**Credit Card Fraud Detection**

A Project Report

submitted in partial fulfillment of the requirements

of

……………. Industrial Artificial Intelligence with Cloud Computing ……

by

**Parth Mehta,**

**Abizer Popat,**

**Nauman Khokhar,**

**Harsh Khandelwal,**

Under the Esteemed Guidance of

**Jay Rathod**

**ACKNOWLEDGEMENT**

We would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this project on "Credit Card Fraud Detection".

Firstly, we would like to thank our supervisor, MR. JAY RATHOD, for being a great mentor and the best guide we could ever have. His guidance, encouragement, and constructive criticism were the source of innovative ideas, inspiration, and the driving force behind the successful completion of this project. The confidence he showed in us was the biggest source of motivation for us. It has been a privilege working with him throughout this project. He always helped us during our project and many other aspects related to academics. His talks and lessons not only helped in the project work and other activities of college but also made us good and responsible professionals.

We would also like to thank our colleagues and peers who contributed to the project in various ways, providing valuable insights and suggestions that helped us improve our work.

Finally, we would like to acknowledge the support of our institution, which provided us with the necessary resources and infrastructure to complete this project.

#### ABSTRACT

*As the payment method is simplified by the combination of the financial industry and IT technology, the payment method of consumers is changing from cash payment to electronic payment using credit card, mobile micropayment, and app card. As a result, the number of cases in which anomalous transactions are attempted by abusing e-banking has increased and financial companies started establishing a Fraud Detection System (FDS) to protect consumers from abnormal transactions. The abnormal transaction detection system aims to identify abnormal transactions with high accuracy by analyzing user information and payment information in real time. Although FDS has shown good results in reducing fraud, but the majority of cases being flagged by this system are False Positives that resulting in substantial investigation costs and cardholder inconvenience. The possibilities of enhancing the current operation constitute the objective of this research. Based on variations and combinations of testing and training class distributions, experiments were performed to explore the influence of these parameters. In this study, we investigated the trend of abnormal transaction detection using payment log analysis and data mining, and summarized the data mining algorithm used for abnormal credit card transaction detection. We used python programming with Apache spark for advanced processing of data and high accuracy.*

**TABLE OF CONTENTS**

Abstract 3

List of Figures 5

**Chapter 1.**  **Introduction**  **6**

1.1 Problem Statement 7

1.2 Problem Definition 7

1.3 Expected Outcomes 7

1.4 Organization of the Report 8

**Chapter 2.**  **Proposed Methodology** **9**

2.1 System Design……………………………………………………………………………………………. 10

2.2 Use Case Diagram………………………………………………………………………………………. 11

2.3 Data Flow Diagram………………………………………………………………………………………. 11

2.4 Advantages……….………………………………………………………………………………………. 12

2.5 System Specifications…………………………………………………………………………………. 12

2.5 .1 Hardware Specifications…………………………………………………………………………. 12

2.5 .2 Software Specifications……………………………………………………………………………. 12

**Chapter 3.**  **Implementation and Results** ……………………………………………………………………………… **13**

3.1 Implementation…………………………………………………………………………………………. 14

3.1.1 Data Preprocessing and Visualization……………………………………………………. 14

3.1.2 Under Sampling……………….…….……………………………………………………………… 15

3.1.1 Model Training and Evaluation.……………………………………………………………... 15

3.2 Results………………………………………………….………………………………………………….…. 15

**Chapter 4. Conclusion**……………………………………………………………………………….………………………… **18**

3.1.1 Advantages…………………………….………………………………………………………………. 19

3.1.1 Scope……………………………………….……………………………………………………………. 19

**GitHub Link**……………………………………………………………………………….………………………….……………… **18**

**Video Link**……………………………………………………………………………….…………………………………………… **18**

**References**……………………………………………………………………………….…………………………………………. **20**

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| Figure No. | Figure Title | Page No. |
| 1 | System Architecture | 10 |
| 2 | Use Case Diagram | 11 |
| 3 | DFD Level 1 | 12 |
| 4 | Distributed Class | 15 |
| 5 | Correlation Matrix | 16 |
| 6 | Random Forest Classifier Confusion Matrix | 17 |
| 7 | Comparison with the other Algorithm | 21 |

**CHAPTER 1**

**INTRODUCTION**

1. **Problem Statement:**

The problem statement is to develop a machine learning model that accurately identifies fraudulent transactions in a credit card dataset. The goal is to reduce the risk of financial losses for credit card companies by effectively classifying transactions as either legitimate or fraudulent.

1. **Problem Definition:**

The problem can be defined as a binary classification problem where the dataset contains various features such as time, transaction amounts, and other variables that can help in distinguishing legitimate transactions from fraudulent ones. The dataset is highly unbalanced, with a significant proportion of legitimate transactions and a smaller proportion of fraudulent transactions. The model needs to be trained to handle this imbalance and accurately predict the class of a transaction based on the available features.

1. **Expected Outcomes:**

1. High Accuracy: The model should be able to accurately classify transactions as either legitimate or fraudulent with a high level of precision.

2. Handling Imbalance: The model should be able to effectively handle the imbalance in the dataset by focusing on the most informative features and reducing the impact of the class imbalance.

3. Feature Selection: The model should be able to identify the most relevant features that contribute to the classification of transactions as legitimate or fraudulent.

4. Visualization: The model should provide a clear visualization of the correlation between features and the class labels, helping in understanding the relationships between variables.

5. Improved Detection: The model should be able to detect fraudulent transactions more effectively, reducing the risk of financial losses for credit card companies.

1. **Organization of the Report**

The remaining report is organized as follows:

Chapter 2 Proposed Methodology

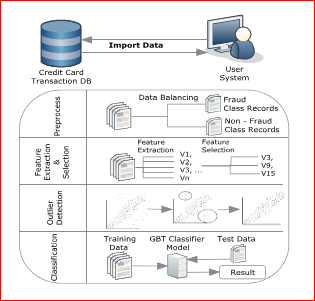
Chapter 3 Implementation and Result

Chapter 4 Conclusion

**CHAPTER 2**

**PROPOSED METHODOLOGY**

* 1. **System Design**

**Figure 1:** **System Architecture**

Above fig shows the process of CCFDS. This system model accepts real time customer credit card transaction database. It is more important to find fraud rate of credit card.

**Data Collection:** The system collects credit card transaction data from various sources, including credit card companies and financial institutions.

**Data Balancing:** The system preprocesses the data by handling missing values, converting categorical variables into numerical variables, and normalizing the data.

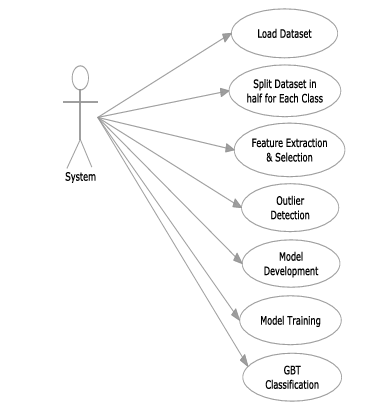
**Feature Extraction and Selection:** The system extracts relevant features from the data, such as time, transaction amounts, and other variables that can help in distinguishing legitimate transactions from fraudulent ones.

**Model Training**: The system trains a machine learning model using the preprocessed and engineered data. The model is trained to predict whether a transaction is legitimate or fraudulent.

**Model Evaluation:** The system evaluates the performance of the trained model using various metrics, such as accuracy, precision, recall, and F1-score.

**Deployment:** The system deploys the trained model in a production environment to detect fraudulent transactions in real-time.

* 1. **Use Case Diagram:**

**Figure 2: Use Case Diagram**

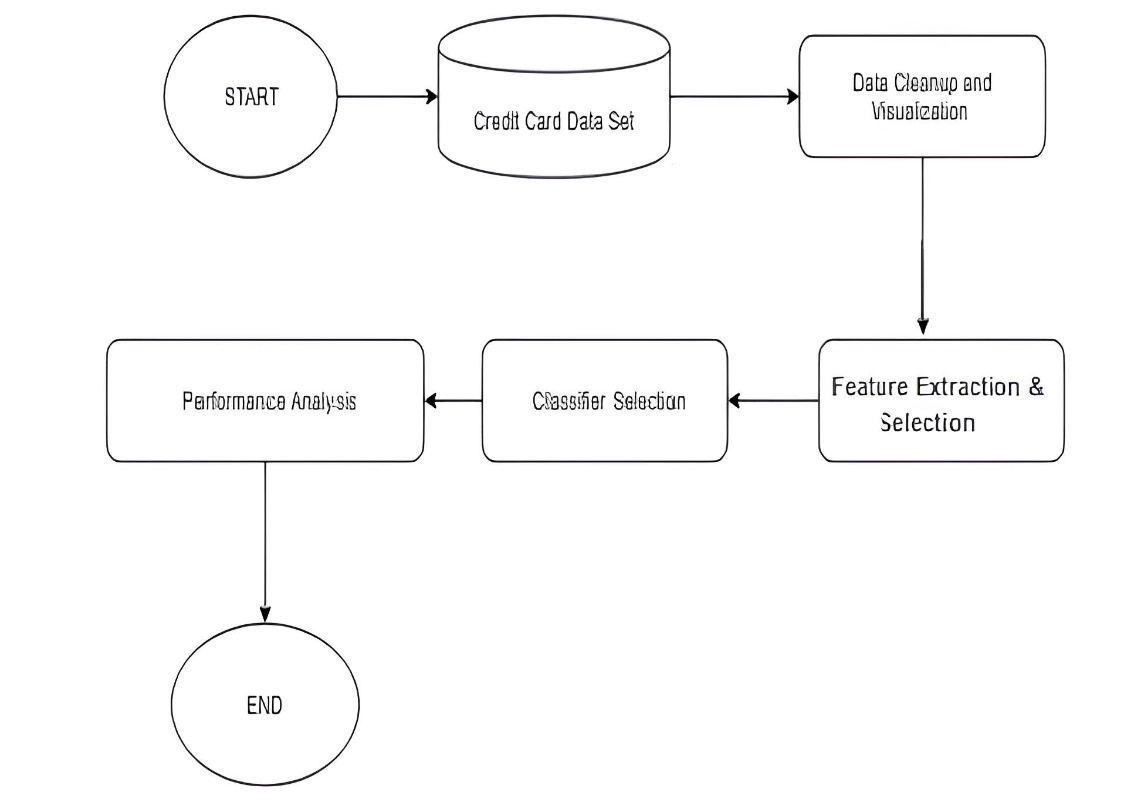
* 1. **Data Flow Diagram**

The DFD used as communication tool between system and user.it is a simple representation of the complete project process. Transaction detection activity follows three phases.:

1. Data exploration

2. Data prepressing

3. Data classifications.

 **Figure 3: DFD Level 1**

* 1. **Advantages**
* The proposed method overcomes the low accuracy forecast problem.
* Utilizing latest AI methods, the fraudulent transactions are recognized and the false alerts are reduced.
* Fast and reliable solution is attained.
  1. **Requirement Specification**
     1. **Hardware Specifications:**

Processor: Pentium i3 or higher.

RAM: 4 GB or higher.

Hard Disk Drive: 20 GB (free).

Peripheral Devices: Monitor, Mouse and Keyboard.

* + 1. **Software Specifications:**

Operating system: Windows 11.

IDE Tool: Jupyter Notebook.

Coding Language: Python 3.6.

APIs : Numpy, Pandas, Matplotlib.

**CHAPTER 3**

**IMPLEMENTATION AND RESULT**

**3.1** **Implementation**

The dataset used for this project contains 284,807 transactions, with 492 of them being fraudulent. The dataset includes 30 features, including the amount of the transaction, the time of the transaction, and various other details.

1. Data Preprocessing: The dataset was loaded into a Pandas Data Frame and the first five rows were displayed to understand the data. The dataset information was printed, and the number of missing values in each column was checked.

2. Data Separation: The data was separated into legitimate and fraudulent transactions.

3. Statistical Analysis: Statistical measures of the data were calculated for both legitimate and fraudulent transactions.

4. Visualization: A count plot was created to visualize the distribution of legitimate and fraudulent transactions. A correlation matrix heatmap was generated to show the correlation between features. A scatter plot was created to visualize the relationship between the class and time of transactions.

5. Under-Sampling: The dataset was under-sampled to balance the number of legitimate and fraudulent transactions.

6. Model Training: A Logistic Regression model was trained on the training data.

7. Model Evaluation: The accuracy of the Logistic Regression model was calculated on both the training and testing data.

**3.1.1 Data Preprocessing and Visualization:**

- First Five Rows: The first five rows of the dataset were displayed to understand the data.

- Dataset Information: The dataset information was printed, and the number of missing values in each column was checked.

- Distribution of Legitimate and Fraudulent Transactions: The distribution of legitimate and fraudulent transactions was analyzed.

- Correlation Matrix Heatmap: A correlation matrix heatmap was generated to show the correlation between features.

- Scatter Plot: A scatter plot was created to visualize the relationship between the class and time of transactions.

**3.1.2 Under-Sampling:**

- Number of Fraudulent Transactions: The number of fraudulent transactions was 492.

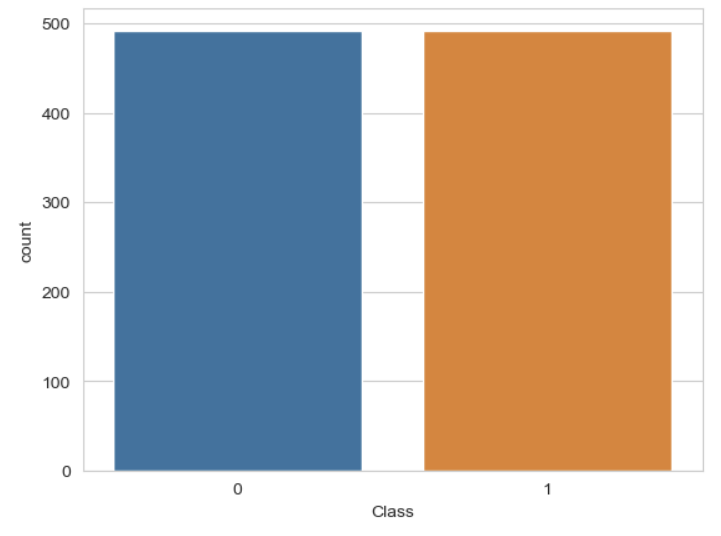
- Under-Sampling: The dataset was under-sampled to balance the number of legitimate and fraudulent transactions.

**3.1.3 Model Training and Evaluation:**

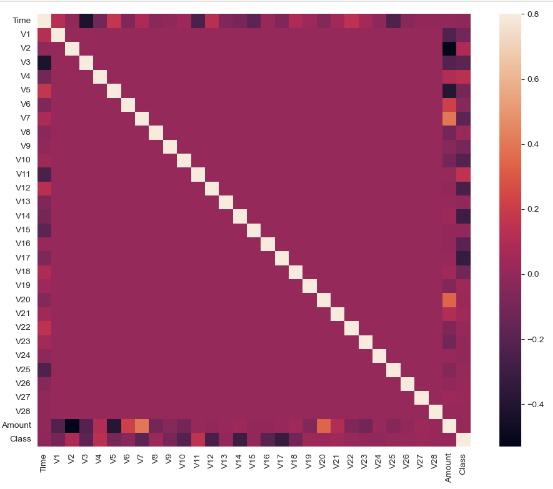
- Logistic Regression Model: A Logistic Regression model was trained on the training data.

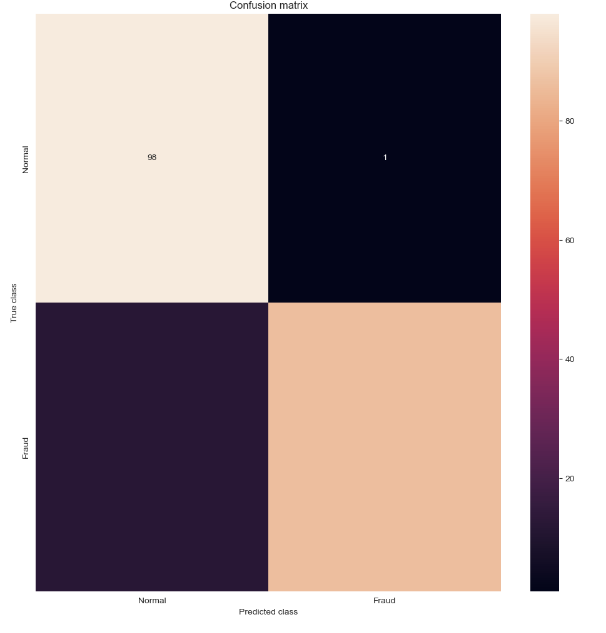
- Accuracy on Training Data: The accuracy of the Logistic Regression model on the training data was 0.98.

- Accuracy on Testing Data: The accuracy of the Logistic Regression model on the testing data was 0.96.

**3.2 Results**

**Figure 4: Distributed class**

**Figure 5: Correlation Matrix**

**Figure 6: Random Forest Classifier Confusion Matrix**

**CHAPTER 4**

**CONCLUSION**

**4.1 Advantages:**

With the development of electronic financial transaction technology and the emergence of simple payment, the risk of fraudulent payment and fraudulent payment increases as the authentication process is simplified. The types of fraudulent use of credit cards include theft and loss, identity theft, new card not received, card forgery, and card information theft. In particular, as phishing, pharming as well as card information leakage due to card information leakage, card information theft accidents are occurring. In response, the government tried to deal with electronic financial fraud by implementing the 'E-Financial Fraud Prevention Service'. It is difficult to cope with financial fraud by simply setting the existing keyboard security, public certificate, and additional password. The abnormal transaction detection system is used to analyze the user's data and payment data in real time to inform the financial institution and the user of the detection if it is different from the usual pattern, and further to arbitrarily stop the transaction. Therefore, an abnormal transaction detection system is important for fast and accurate detection, and research is needed to improve the algorithm. In this study, the method of detecting anomalous transactions using the electronic payment log analysis and machine learning technique was investigated. Results show the significance of algorithms used over the dataset and efficient classification is performed.

**4.2 Scope:**

In future deep learning concepts can be applied using convolution networks for improved accuracy. Also, some other datasets can be used for further testing of proposed mechanisms.

**Drive Link:**[**https://drive.google.com/drive/folders/1mEiiCfQd9IalJuNx\_9SshqEtps-BAFaX**](https://drive.google.com/drive/folders/1mEiiCfQd9IalJuNx_9SshqEtps-BAFaX)

**GitHub Link:**[**https://github.com/harshkhandelwal23/Credit\_Card\_Fraud\_Detection**](https://github.com/harshkhandelwal23/Credit_Card_Fraud_Detection)

**REFERENCES**

[1] Srivastava, A., Kundu, A., Sural, S., & Majumdar, A. K. (2020). Credit Card Fraud Detection Using Hidden Markov Model. International Research Journal of Engineering and Technology (IRJET), 7(8), 1251-1256.

[2] Donald V. Macdougall, Richard G. Mosley, Garioch J. l. Saunders; Credit card crime in Canada: Investigation - Prosecution; The Canadian Association of Crown Counsel; page 1-56; January 1985.

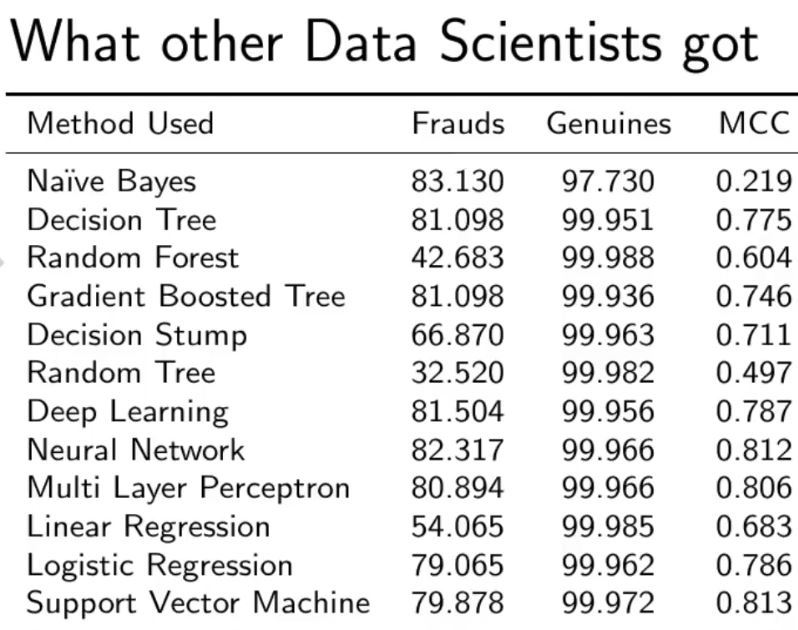
[3] Isabelle Sender; Detecting and combating fraud; Chain Store Age; New York; Vol. 74; Issue 7; Page 162; July 1998.

[4] Elford Dean, Raj Thomas, Lorry; Visa security center; Personal meetings; January 7 and February 11,1999.

[5] Gyusoo Kim and Seulgi Lee, “2014 Payment Research”, Bank of Korea, Vol. 2015, No. 1, Jan. 2015.

[6] EWT Nagi, Yong Hu, HY Wong, Yijun Chen, Xin Sun, “The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature,” Decision Support Systems, Vol. 50, No. 3, Feb. 2011.

**APPENDIX**

**Appendix A. Comparison with other algorithms without dealing with the imbalancing of the data.**

**Figure 7: Comparison with the other Algorithm**

As you can see with our Random Forest Model, we are getting a better result even for the recall which is the trickiest part.