International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 9, Issue 2, pp 152-166, February-2025
https://doi.org/10.47001/IRJIET/2025.902024

Design and Implementation of a Credit Card Fraud Detection System Using Random Forest and Logistics Regression Models

Anusiuba, Overcomer Ifeanyi Alex

Department of Computer Science, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria
E-mail: oi.anusiuba@unizik.edu.ng

***Abstract* - Credit card fraud poses a significant challenge in the digital era, necessitating advanced techniques for early detection and prevention. This study presents a comprehensive exploration into the design and implementation of a credit card fraud detection system leveraging machine learning models, specifically Random Forest and Logistic Regression. The research methodology involves preprocessing a diverse and extensive credit card transaction dataset, encompassing various transaction features. Through careful feature engineering, the dataset is prepared for training and testing the Random Forest and Logistic Regression models. The Random Forest model, employing ensemble learning, amalgamates multiple decision trees to enhance predictive accuracy and resilience against over fitting. Concurrently, Logistic Regression, a classical statistical method, analyzes the relationship between input features and the likelihood of fraudulent transactions. The comparative analysis of these models provides insights into their respective strengths and weaknesses, aiding in the selection of the most effective model for credit card fraud detection. The evaluation phase assesses the performance of the models using key metrics such as accuracy, precision, recall, and F1-score. A detailed examination of these metrics under various scenarios sheds light on the models' ability to distinguish between legitimate and fraudulent transactions. Real-world implications of implementing these models in financial institutions or credit card companies are discussed, emphasizing the potential for enhanced security and reduced financial losses. Moreover, this study discusses the ethical considerations and challenges associated with deploying machine learning models in fraud detection systems. Privacy concerns, model interpretability, and the dynamic nature of fraud patterns are acknowledged, providing a holistic view of the practical implications of implementing such systems. Finally, the findings of this research contribute valuable insights to the ongoing efforts in combating credit card fraud. The comprehensive analysis of Random Forest and Logistic Regression models, coupled with real-world applicability and ethical considerations, positions this study as a significant advancement in the field of financial security and fraud prevention.**

Keywords: Credit Card, Fraud Detection System, Credit Card Fraud Detection, Random Forest, Logistics Regression Models.

I. BACKGROUND OF THE STUDY

Credit card fraud is a serious problem, and it's only getting worse. In 2021 alone, there were over 1.3 billion dollars in losses due to credit card fraud in the United States. In Nigeria, according to the Nigerian Economic and Financial Crimes Commission, there were over 100,000 cases of credit card fraud reported in 2021. The total financial loss from these cases was estimated to be over one billion naira. The problem is especially severe in Nigeria because of the lack of robust financial regulations and enforcement. In addition, there is a lack of education and awareness about the dangers of credit card fraud. Current methods for detecting and preventing credit card fraud include the use of algorithms, data mining, and artificial intelligence. However, these methods have limitations and often result in false positives or false negatives. There is need for new and innovative methods that can accurately detect and prevent credit card fraud.

The design and implementation of a credit card fraud detection system using Random Forest and Logistic Regression models is a topic of interest in the field of machine learning. The existing system for credit card fraud detection uses various machine learning algorithms, including Random Forest and Logistic Regression, to detect fraudulent transactions (Anusiuba et al, 2022)

Random Forest is a supervised machine learning technique that builds several decision trees to improve classification performance in credit card fraud detection. It has been used in hybrid approaches with other methods like isolation forest and SVM to address the challenge of detecting fraudulent transactions in big imbalanced datasets (Dornadul & Geetha, 2019).

© 2025 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 2, pp 152-166, February-2025

https://doi.org/10.47001/IRJIET/2025.902024

Logistic Regression, on the other hand, is another algorithm used for credit card fraud detection. It has been analyzed alongside other methods such as Gradient Boosting Classifier and Random Forest to obtain the best model for electronically detecting unauthorized financial transactions in bank payment systems (Sulaiman, et al., 2022).

The design and implementation of a credit card fraud detection system using Random Forest and Logistic Regression models involves analyzing the dataset, building and optimizing machine learning models, and comparing the results to determine the best model for fraud detection. The Random Forest model has been found to perform well in terms of fraud detection rate, achieving 54% in out-of-time tests (Han, et al., 2020).

The design and implementation of a credit card fraud detection system using Random Forest and Logistic Regression models involves analyzing the dataset, building and optimizing machine learning models, and comparing the results to determine the best model for fraud detection. Random Forest and Logistic Regression have shown promising results in improving classification performance and detecting fraudulent transactions, making them valuable tools for banking and financial institutions in combating credit card fraud.

1.1 Statement of the Problem

The statement of the problem in the design and implementation of a credit card fraud detection system using Random Forest and Logistic Regression models involves addressing the need for a secure credit card fraud detection system due to the prevalence of fraudulent transactions and the resulting financial losses for banks and customers. The high volume of credit card transactions, coupled with the existence of fraudulent transactions, necessitates the development of effective fraud detection systems. Various machine learning algorithms, including Naïve Bayes, Logistic Regression, SVM, Decision trees, Random Forest, Genetic algorithm, J48, and AdaBoost, are used for credit card fraud detection.

The challenges to be addressed in the design and implementation of the system include the highly unbalanced nature of fraud data, the need to improve the performance of machine learning algorithms such as Random Forest, Logistic Regression, K-Nearest Neighbor, and Naïve Bayes, and the absence of benchmarks and standard evaluation metrics for identifying better performing classifiers. Additionally, the uniqueness of frauds and the ingenuity of fraudsters pose challenges to the accurate detection of fraudulent transactions.

The goal of the study is to provide insight into credit card fraud, analyze the dataset, and discuss the application of Decision tree and Random Forest algorithms in credit card fraud detection. The study aims to develop and optimize machine learning models, compare the results, and determine the best model for fraud detection. The ultimate objective is to develop a robust credit card fraud detection system that effectively identifies and mitigates fraudulent transactions, thereby reducing financial losses for banks and customers.

1.2 Aim and Objectives of the Study

The aim of the design and implementation of a credit card fraud detection system using Random Forest and Logistic Regression models is to develop an effective and secure system for detecting fraudulent credit card transactions. The primary objective is to leverage machine learning algorithms, specifically Random Forest and Logistic Regression, to improve the performance of fraud detection systems and minimize financial losses due to credit card fraud. The study aims to provide insight into credit card fraud, analyze the dataset, and discuss the application of Decision tree and Random Forest algorithms in credit card fraud detection

The objectives of the study include:

1. Analyzing the dataset to understand the patterns and characteristics of fraudulent credit card transactions.
2. Building and optimizing machine learning models, particularly Random Forest and Logistic Regression, to enhance the accuracy and efficiency of fraud detection.
3. Comparing the performance of different machine learning models to determine the most effective approach for detecting fraudulent transactions.
4. Developing a robust credit card fraud detection system that can be applied in banking and financial institutions to mitigate the occurrence of fraudulent activities and reduce financial losses.

II. REVIEW OF RELATED LITERATURE

Financial fraud is a serious problem that is only getting worse and has far-reaching effects on the financial sector, businesses, and the government (Ekwealor, et al., 2021). Fraud is defined as criminal deception done with the intention of making money (Ekwealor et al, 2021). Credit card transactions have surged thanks to a high reliance on internet technology. The rate of credit card fraud is rising as credit card transactions take over as the preferred method of payment for both online and offline transactions .

Banks used to provide only in-person services to customers until 1996 when the first internet banking application was introduced in the United States of America by Citibank and Wells Forgo Bank (Yak & Tudeal, 2011). After the introduction of internet banking, the use of credit cards

© 2025 IRJIET All Rights Reserved

ISSN (online): 2581-3048
Volume 9, Issue 2, pp 152-166, February-2025
https://doi.org/10.47001/IRJIET/2025.902024

over the internet was adopted. This has increased rapidly during the past decade and services like e-commerce, online payment systems, working from home, online banking, and social networking have also been introduced and widely used (Madan,et al., 2021). Due to this, fraudsters have intensified their efforts to target online transactions utilizing various payment systems (Yann-a, 2018). In recent times, improvements in digital technologies, particularly for cash transactions, have changed the way people manage money in their daily activities. Many payment systems have transitioned tremendously from physical pay points to digital platforms (Nath, 2020). To sustain productivity and competitive advantage, the use of technology in digital transactions has been a game-changer and many economics have resorted to it (Pencarelli, 2019). Hence, internet banking and other online transactions has been a convenient avenue for customers to carry out their financial and other banking transactions from the comfort of their homes or offices, particularly through the use of credit cards.

According to Raj, et al., (2011), a credit card is designed as a piece of plastic with personal information incorporated and issued by financial service providers to enable customers to purchase goods and services at their convenience worldwide. The unlawful use of another person's credit card to get money or property either physically or digitally is known as credit card fraud (Anusiuba et al, 2022). Events involving credit card fraud occurs often end in enormous financial losses (Anusiuba et al, 2022). It is simpler to commit fraud now than it was in the past because an online transaction environment does not require the actual card and the card's information suffices to complete a payment (Vlasselaer, et al., 2015). Faisal, et al., (2021), postulate that monetary policy as well as business plans and methods used by big and small businesses alike have been impacted by the introduction of credit cards.

There are two types of credit card fraud: internal and external (Aihua, et al., 2007). While external card fraud entails using a stolen credit card to obtain money through illegal ways, inner card fraud happens as a result of an agreement between cardholders and the bank and involves using a fake identity to commit fraud. Most credit card frauds are external card fraud, which has been the subject of much investigation. Another classification has been made into three categories: classic card-related frauds (application, stolen, account takeover, fake, and counterfeit), frauds involving retailers (merchant collusion and triangulation), and frauds involving the internet (site cloning, credit card generators, and false merchant sites) (Delamaire, et al.,2022).Due to their time-consuming nature and ineffectiveness, manual methods of fraud detection have become increasingly impracticable with the introduction of big data. The challenge of credit card fraud, however, has drawn the attention of financial institutions to current computational approaches.

Following similar patterns, compliance and risk management services employed to identify online fraud have shown a lot of interest in AI and machine learning models (Kurshan, et al., 2020). Some of these models include, Decision Tree, Logistic Regression, Random Forest, Ada Boost, XG Boost, Support Vector Machine (SVM) and Light GBM.

This has become necessary because credit card fraud detection is a classification and prediction problem. Supervised machine learning models have been proved as the best models to detect fraud using the above-mentioned algorithms (Lebichot, et al., 2021). This study therefore seeks to compare two classification and prediction techniques, namely; Logistic Regression, and Random Forest in classifying and predicting financial transactions as either fraudulent or not fraudulent.

Afriyie et.al, (2023) examines the application of supervised machine learning (ML) algorithms in detecting and predicting credit card fraud. It evaluates models such as logistic regression, random forest, and decision trees, emphasizing their efficacy in fraud prevention. The article highlights the adaptive nature of ML algorithms, which improve continuously as they process new data, thereby enhancing their ability to detect emerging fraud patterns. Furthermore, it explores both supervised and unsupervised learning techniques, showcasing the significant advantages of machine learning in tackling fraud across various industries, particularly in financial transactions.

Krishna & Praveenchandar, (2022) in their study presents a comparative analysis of two machine learning models—logistic regression and random forest—specifically for detecting credit card fraud. The research evaluates the performance of both algorithms in identifying fraudulent transactions, with findings indicating that random forest delivers significantly better accuracy compared to logistic regression. The article underscores the importance of machine learning algorithms in enhancing fraud detection capabilities and improving the prediction accuracy for financial institutions, particularly in the realm of credit card fraud detection.

Khyati, Yadav, & Mallick, (2012) provides a comprehensive overview of various machine learning techniques applied to the detection of credit card fraud in online transactions. It discusses the rising incidence of fraud and the role of machine learning methods in mitigating this issue. The paper emphasizes the critical importance of feature selection in fraud detection and evaluates a range of

© 2025 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Issue 2, pp 152-166, February-2025
https://doi.org/10.47001/IRJIET/2025.902024

supervised learning algorithms, including Decision Trees, Random Forest, Artificial Neural Networks (ANN), Naive Bayes, and Logistic Regression. Additionally, the review highlights the necessity of large, high-quality datasets for effective training and testing of these models. The article concludes by reinforcing the growing importance of machine learning in preventing credit card fraud.

RamaKalyani & Uma (2012) applied genetic algorithms to reduce false alerts by incorporating customer behavior patterns, showing that this approach can effectively predict fraudulent transactions soon after they occur. Rawat, (2022) conducted a comparative analysis of machine learning algorithms and concluded that ML techniques offer higher accuracy and detection rates compared to traditional methods. Maniraj et al. (2019) utilized local outlier factors and isolation forest algorithms, achieving 99.6% accuracy on a smaller dataset. Meanwhile, Aman, et al , (2017) incorporated customer behavior and location data to assess the likelihood of fraud, although their system struggled to detect fraud committed by new users. The review suggests that while machine learning has made significant strides, the need for larger, more diverse datasets and the development of advanced methods, such as deep learning, is critical for improving the effectiveness of fraud detection systems.

Although machine learning algorithms have demonstrated their ability to identify fraudulent credit card transactions, there is still room for improvement. While these models can detect, classify, and potentially prevent fraud, their accuracy is often constrained by the limitations of the training data. To further enhance fraud detection capabilities, the integration of deep learning techniques is recommended. Deep learning models have the potential to significantly improve the accuracy, reliability, and efficiency of fraud detection systems, providing more robust solutions for financial institutions in combating fraud. The evolving nature of fraudulent activities necessitates the development of increasingly sophisticated detection methods to address the challenges faced in the financial sector.

III. METHODOLOGY ADOPTED

The methodology adopted in the design and implementation of a credit card fraud detection system using random forest and logistic regression models involves the use of supervised machine learning algorithms to detect and predict fraud in credit card transactions. The machine learning models of logistic regression, random forest, and decision trees are evaluated for detecting fraudulent credit card transactions. The random forest algorithm is chosen because it acts as a binary classifier, making it suitable for credit card fraud detection, classifying transactions as either fraud or not fraud. The workflow for training the model involves reading, partitioning, random forest training, random forest prediction generation, threshold application, and performance scoring.

The selected methodology focuses on achieving the highest possible accuracy and preventing financial losses due to fraud. It involves the evaluation of machine learning models, training the models, and simulating them to electronically detect unauthorized financial transactions in bank payments. Additionally, the methodology considers the use of supervised machine learning algorithms for detecting and predicting fraud in credit card transactions, with a particular emphasis on the analysis of the dataset and the selection of appropriate algorithms for accurate fraud detection.

3.1 Analysis of the Existing System

The existing system is about credit card fraud detection systems with the help of comparative analysis of KNN and Logistic Regression algorithms in addition to classification and regression algorithms, aiming to procure optimal elucidation as time progresses. Here, they aim to diminish false alerts with the help of a Machine Learning algorithm while optimizing a group of interval-valued parameters. For that reason, through this work, they have tried to evolve a fraud detection system using K-NN and Logistic Regression algorithm. By using the existing system, they can detect malicious activities and can raise false alerts while making credit card transactions. The parameters considered for comparative analysis are precision, recall, and accuracy.

3.2 Dataflow of the Existing System

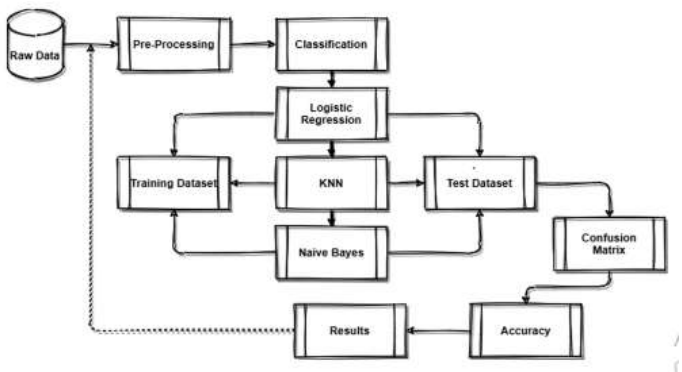

Figure 3.1: A dataflow of an existing system

3.3 Analysis of the Proposed System

The method used in this research is the sample data method. The reason for using this method is because it is the best way to collect datasets from search results and learn datasets from Kaggle datasets. This research is systematically divided into several stages of research consisting of data

© 2025 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 2, pp 152-166, February-2025

https://doi.org/10.47001/IRJIET/2025.902024

collection (dataset), data processing and reading, histogram check, feature engineering, model training, evaluation and validation:

A. Data collection (Dataset): The dataset contains transactions made with credit cards by cardholders in Europe from the Kaggle dataset. This register represents transactions that occurred in the last two days; from the information the dataset has 492 fraudulent transactions. For some information about the characteristics of datasets like V1, V2,..V28 is the main component obtained by the PCA process. The "Time" attribute contains the seconds that elapsed between each transaction in the log data. Attribute "Amount" is the number of transactions, this attribute can be used as paid learning. The 'Class' feature is a response variable and takes a value of 1 if there is fraud and 0 if there is no fraud.

B. Data processing and reading: The data is processed based on the results of data collection and data cleaning processes to overcome data problems such as data anomalies, missing data values, data redundancy, and inappropriate data. The data is then selected and grouped by type and function to divide it into training and testing data so that it can be applied to the classification algorithm that will be tested. The development carried out in this research is the addition of Machine Learning (Supervised Learning) algorithms, namely the Random Forest Classifier (RFC) and Logistic Regression (LGR).

C. Histogram check: Created a function to check the distribution of the values of the features over time.

D. Feature Engineering: Identifying and selecting relevant features from the dataset that is crucial for detecting fraudulent transactions and creating independent and dependent features. Defining the target variable and the target variable is a feature called class.

E. Model Training: Utilizing machine learning algorithms such as random forest and logistic regression to build the fraud detection model. The dataset is divided into training and testing sets for model evaluation.

F. Model Evaluation and Validation: Assessing the performance of the trained models using metrics such as precision, recall, and accuracy. The models are validated on test datasets to ensure their effectiveness in detecting fraudulent transactions.

3.4 Dataflow of the Proposed System

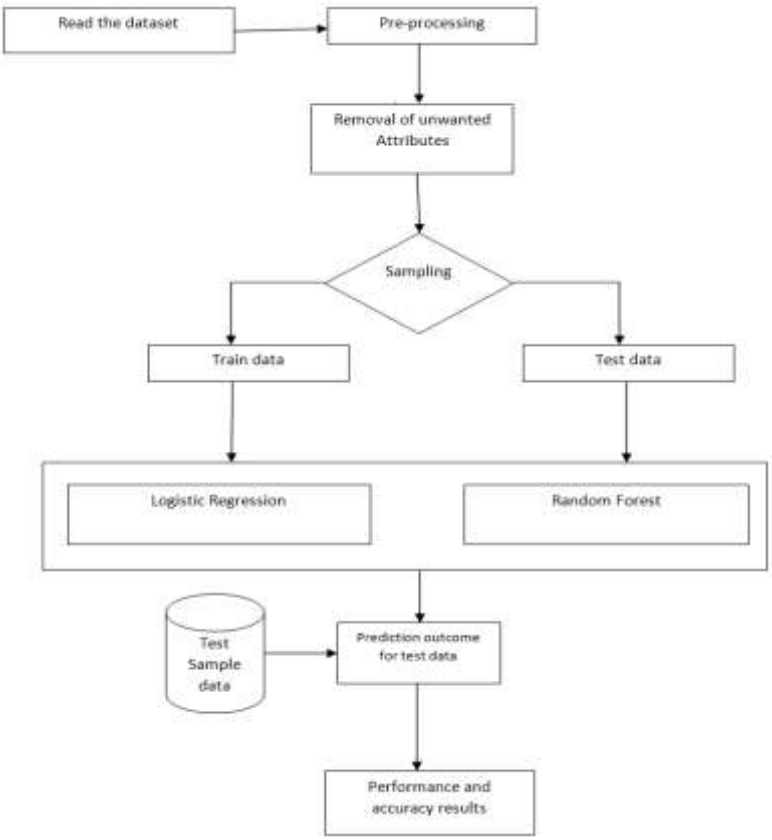

Figure 3.2: A data flow of the proposed system

© 2025 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Issue 2, pp 152-166, February-2025
https://doi.org/10.47001/IRJIET/2025.902024

3.4 High Level Model of the Proposed System

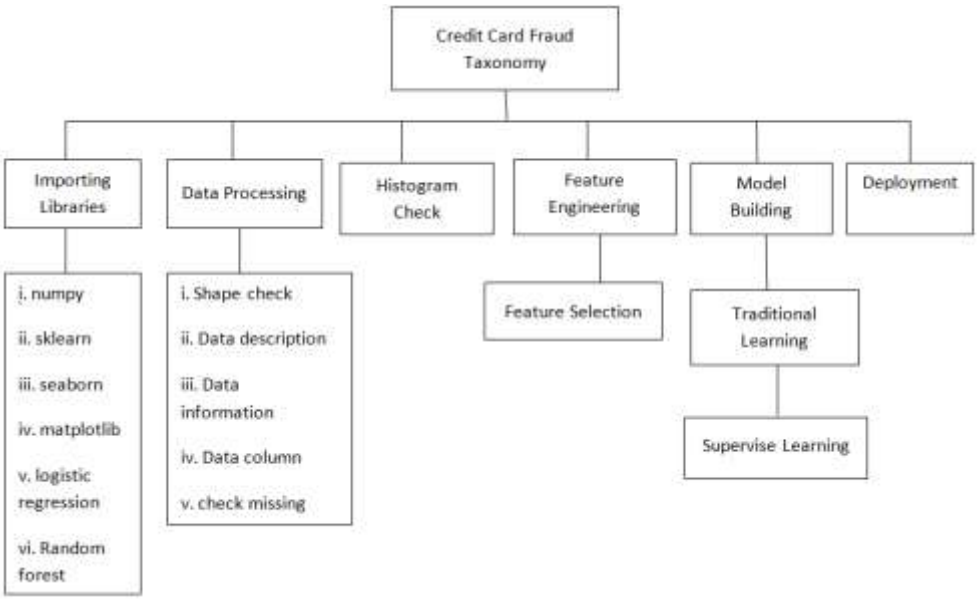

Figure 3.3 High level model of the proposed system

3.5 Control Centre/Main Menu

The focus is primarily on the algorithms and the data processing. Designing a main menu for a credit card fraud detection system involves presenting key functionalities and options in a user-friendly and intuitive manner. Below is the main menu:

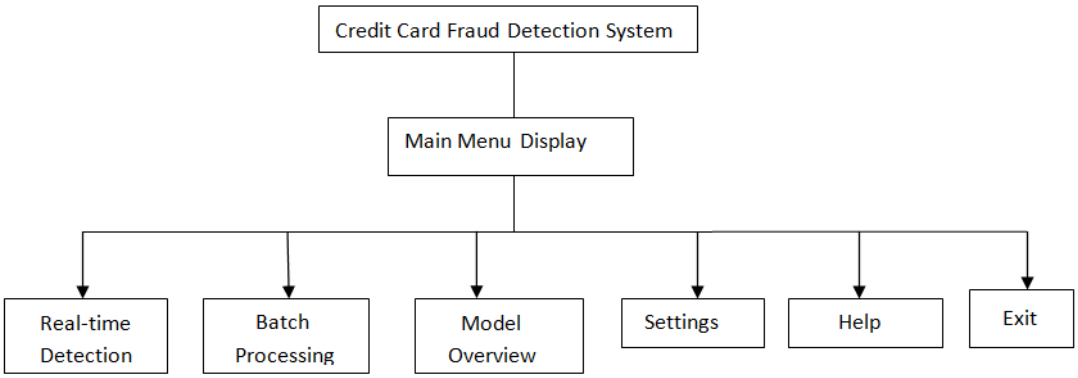

Figure 3.4: System Main Menus

3.6 The Submenus/Subsystem

3.6.1 Real-time Detection Subsystems and Brief Description of each item

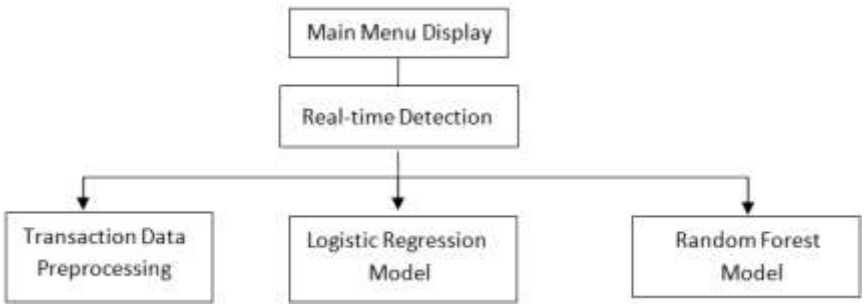

Figure 3.5: Subsystem for Real-time Detection Menu

© 2025 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Issue 2, pp 152-166, February-2025
https://doi.org/10.47001/IRJIET/2025.902024

3.6.2 Input/output format

The input module has an input which is text provided in a form field. The following input is expected to be provided by the user:

i. Time
ii. v1, v2, v3, v4, v5, v6, v7 & v8
iii. Amount
iv. Class

The output module provides details if the transaction is fraudulent or not.

3.7 Algorithm

```
# Import necessary libraries
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report
# Load the dataset
# Assuming you have a CSV file named 'credit_card_data.csv'
data = pd.read_csv('credit_card_data.csv')
# Separate features and labels
X = data.drop('fraud_label', axis=1)
y = data['fraud_label']
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# Standardize features
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
# Train Logistic Regression model
logistic_model = LogisticRegression()
logistic_model.fit(X_train, y_train)
# Train Random Forest model
random_forest_model = RandomForestClassifier()
random_forest_model.fit(X_train, y_train)
# Ensemble model predictions
logistic_predictions = logistic_model.predict_proba(X_test)[:, 1]
random_forest_predictions = random_forest_model.predict_proba(X_test)[:, 1]
# Combine predictions (e.g., simple average)
ensemble_predictions = (logistic_predictions + random_forest_predictions) / 2
# Evaluate the model
accuracy = accuracy_score(y_test, final_predictions)
classification_report_output = classification_report(y_test, final_predictions)
# Print results
print(f'Accuracy: {accuracy:.2f}')
print('Classification Report:')
print(classification_report_output)
```

3.8 System Flowchart

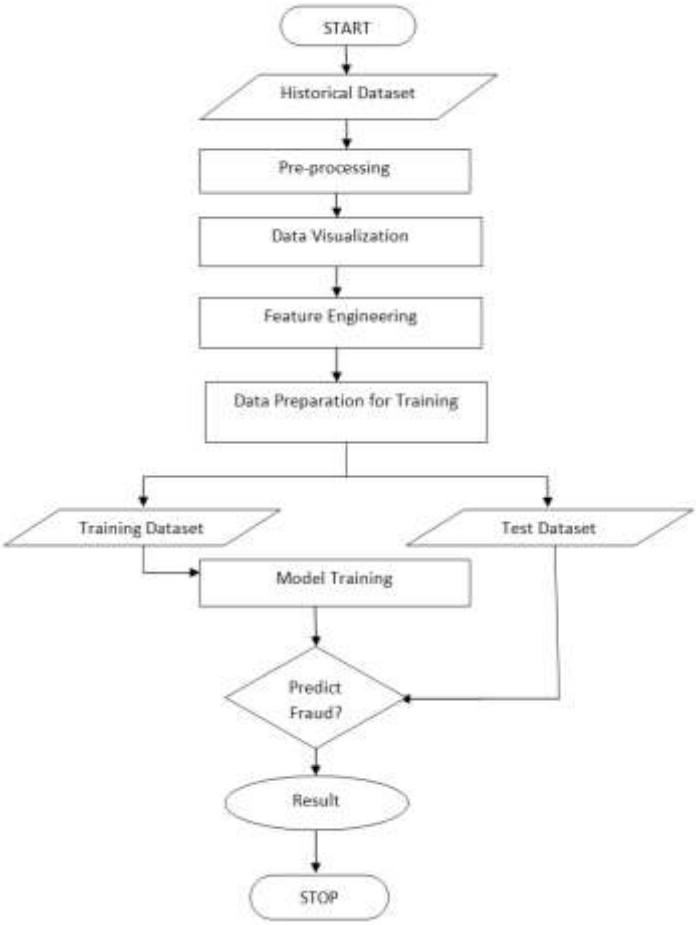

Figure 3.6: System Flowchart of the System

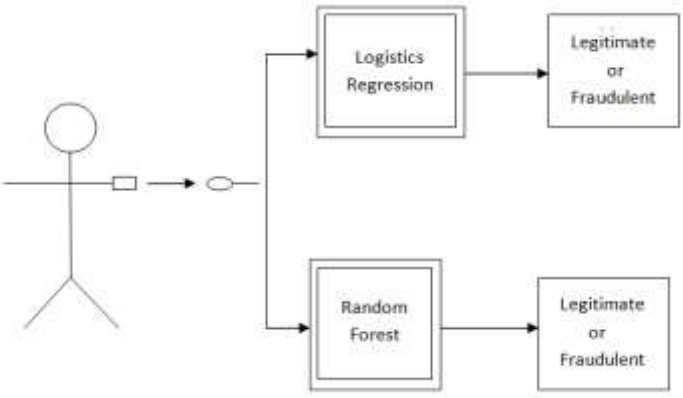

Figure 3.7: Use case diagram of the system

© 2025 IRJIET All Rights Reserved

ISSN (online): 2581-3048
Volume 9, Issue 2, pp 152-166, February-2025
https://doi.org/10.47001/IRJIET/2025.902024

IV. Summary

The study explores the development and implementation of a credit card fraud detection system using machine learning algorithms. It begins by outlining the context, significance, and objectives of the research, addressing gaps in existing knowledge and defining the scope and limitations of the study. A thorough literature review is conducted, highlighting theoretical insights and previous works, and identifying areas for further investigation. The research methodology includes an analysis of both existing and proposed systems, with a focus on data flow and system models.

In the context of fraud detection, machine learning algorithms like random forest and logistic regression are examined for their effectiveness. Random forest, which uses an ensemble of decision trees, is particularly adept at handling imbalanced datasets and detecting fraudulent transactions. Logistic regression is employed to predict the likelihood of fraud by analyzing patterns and anomalies in transaction data. The system is built using tools such as Jupyter Notebook, Visual Studio Code, and FastAPI, allowing for real-time data analysis, model training, and continuous monitoring to ensure accurate detection.

In conclusion, the study demonstrates that integrating machine learning techniques and modern development tools significantly enhances credit card fraud detection, reducing financial risks and losses for organizations.

4.1 Conclusion

Credit card fraud detection systems are crucial in preventing fraudulent activities, and machine learning algorithms like random forest and logistic regression are highly effective in this domain. Random forest uses an ensemble of decision trees to classify data, making it particularly useful for handling imbalanced datasets and providing high accuracy in identifying fraudulent transactions. Logistic regression, a statistical method, predicts the probability of events and is widely used to detect fraud by identifying patterns and anomalies in transaction data. To build a real-time fraud detection system, tools such as Jupyter Notebook, Visual Studio Code, and FastAPI are employed. The process involves analyzing raw data, creating features to capture fraud patterns, training the model, and deploying it for continuous monitoring to ensure optimal performance and accurate fraud detection. Overall, machine learning techniques like random forest and logistic regression, combined with tools like Jupyter Notebook, enhance the ability to detect and prevent credit card fraud, reducing risks and financial losses for organizations.

REFERENCES

[1] Afriyie J.K., Kassim T., Wilhemina A.P., Addai-Henne S., Dwamena A.H., Owiredu E.O., Ayeh S.A., & Eshun J. (2023). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. 6(2023) 100163, https://doi.org/10.1016/j.dajour.2023.100163

[2] Aihua, S., Rencheng, T., & Yaochen, D. (2007). Application of classification models on credit card fraud detection: Proceedings - ICSSSM'07: International Conference on Service Systems and Service Management. doi: 10.1109/ICSSSM.2007.4280163.

[3] Aman, G., Prakash, D., Norman J., & Mangayarkarasi, R. (2017). Credit card fraud detection using neural network and geo-location in the 14th ICSET-2017. Conf. Series: Materials Science and Engineering 263 (2017) 042039 doi:10.1088/1757-899X/263/4/042039

[4] Anusiuba O. I. A, Okechukwu O. P., Ekwealor O. U, Anusiuba A. A.,(2022), The Application of Hidden Markov Model in Credit Card Fraud Detection System, iJournals: International Journal of Software & Hardware Research in Engineering (IJSHRE), 10(2), 1-20

[5] Ekwealor O.U, Anusiuba O. I. A, Ezuruka E. O, Uchefuna C.I., (2021). An Intelligent Credit Card Fraud Detection System, iJournals: International Journal of Software & Hardware Research in Engineering (IJSHRE) 9(2), 25-54

[6] Carcillo, F., Borgne, Le, Y., Caelen, O., Kessaci, Y., & Oblé, F. (2021). Combining unsupervised and supervised learning in credit card fraud detection: Information Science, 557, 317–331, http://dx.doi.org/10.1016/j.ins.2019.05.042.

[7] Delamaire L., Abdou H., & B. systems, and undefined. (2009). Credit card fraud and detection techniques: a review. eprints.hud.ac.uk, Accessed: Dec. 25, 2022

[8] Dornadul, V.N., & Geetha, S. (2019). Credit card fraud detection using machine learning algorithm 165 631-641 http://creativecommons.org/licenses/by-nc-nd/4.0/

[9] Faisal L.E., Tayachi, T., Arabia, S., & Banking, O. (2021) The role of internet banking in society. 18 (13) (2021) 249–257.

[10] Han, Y., Yao, S., Wen, T., Tian, Z., Wang C., & Gu, Z. (2020). Detection and Analysis of Credit Card Application Fraud Using Machine Learning Algorithms. 1693(2020) 25-27 https://doi.org/10.1088/1742-6596/1693/1/012064

[11] Khyati C., Yadav J., & Mallick B. (2012). A review of Fraud Detection Techniques: Credit Card, International Journal of Computer Applications (0975 – 8887) 45(1).

© 2025 IRJIET All Rights Reserved

ISSN (online): 2581-3048
Volume 9, Issue 2, pp 152-166, February-2025
https://doi.org/10.47001/IRJIET/2025.902024

[12] Krishna, M.V., & Praveenchandar, J. (2022). Comparative Analysis of Credit Card Fraud Detection using Logistic regression with Random Forest towards an Increase in Accuracy of Prediction. : International Conference on Edge Computing and Applications (ICECAA), Tamilnadu, India, pp.1097-1101, https://doi.org/10.1109/ICECAA55415.2022.9936488

[13] Kurshan, E., Shen, H., & Yu, H. (2020). Financial crime & fraud detection using graph computing: Application considerations & outlook, in: 2020 Second International Conference on Transdisciplinary AI (TransAI), IEEE, pp. 125–130.

[14] Lebichot, B., Siblini, G.M.P.W., & Bontempi, L.H.F.O.G. (2021). Incremental learning strategies for credit cards fraud detection: International Journal of Data Science and Analytics, 12(2), 165–174.

[15] Madan, S., Sofat, S., & Bansal, D. (2021). Tools and Techniques for Collection and Analysis of Internet-of-Things malware: A systematic state-of-art review, J. King Saud Univ. - Comput. Inf. Sci. http://dx.doi.org/10.1016/j.jksuci.2021.12.016.

[16] Maniraj, S.P., Saini, A., Sarkar, S.D., Ahmed, S. (2019). Credit Card Fraud Detection Using Machine Learning and Data Science in the International Journal of Engineering Research and Technology (IJERT), ISSN: 2278- 0181,Vol. 8

[17] Nath, N. (2020). Credit card fraud detection using machine learning algorithms credit card fraud detection using machine learning algorithms: Procedia Comput. Sci. vol.165 631–641, http://dx.doi.org/10.1016/j.procs.2020.01. 057.

[18] Pencarelli, T. (2019). The digital revolution in the travel and tourism industry, Inf. Technol. Tourism. 0123456789, http://dx.doi.org/10.1007/s40558-019-00160-3.

[19] RamaKalyani, K., & Uma D.D. (2012). Fraud Detection of Credit Card Payment System by Genetic Algorithm in the International Journal of Scientific & Engineering Research. Volume 3 ISSN 2229- 5518

[20] Rawat, T. (2022). Machine Learning For Credit Card Fraud Detection System. 10(5) PP. 08-14 https://doi.org/ 2320-9364

[21] Sulaiman, B.R., Schetinin, V. & Sant, P. (2022). Review of Machine Learning Approach on Credit Card Fraud Detection: Human-Centered Intelligent Systems, 2, 55–68. https://doi.org/10.1007/s44230-022-00004-0

[22] Vlasselaer, V, Van, Bravo, C., Caelen, O., Eliassi-rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2015). A novel approach for automated credit card transaction fraud detection using network-based extensions: Decis. Support Syst. 75 38–48, http://dx.doi.org/10.1016/j.dss.2015.04.013.

[23] Yak K., & Tudeal D. (2011). Internet Banking Development as A Means of Providing Efficient Financial Services in South Sudan. 2 (2011) 139–148

[24] Yann-a, F.C. (2018). Streaming active learning strategies for real-life credit card fraud detection: Assessment and visualization, 2018.

Citation of this Article:

Anusiuba, Overcomer Ifeanyi Alex. (2025). Design and Implementation of a Credit Card Fraud Detection System Using Random Forest and Logistics Regression Models. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 9(2), 152-166. Article DOI https://doi.org/10.47001/IRJIET/2025.902024

© 2025 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 2, pp 152-166, February-2025

https://doi.org/10.47001/IRJIET/2025.902024

APPENDIX A

SOURCE CODE LISTING (PYTHON PROGRAMMING LANGUAGE)

```
# Importing libraries
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
# loading the dataset
data = pd.read_csv('creditcard.csv')
# reading the first five dataset
data.head()
# Data preprocessing
#checking for shape
data.shape
data.columns
data.describe()
#checking for value_counts
data['Class'].value_counts()
# checking for info
data.info()
# checking for missing values
sns.heatmap(data.isna(),yticklabels=False,cbar=False,cmap='viridis')
data.isnull().sum()
# Historical check
def draw_histograms(dataframe, features, rows, cols):
    fig=plt.figure(figsize=(20,20))
    for i, feature in enumerate(features):
        ax=fig.add_subplot(rows, cols, i + 1)
        dataframe[feature].hist(bins = 20, ax = ax, facecolor = 'midnightblue')
        ax.set_title(feature +'Distribution', color='DarkRed')
        ax.set_yscale('log')
    fig.tight_layout()
    plt.show()
draw_histograms(data, data.columns, 8, 4)
# Feature engineering
## independent and dependent features
X = data.drop('Class',axis=1)
y = data.Class
#Model Building
# Logistic Regression
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score,confusion_matrix,classification_report
from sklearn.model_selection import KFold
import numpy as np
from sklearn.model_selection import GridSearchCV
log_Class=LogisticRegression()
grid={'C':10.0**np.arange(-2,3),'penalty':['l1','l2']}
cv=KFold(n_splits=5,random_state=None,shuffle=False)
from sklearn.model_selection import train_test_split
```

© 2025 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 2, pp 152-166, February-2025

https://doi.org/10.47001/IRJIET/2025.902024

APPENDIX B

SAMPLE OUTPUT

```
In [3]: # reading the first five dataset
        data.head()
```

Out[3]:

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V24	V2
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.066928	0.12853
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288	-0.339846	0.16717
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	0.247998	0.771679	0.909412	-0.689281	-0.32764
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...	-0.108300	0.005274	-0.190321	-1.175575	0.64737
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...	-0.009431	0.798278	-0.137458	0.141267	-0.20601

5 rows × 31 columns

Plate 1: Reading the first five dataset

```
In [4]: #checking for shape
        data.shape
Out[4]: (284807, 31)

In [5]: data.columns
Out[5]: Index(['Time', 'V1', 'V2', 'V3', 'V4', 'V5', 'V6', 'V7', 'V8', 'V9', 'V10',
               'V11', 'V12', 'V13', 'V14', 'V15', 'V16', 'V17', 'V18', 'V19', 'V20',
               'V21', 'V22', 'V23', 'V24', 'V25', 'V26', 'V27', 'V28', 'Amount',
               'Class'],
              dtype='object')

In [6]: data.describe()
```

Out[6]:

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...
count	284807.000000	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	...
mean	94813.859575	3.918649e-15	5.682686e-16	-8.761736e-15	2.811118e-15	-1.552103e-15	2.040130e-15	-1.698953e-15	-1.893285e-16	-3.147640e-15	...
std	47488.145955	1.958696e+00	1.651309e+00	1.516255e+00	1.415869e+00	1.380247e+00	1.332271e+00	1.237094e+00	1.194353e+00	1.098632e+00	...
min	0.000000	-5.640751e+01	-7.271573e+01	-4.832559e+01	-5.683171e+00	-1.137433e+02	-2.616051e+01	-4.355724e+01	-7.321672e+01	-1.343407e+01	...
25%	54201.500000	-9.203734e-01	-5.985499e-01	-8.903648e-01	-8.486401e-01	-6.915971e-01	-7.682956e-01	-5.540759e-01	-2.086297e-01	-6.430976e-01	...
50%	84692.000000	1.810880e-02	6.548556e-02	1.798463e-01	-1.984653e-02	-5.433583e-02	-2.741871e-01	4.010308e-02	2.235804e-02	-5.142873e-02	...
75%	139320.500000	1.315642e+00	8.037239e-01	1.027196e+00	7.433413e-01	6.119264e-01	3.985649e-01	5.704361e-01	3.273459e-01	5.971390e-01	...
max	172792.000000	2.454930e+00	2.205773e+01	9.382558e+00	1.687534e+01	3.480167e+01	7.330163e+01	1.205895e+02	2.000721e+01	1.559499e+01	...

Plate 2: Data Preprocessing

```
 2   V2       284807 non-null   float64
 3   V3       284807 non-null   float64
 4   V4       284807 non-null   float64
 5   V5       284807 non-null   float64
 6   V6       284807 non-null   float64
 7   V7       284807 non-null   float64
 8   V8       284807 non-null   float64
 9   V9       284807 non-null   float64
 10  V10      284807 non-null   float64
 11  V11      284807 non-null   float64
 12  V12      284807 non-null   float64
 13  V13      284807 non-null   float64
 14  V14      284807 non-null   float64
 15  V15      284807 non-null   float64
 16  V16      284807 non-null   float64
 17  V17      284807 non-null   float64
 18  V18      284807 non-null   float64
 19  V19      284807 non-null   float64
 20  V20      284807 non-null   float64
 21  V21      284807 non-null   float64
 22  V22      284807 non-null   float64
 23  V23      284807 non-null   float64
 24  V24      284807 non-null   float64
 25  V25      284807 non-null   float64
 26  V26      284807 non-null   float64
 27  V27      284807 non-null   float64
 28  V28      284807 non-null   float64
 29  Amount   284807 non-null   float64
 30  Class    284807 non-null   int64
dtypes: float64(30), int64(1)
memory usage: 67.4 MB
```

Plate 3: Checking for data types

© 2025 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 2, pp 152-166, February-2025

https://doi.org/10.47001/IRJIET/2025.902024

```
In [9]: # checking for missing values
        sns.heatmap(data.isna(),yticklabels=False,cbar=False,cmap='viridis')
Out[9]: <AxesSubplot:>
```

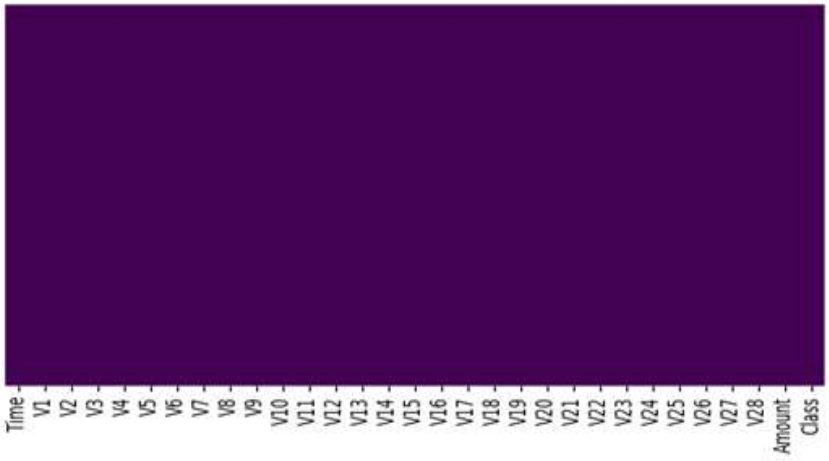

Plate 4: Checking for missing values

```
Out[10]: Time      0
         V1        0
         V2        0
         V3        0
         V4        0
         V5        0
         V6        0
         V7        0
         V8        0
         V9        0
         V10       0
         V11       0
         V12       0
         V13       0
         V14       0
         V15       0
         V16       0
         V17       0
         V18       0
         V19       0
         V20       0
         V21       0
         V22       0
         V23       0
         V24       0
         V25       0
         V26       0
         V27       0
         V28       0
         Amount    0
         Class     0
```

Plate 5: Checking missing values

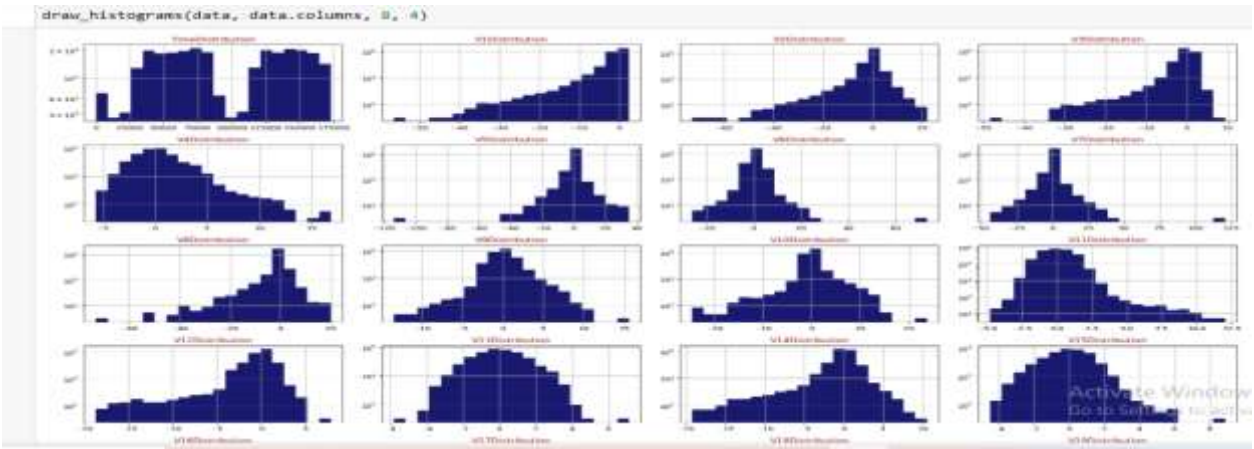

Plate 6: Histogram check

© 2025 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 2, pp 152-166, February-2025

https://doi.org/10.47001/IRJIET/2025.902024

```
In [16]: clf=GridSearchCV(log_Class,grid,cv=cv,n_jobs=-1,scoring='f1_macro')
         clf.fit(X_train,y_train)

C:\Users\USER\anaconda3\lib\site-packages\sklearn\model_selection\_validation.py:372: FitFailedWarning:
25 fits failed out of a total of 50.
The score on these train-test partitions for these parameters will be set to nan.
If these failures are not expected, you can try to debug them by setting error_score='raise'.

Below are more details about the failures:
--------------------------------------------------------------------------------
25 fits failed with the following error:
Traceback (most recent call last):
  File "C:\Users\USER\anaconda3\lib\site-packages\sklearn\model_selection\_validation.py", line 680, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "C:\Users\USER\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py", line 1461, in fit
    solver = _check_solver(self.solver, self.penalty, self.dual)
  File "C:\Users\USER\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py", line 447, in _check_solver
    raise ValueError(
ValueError: Solver lbfgs supports only 'l2' or 'none' penalties, got l1 penalty.

  warnings.warn(some_fits_failed_message, FitFailedWarning)
C:\Users\USER\anaconda3\lib\site-packages\sklearn\model_selection\_search.py:969: UserWarning: One or more of the test scores a
re non-finite: [       nan 0.81670599        nan 0.82133055        nan 0.82648408
        nan 0.83494562        nan 0.82920138]
  warnings.warn(
C:\Users\USER\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:814: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
```

Plate 7: Feature Engineering

```
In [18]: # confusion matrix
         cm=confusion_matrix(Y_test,y_pred)
         conf_matrix=pd.DataFrame(data=cm,columns=['Predicted:0','Predicted:1'],index=['Actual:0','Actual'])
         plt.figure(figsize = (8,5))
         sns.heatmap(conf_matrix, annot=True,fmt='d',cmap='YlGnBu')

Out[18]: <AxesSubplot:>
```

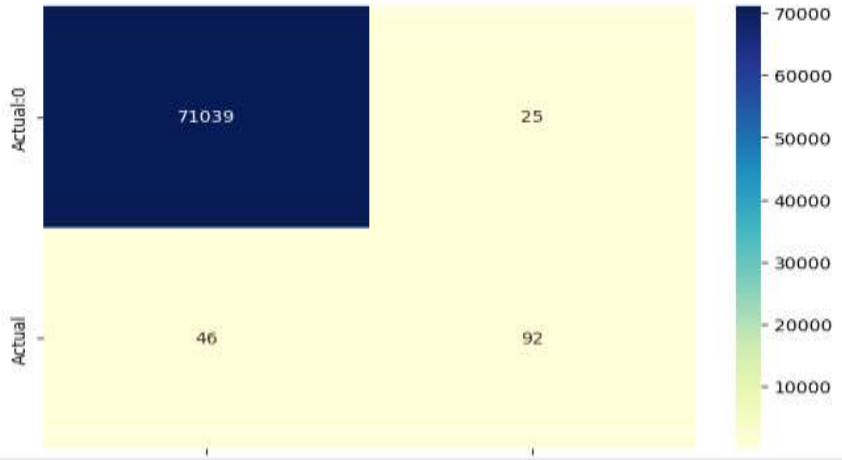

Plate 8: Validation of logistic regression model

```
In [19]: print(accuracy_score(Y_test,y_pred))
0.9990028369989608

In [20]: print(classification_report(Y_test,y_pred))
                precision    recall  f1-score   support

           0       1.00      1.00      1.00     71064
           1       0.79      0.67      0.72       138

    accuracy                           1.00     71202
   macro avg       0.89      0.83      0.86     71202
weighted avg       1.00      1.00      1.00     71202

the logistic Regression model predicted 100% accurately
```

Plate 9: Accuracy of the logistic regression model which gives 100% accuracy

© 2025 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 2, pp 152-166, February-2025

https://doi.org/10.47001/IRJIET/2025.902024

```
In [23]: # confusion matrix
cm=confusion_matrix(Y_test,y_pred)
conf_matrix=pd.DataFrame(data=cm,columns=['Predicted:0','Predicted:1'],index=['Actual:0','Actual:1'])
plt.figure(figsize = (8,5))
sns.heatmap(conf_matrix, annot=True,fmt='d',cmap="YlGnBu");
```

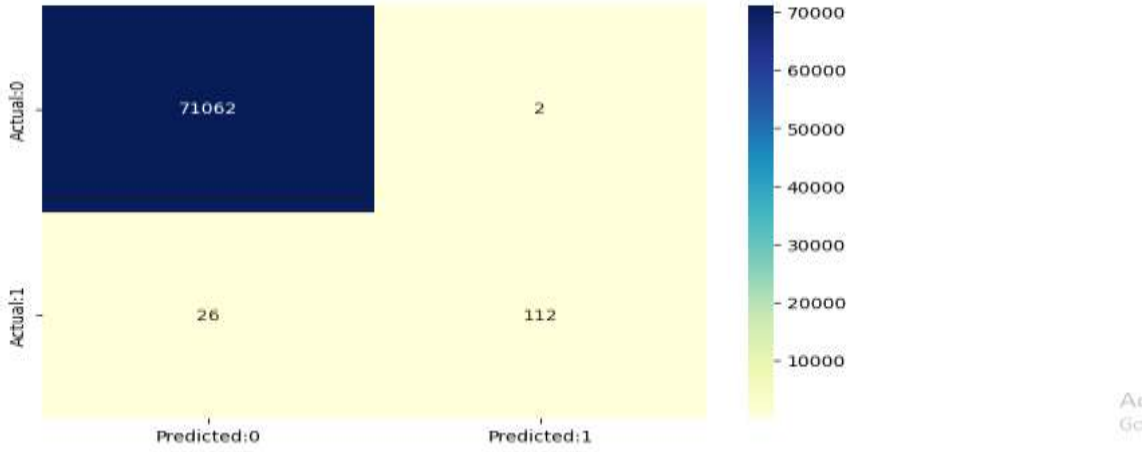

Plate 10: Validation of random forest model

```
In [24]: print(accuracy_score(Y_test,y_pred))

         0.9996067526193084

In [25]: import pickle

In [26]: filename = 'Credit_card_Detection_model.pkl'
         pickle.dump(classifier, open('Credit_card_Detection_model.pkl', 'wb'))

In [27]: # loading the saved model
         loaded_model = pickle.load(open('Credit_card_Detection_model.pkl', 'rb'))
```

Plate 11: Accuracy of the random forest model which gives 100% accuracy

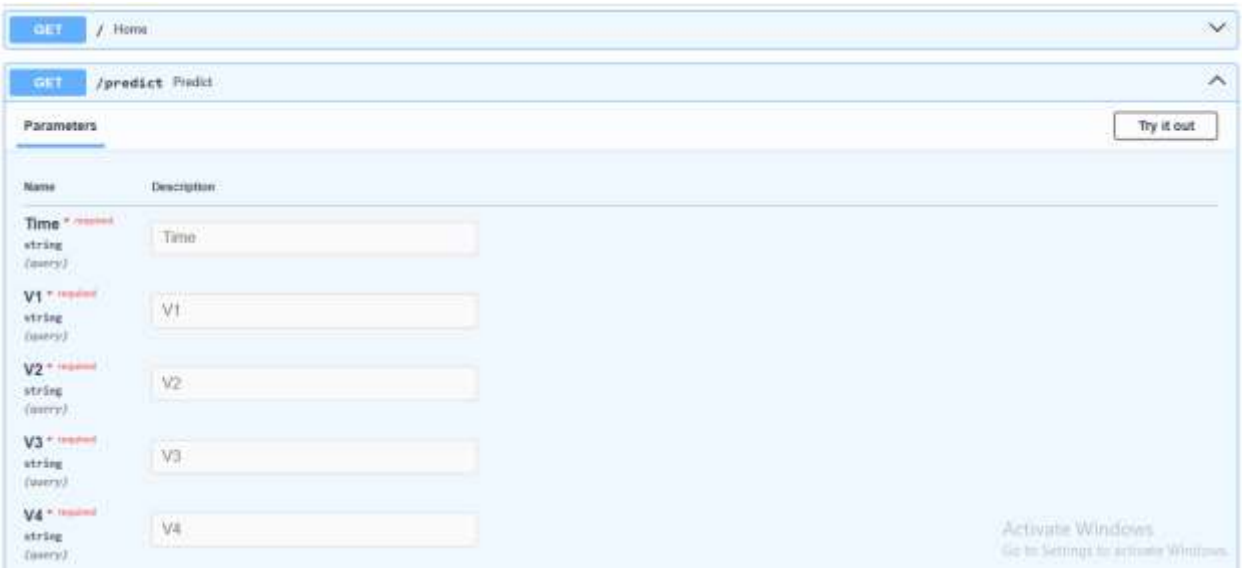

Plate 12: Deployment Stage

© 2025 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 2, pp 152-166, February-2025

https://doi.org/10.47001/IRJIET/2025.902024

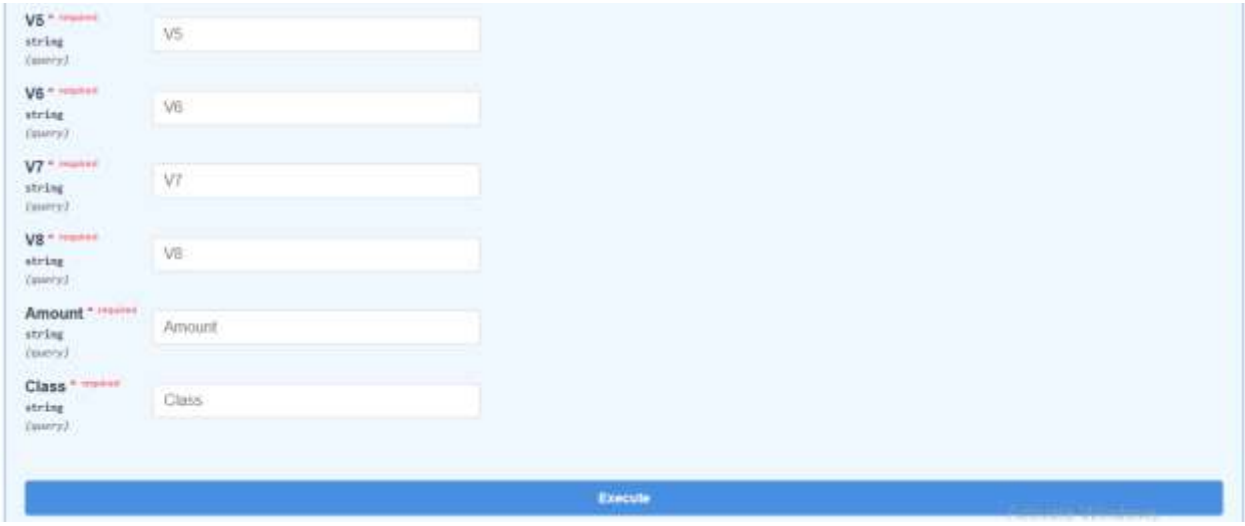

Plate 13: Deployment Stage

****

© 2025 IRJIET All Rights Reserved