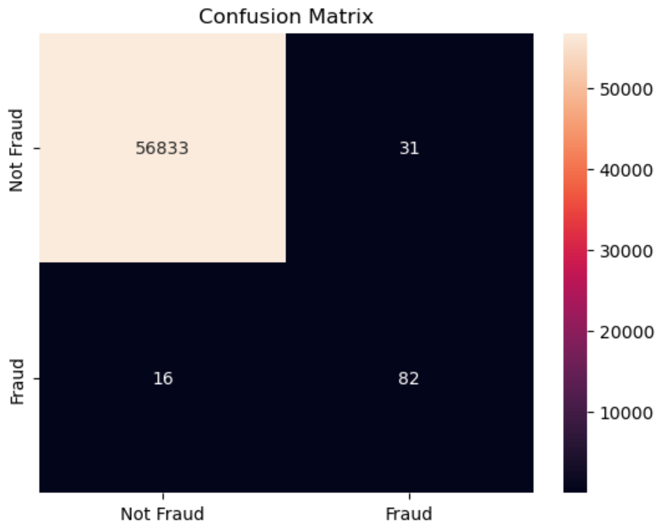

Figure 3: Confusion Matrix of Decision Tree Classifier Algorithm

Figure 4 shows the confusion matrix of logistic regression. In Fraudulent transactions it has a precision of 0.86, recall of 0.58 and f1_score of 0.70, this indicates that the model performed well in detecting fraudulent transactions. The precision for not fraudulent transactions is 1.00 indicating a high proportion of correctly predicted non-fraudulent transactions.

```
                 Logistic Regression:
                   [[56855    9]
                    [   41   57]]
                     precision   recall  f1-score   support


Not fraud          1.00     1.00     1.00    56864
 Fraud              0.86     0.58     0.70       98


accuracy                                   1.00    56962
```

```
        macro avg        0.93     0.79     0.85     56962

weighted avg                 1.00     1.00     1.00     56962
```

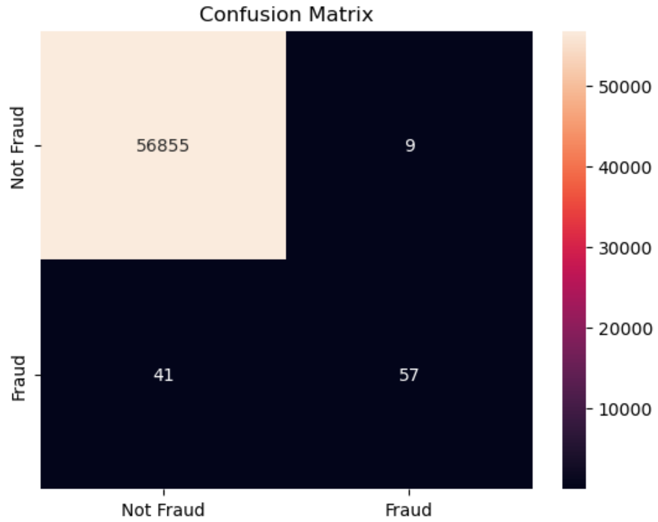

Figure 4: Confusion Matrix of Logistic Regression Algorithm

Figure 5 shows the confusion matrix of Random Forest. In Fraudulent transactions it has a precision of 0.97, recall of 0.80 and f1_score of 0.88, this indicates that the model performed well in detecting fraudulent transactions. The precision for not fraudulent transactions is 1.00 indicating a high proportion of correctly predicted non-fraudulent transactions.

```
                        Random Forest:

                        [[56862     2]

                         [   20    78]]

                          precision    recall  f1-score   support


       Not fraud              1.00     1.00     1.00     56864
```

```
Fraud                    0.97    0.80    0.88     98


     accuracy                           1.00    56962
    macro avg          0.99    0.90    0.94    56962
weighted avg             1.00    1.00    1.00    56962
```

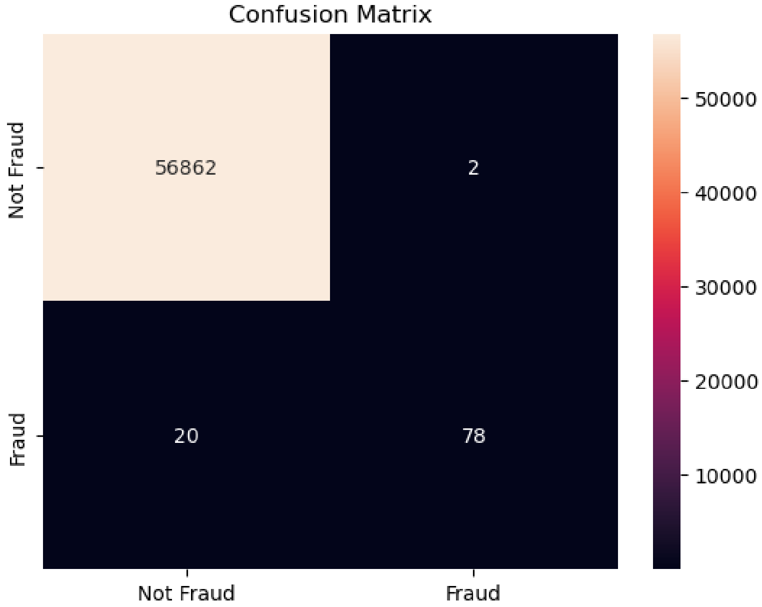

Figure 5: Confusion Matrix of Random Forest Algorithm

Figure 6 shows the confusion matrix of Support Vector Machines. In Fraudulent transactions it has a precision of 0.97, recall of 0.62 and f1_score of 0.76, this indicates that the model performed well in detecting fraudulent transactions. The precision for not fraudulent transactions is 1.00 indicating a high proportion of correctly predicted non-fraudulent transactions.

Support Vector Machine:

```
                  [[56862     2]
                   [   37    61]]
                    precision    recall  f1-score   support


     Not fraud           1.00      1.00      1.00     56864
  Fraud                         0.97      0.62      0.76        98


     accuracy                              1.00     56962
     macro avg           0.98      0.81      0.88     56962
     weighted avg        1.00      1.00      1.00     56962
```

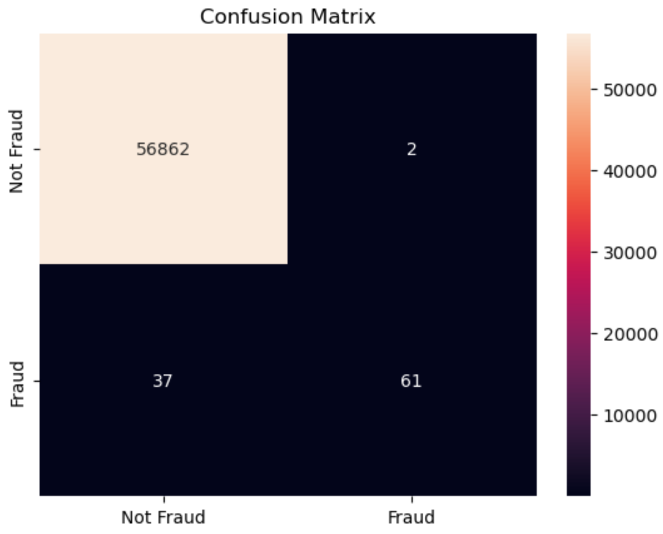

Figure 6: Confusion Matrix of Support Vector Machine Algorithm

Figure 7 shows the confusion matrix of the K-Neighbors Classifier. In Fraudulent transactions it has a precision of 0.94, recall of 0.78 and f1_score of 0.85, this indicates that the model performed well in detecting fraudulent transactions. The precision for not fraudulent transactions is 1.00 indicating a high proportion of correctly predicted non-fraudulent transactions.

```
                         KNeighborsClassifier:

                           [[56859     5]

                            [   22    76]]

                           precision    recall  f1-score   support


           Not fraud            1.00      1.00      1.00     56864

             Fraud              0.94      0.78      0.85        98


           accuracy                                 1.00     56962

           macro avg            0.97      0.89      0.92     56962

    weighted avg                        1.00      1.00      1.00     56962
```

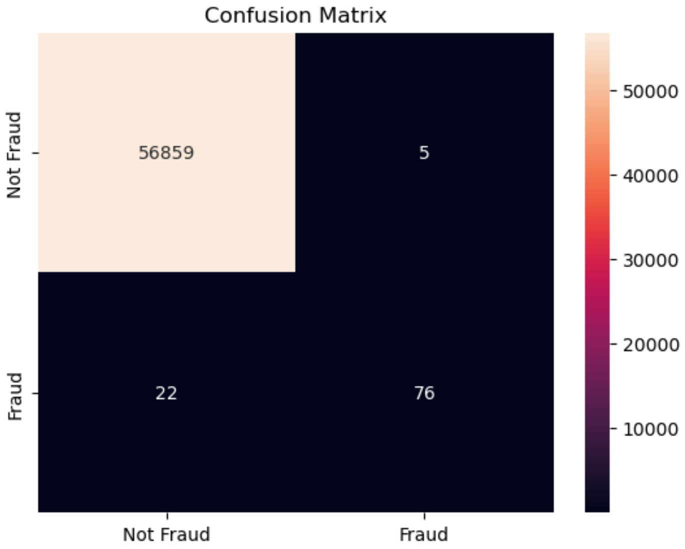

Figure 7: K- Neighbor Classifier Algorithm

C. **PERFORMANCE MATRIX**

Machine Learning Algorithm	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1_Score (%)**	**Roc-Auc (%)**
Logistic Regression	99.91	86.36	58.16	69.51	79.07
Random Forest	99.96	97.50	79.59	87.64	89.79
Support Vector Machine	99.93	96.83	62.24	75.78	81.12
K-Neighbor Classifier	99.95	93.83	77.55	84.92	88.77
Decision Tree Classifier	99.92	72.57	83.67	77.73	91.81

Table 5: The performance matrix of the Algorithm

1. The Random Forest demonstrated the highest accuracy at 99.96%, closely followed by the K-Neighbor Classifier with 99.95%, the lowest accuracy is Logistic Regression. These algorithms demonstrated that Random Forest is mostly used for the detection of credit card fraud.

2. Random Forest exhibited a notable precision of 97.50% which is the highest among the five machine learning algorithms and ensures comprehensive fraud detection.

3. The Decision Tree Classifier exhibited a notable recall of 83.67%, indicating its capability to minimize false positives while capturing a higher proportion of actual fraudulent transactions.

4. Random Forest achieved an F1-Score of 87.64%, reflecting its balanced performance in harmonizing Accuracy and Precision. This balanced measure demonstrates the algorithm's ability to maintain a reasonable trade-off between false positives and false negatives in identifying fraudulent activities.

5. The Decision Tree Classifier exhibited a notable ROC-AUC of 91.81% and Logistic Regression has the lowest score of 79.07%.

6. Logistic Regression demonstrated slightly lower metrics across Accuracy, Recall, F1_Score and ROC-AUC having 99.91%, 58.16%, 69.51% and 79.07% compared to supervised learning algorithms, it's important to note that clustering algorithms like Decision Tree Classifier can provide insights into data patterns but might not be as effective for precise fraud identification.

D. ROC_AUC CURVE

From Figure 10, we see that the Decision Tree Classifier have the highest score of 0.9181 and Logistic Regression have the lowest score of 0.7907

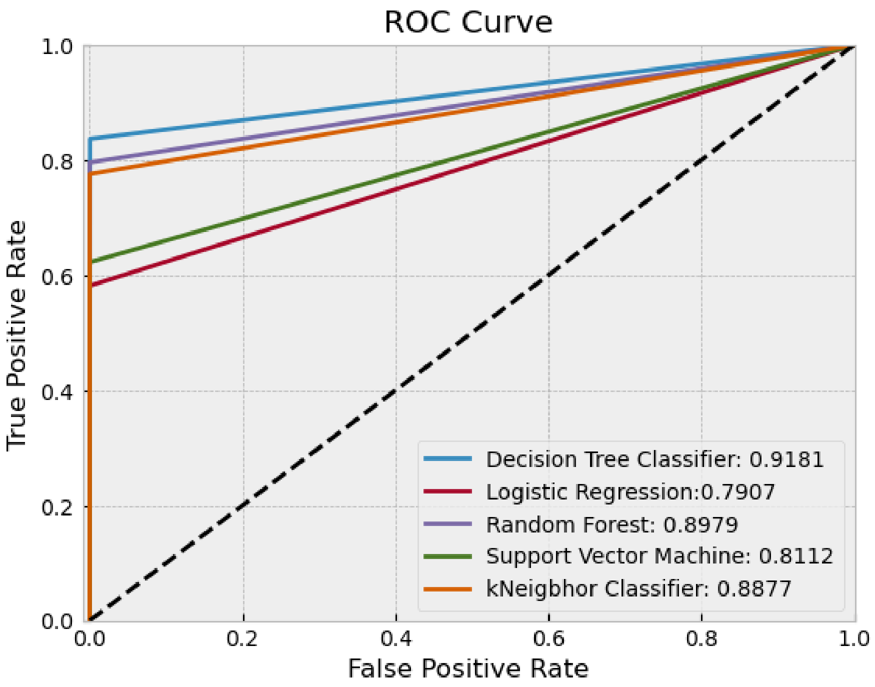

Figure 8: The combination of Machine Learning Algorithms of the AUC_ROC Curve Indicating the True Positive Rate and False Positive Rate.

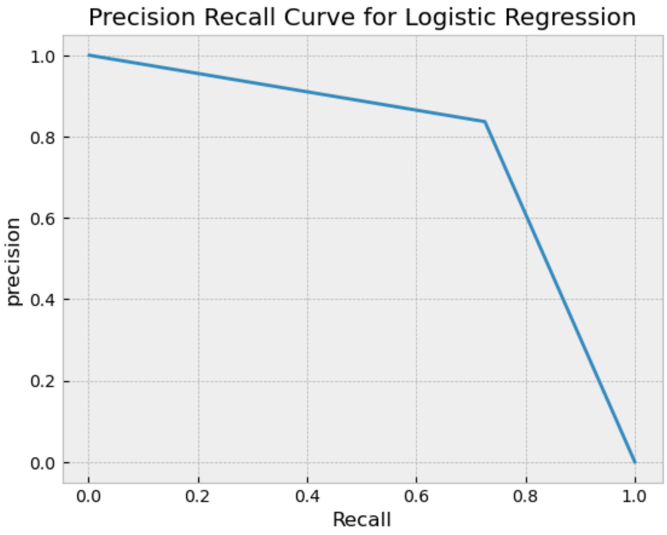

Figure 9: Precision-Recall Curve for Logistic Regression

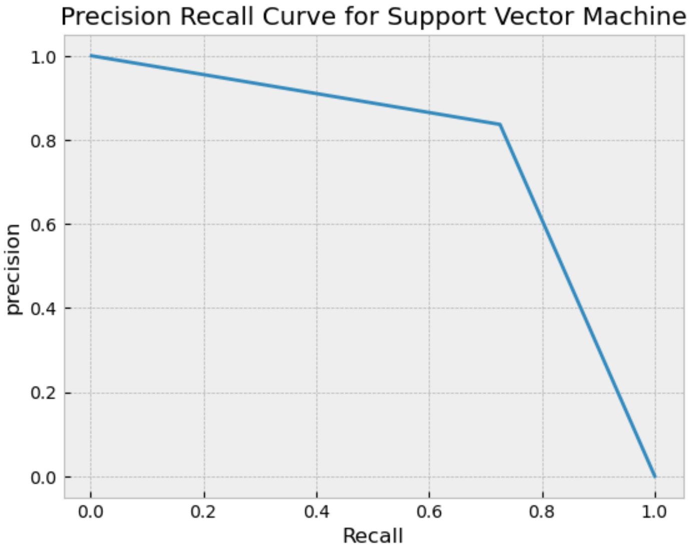

Figure 10: Precision-Recall Curve for Support Vector Machine

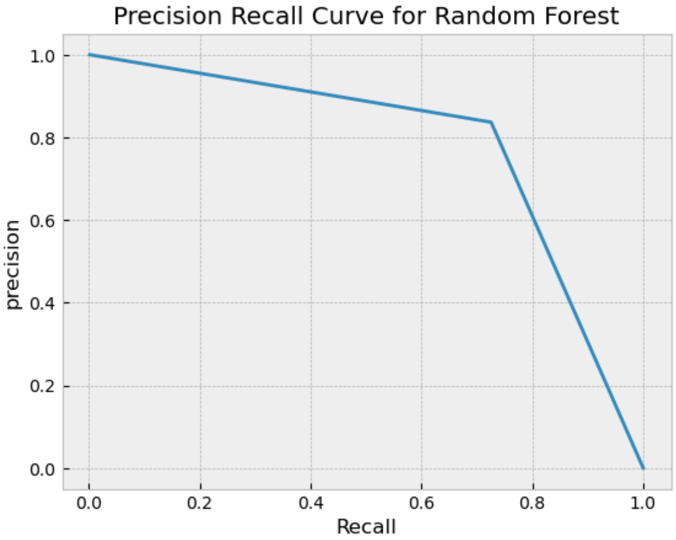

Figure 11: Precision-Recall Curve for Random Forest

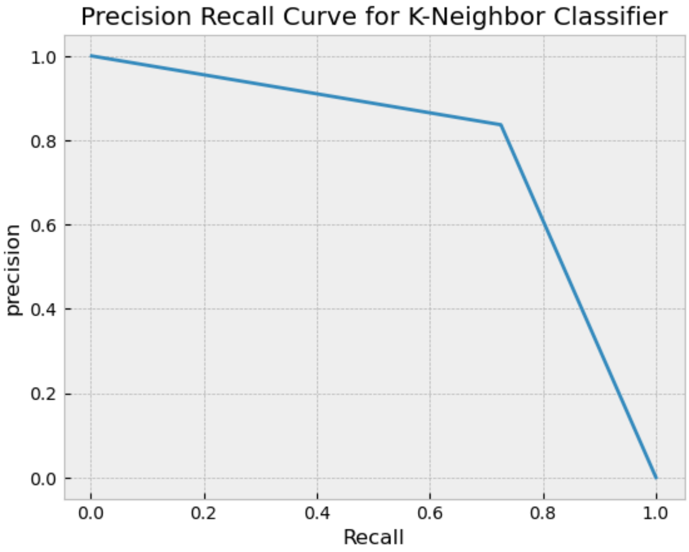

Figure 12: Precision-Recall Curve for K-Neighbor Classifier

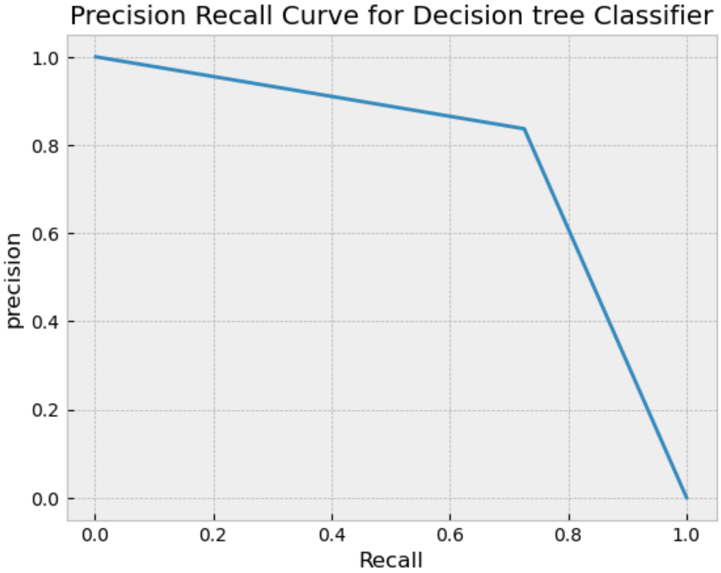

Figure 13: Precision-Recall Curve for Decision Tree Classifier

5.2 LIMITATION OF THE RESEARCH

From this research, we have seen that the Random Forest technique is rapidly increasing and the best, government should make use of it in other to develop machine learning for credit card fraud in the banking system. Also, there could be a problem when the fraudulent transaction is increasing the fraud and not fraud will continue to increase this is why the government should take serious action to control and promote the development of machine learning by ensuring which of the models is the best such as random Forest. For further development, they have to work to solve this problem by using various models and performances to select the best one.

CHAPTER SIX

CONCLUSION

This research has demonstrated how the government should control and promote the development of credit card fraud in the banking system by using machine learning and AI (Kazeem, 2023), for proper investigation of fraudulent and legitimate transactions in society and the environment (Muhammad , et al., n.d.).

Also, to evaluate the dataset of credit card detection from the Kaggle by using test data we have observed that there are 284807 rows and thirty-one (31) column transactions. The fraud has 492 and non-fraud (does not have fraud) of 284315 transactions and there is a small whole number to separate the one that has the fraud and non-fraud transaction by the use of binary integer to represent which are 0 and 1. 0 is for those that do not have fraud and 1 is for those who have fraud transaction. The pie chart was used to know the percentage of fraudulent transactions which are 99.8% and 0.2%. A confusion matrix is a major problem in calculating the data which are divided into four parts: True positive, True negative, False positive, and false negative.

There are different machine learning techniques like logistic regression, Decision Tree, Support Vector Machine, K-neighbor classifier, and Random Forest were used to detect fraud in credit card systems. Precision, recall, f1_Score, ROC-AUC, and accuracy are used to evaluate the performance and models for the proposed system. After implementing an algorithm, Random Forest has the highest performance in the accuracy of 99.6%, precision of 97.50%, and F1_Score of 87.64%. The Government should ensure that random forest is perfectly used to control and develop the credit card in the banking system to decide whether a transaction is legitimate or fraudulent. The results show the accuracy for Logistic Regression, Random Forest, Support Vector Machines, K-Neighbor classifiers, and Decision Tree Classifier are 99.91%, 99.96%, 99.93%, 99.95% and 99.92% respectively. Logistic Regression has the lowest performance in the accuracy of 99.91%, recall of 58.16%, and F1_Score of 69.51, Decision Tree Classifier has the lowest recall of 72.57%. Therefore, the comparative results show that Random Forest performs better than Logistic Regression, Support Vector Machines, Decision Tree Classifiers, and K-neighbor Classifiers techniques (Andrea , et al., 2014). So, the Random Forest technique can be used for credit card detection to promote and control the development of machine learning.

We have used the graphical representation of machine learning algorithms models to predict the sample of data analysis for the fraudulent credit card transaction using the ROC -AUC curve, precision curve, and Recall curve, to tackle the detection problem in the society (Andrea , et al., 2014). For the AUC-ROC, Decision Tree Classifier have the highest percentage of 91.81% followed by Random Forest at 89.79%, Logistic Regression has the lowest percentage of 79.07%.

Machine learning and AI techniques are capable for the government to handling credit card fraudulent cases in an efficient manner. However, there is a limitation occurred that how their performance will be found when the total number of transactions is increased to some extreme level.

REFERENCES

Aarvik, P., 2019. Artificial Intelligence as an anti-Corrruption tool in a development setting? p. 1.

Alexander, T. F., 2020. Early History of Machine Learning.

Andrea, D. P. et al., 2014. Learned lessons in credit card fraud detection from a practitioner's perspective. *Expert Systems with Applications,* p. 4927.

Anon., n.d.

Bart van Liebergen, n.d. Machine Learning: A Revolution in Risk Management and Compliance?. *THE CAPCO INSTITUTE JOURNAL OF FINANCIAL TRANSFORMATION.*

Bernard, W. & Marcian, E. H., n.d. Adaptive Switching Circuits. pp. 96 - 97.

Bonyun, Sean C., 2020. The Federal Register / FIND; Washington. 7 Jan, 85(012), pp. 4-5.

Claire C. Austin, S. B. N. F. C. H. A. L. P. W., n.d. Research Data Repositories: *Review of Current Features, Gap Analysis, and Recommendations for Minimum Requirements,* p. 24.

Dr, J. S., D.G, . S., S, A. & R.S, S., 2023. History of Machine Learning with Its Advantages and Its Applications. *INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY,* August, 10(3), p. 173.

Faraji, Z., 2022. A Review of Machine Learning Applications for Credit Card Fraud Detection with A Case Study. *SEISENSE Journal of Managemen,* Volume 5, pp. 49-59.

Fumiko Hayashi, n.d. Payment Card Fraud Rates in the United States Relative to Other Countries after Migrating to Chip Cards.

Henderson, M. E., 2016. Data Management: A Practical Guide for Librarians. pp. 65 - 66.

K. Ratna, S. V., P.Jyothi, G. Varun, S. & R.Rohith, S. S., n.d. Credit card fraud detection using Machine learning algorithms. *Journal of Research in Humanities and Social Science*, 8(2), pp. 04-11.

Kate Harrison, 2014. A Guide to Data Management in Ecology and Evolution. p. 33.

Kazeem, O., 2023. FRAUD DETECTION USING MACHINE LEARNING. 21 September .p. 18.

Krithika Prakash, N. K.-A. L. I. M. L. a. T. G., 2023. Datasharingandre-useinthetraumaticstress field: An international survey of trauma researchers. *EUROPEAN JOURNAL OF PSYCHOTRAUMATOLOGY,* Volume 14, p. 1.

Lakshmi S V, S. S. & Selvani, D. K., 2018. Machine Learning For Credit Card Fraud Detection System. *International Journal of Applied Engineering Research ISSN 0973-4562,* Volume 13, pp. 16819-16824.

Lisa Nichols, 2020. Federal Government Documents and Publications: Washington. *Request for Public Comment on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting From Federally Funded Research,* 5 Mar. pp. 3-4.

Longley, R., 2022. Definition and Examples of Fraud. 01 December.

M. Ummul Safa & R. M. Ganga, 2019. Credit Card Fraud Detection Using Machine Learning. *International Journal of Research in Engineering, Science and Management,* November.2(11).

Mahendra Prasad Nath, Pravin Pandey, Karthikeyan Somu & Peter Amalraj, 2018. Artificial Intelligence & Machine Learning: The Emerging Milestones in Software Development. *International Journal of Research and Scientific Innovation (IJRSI),* September, V(IX), p. 39.

MindStream, 2024. Navigating the AI and Machine Learning Revolution: A Comparative Overview. 24 February.

Mr. KOLLI NIKHIL, Mr. BISWAMBHARA VINAY MAHARSHI, Mr. KAMIREDDY TANOOJ & Mr. Dr.T.V.S.SriRam, 2023. CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING ALGORITHMS. *Journal of Engineering Sciences,* 14(04), p. 477.

Muhammad , Z. K. et al., n.d. The Performance Analysis of Machine Learning Algorithms for Credit Card Fraud Detection. *Department of Computer Science,* p. 96.

Musibau Adekunle Ibrahim & Patric Ozoh , 2023. Fraud detection model for illegitimate transactions. *Kabale University Interdisciplinary Research Journal (KURJ),* October, 2(2), pp. 21-37.

Neil Shah, Nandish Bhagat & Manan Shah, 2021. Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. Volume Visual Computing for Industry, Biomedicine, and Art.

Omar, B. K., 2022. Should Governments Promote or Control Development in Machine Learning and Artificial Intelligence AI? *International Conference on Research in Education and Science,* 24-27 March, Volume 1, pp. 42-50.

Osmond, et al., 2023. Impact of Fraud and Financial Crimes on the Growth and Development of Nigeria's Economy. *Direct Research Journal of Social Science and Educational Studies,* September, Volume 11(5), pp. 80-87.

P. Barson, et al., n.d. The Detection of Fraud in Mobile Phone Networks.

Pradheepan, R. & Neamat, E. G., 2019. Fraud Detection using Machine Learning and Deep Learning. *International Conference on Computational Intelligence and Knowledge Economy (ICCIKE),* 11-12 December.

Prerna, . P. & Harish, C., n.d. Research Data Management through Research Data Repositories in the field of Computer Sciences. p. 520.

Rao, H. R., Kumar, N. V. & Morarjee, K., 2021. A study on machine learning approaches to detect credit card fraud. 30 July.

Reza Farishy, 2022. The Use of Artificial Intelligence in Banking Industry. *International Journal of Social Service and Research.*

Robert Ulrich, n.d. Registry of Research Data Registry. *Karlsruhe Institute of Technolgy.*

Rosa, V. E. Q. et al., 2023. Architecture of a Data Portal for Publishing and Delivering Open Data for Atmospheric Measurement. *International Journal of Environmental Research and Public Health,* Volume 20.

S., J. I., Sujatha, S. & G., S., n.d. An Efficient Study of Fraud Detection System Using Ml Techniques. *Intelligent Computing and Innovation on Data Science*, p. 66.

Sayali Saraf & Anupama Phakatkar, 2022. Detection of Credit Card Fraud using a Hybrid Ensemble Model. *(IJACSA) International Journal of Advanced Computer Science and Applications,* Volume 13.

Shah, V., 2021. Machine Learning Algorithms for Cybersecurity: Detecting and preventing Threats. 15(04).

Sheo, K., Vinit, K. G., Mohd, D. A. & Rashmi, P., 2022. Credit Card Fraud Detection Using Support Vector Machine. *Proceedings of the 2nd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications, Lecture Notes in Networks and Systems 237,* p. 34.

Stefan Schulz & Anton Möllerke, 2022. MACE – An Open Access Data Repository of Mass Spectra for Chemical Ecology. *Journal of Chemical Ecology (2022) 48:589–597,* 16 May.

Vaishnavi, N. D. & Geetha, S., 2019. Credit Card Fraud Detection using Machine Learning Algorithms. *INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING*, p. 634.

Vrnika Jain & Palki, 2021. Essential Mathematics for Artificial Intelligence (AI) and Machine Learning (ML). *International Journal of Mechanical Engineering,* 3 November.6(3).

Yolanda Aji Abei, 2021. Impact of Internal Control on Fraud Detection and Prevention in Microfinance Institutions.