Visvesvaraya Technological University, Belagavi-590018
"Jnana Sangama", Belagavi–590018

A PROJECT REPORT ON

" COUNTERFEITING DETECTION OF BANKING USING MACHINE LEARNING"

Submitted in partial fulfillment of the requirements for the 8th semester VTU CBCS Subject namely

MAJOR PROJECT PHASE 2

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE AND ENGINEERING

For the Academic year2023 - 2024

Submitted By

ABDU SAMI	**[4SH20CS001]**
ISHAAN HUSSAIN	**[4SH20CS023]**
MUHAMMED FARWAZ	**[4SH20CS035]**
NABEEL SAYED ANWAR	**[4SH20CS037]**

Under the Guidance of

Mr. ABHISHEK GOWDA R M

Assistant Professor
Department of CSE

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERINGSHREE DEVI INSTITUTE OF TECHNOLOGY
KENJAR, MANGALURU- 574 142

SHREE DEVI INSTITUTE OF TECHNOLOGY
KENJAR, MANGALURU- 574 142

(An Institution under VTU, Belagavi)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

Certified that the project work entitled **"COUNTERFEITING DETECTION OF BANKING USING MACHINE LEARNING"** is a bonafide work carried out by **ABDU SAMI, ISHAAN HUSSAIN , MUHAMMED FARWAZ, NABEEL SAYED ANWAR** bearing USN's **4SH20CS001,4SH20CS023, 4SH20CS035, 4SH20CS037** respectively in partial fulfilment for the VTU CBCS subject **Major Project Phase 2**, and for the award of degree of **Bachelor of Engineering** in **Computer Science and Engineering** of the Visvesvaraya Technological University, Belagavi during the year 2023-2024. It is certified that all corrections / suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the degree of Bachelor of Engineering.

Signature of Guide **Mr.Abhishek Gowda R M** **Asst. Professor** **Dept. Of CSE**	Signature of HOD **Prof. Anand S Uppar** **Head of the Department** **Dept. Of CSE**	Signature of Principal **Dr. K. E. Prakash** **The Principal**

EXTERNAL VIVA

Name of the Examiners	**Signature with Date**
1. _____	_____
2._____	_____

ACKNOWLEDGEMENT

A successful project is a fruitful culmination of the efforts of many people. Some directly involved and others who have quietly encouraged and extended their invaluable support throughout its progress.

We would also like to convey our heartfelt thanks to our **Management** for providing us with good infrastructure, laboratory facility, qualified and inspiring staff whose guidance was of great help in successful completion of this project.

We are extremely grateful and thankful to our beloved **Principal and Director Dr. K E Prakash** for providing a congenial atmosphere and also the necessary facilities for achieving the cherished goal.

We feel delighted to have this page to express my sincere thanks and deep appreciation to **Prof. Anand S. Uppar, Head of the Department, Computer Science and Engineering**, for his valuable guidance, keen interest and constant encouragement throughout the entire period of this project work.

We would like to thank our project guide **Mr. Abhishek Gowda R M, Asst Prof of Computer Science and Engineering Department** for her valuable guidance and constant support throughout the project work.

We are thankful to all the teaching and non-teaching staff for allowing us to successfully carryoutthe project work.

Finally, we also thank our family and friends who provided lot of support in this project work.

ABDU SAMI [4SH20CS001]
ISHAAN HUSSAIN [4SH20CS023]
MUHAMMED FARWAZ [4SH20CS035]
NABEEL SAYED ANWAR [4SH20CS037]

SHREE DEVI INSTITUTE OF TECHNOLOGY
KENJAR, MANGALURU- 574 142

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

We **Abdu Sami, Ishaan Hussain , Muhammed Farwaz, Nabeel Sayed Anwar** bearing USN's **4SH20CS001, 4SH20CS023, 4SH20CS035, 4SH20CS037** respectively, students of 8th semester Bachelor of Engineering, Computer Science and Engineering, Shree Devi Institute of Technology, Mangalore declare that the project work entitled **"Counterfeiting Detection Of Banking Using Machine Learning"** has been duly executed by us under the guidance of **Mr. Abhishek Gowda R M** Asst. Professor, Department of Computer Science and Engineering, Shree Devi Institute of Technology, Mangalore and submitted for the requirements for the 8th semester **Major Project Phase 2** of **Bachelor of Engineering in Computer Science Engineering** during the year 2023-2024.

Date:
Place:

ABDU SAMI **[4SH20CS001]**
ISHAAN HUSSAIN **[4SH20CS023]**
MUHAMMED FARWAZ **[4SH20CS035]**
NABEEL SAYED ANWAR **[4SH20CS037]**

ABSTRACT

Counterfeiting Detection of Banking Data by Machine Learning Techniques is a comprehensive study aimed at enhancing fraud detection processes within the banking sector using advanced machine learning methodologies. In an era where financial fraud poses a substantial threat to both institutions and customers, this research addresses the pressing need for more accurate and efficient fraud detection mechanisms. The study begins by framing the problem statement, emphasizing the critical importance of detecting fraudulent transactions promptly and with precision. It delves into an extensive literature review, summarizing prior research in the domain of fraud detection in banking, with a particular focus on the evolution of machine learning techniques in combating financial fraud.

This project is centered on the detection of credit card fraud transactions using logistic regression as the primary machine learning algorithm and the SMTP (Simple Mail Transfer Protocol) modulefor real-time communication with users. Given the escalating threat of fraudulent activities in creditcard transactions, this project aims to fortify the defenses of banking institutions by swiftly identifying and addressing suspicious activities. Logistic regression, renowned for its simplicity and interpretability, is harnessed as the cornerstone of the fraud detection system. By meticulously analyzing historical credit card transaction data, the logistic regression models excel at distinguishing between legitimate and fraudulent transactions, thereby bolstering security measures.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter-1

INTRODUCTION

1.1 OVERVIEW

Credit card is the most popular mode of payment. As the number of credit card users is rising world-wide, the identity theft is increased, and frauds are also increasing. In the virtual card purchase, only the card information is required such as card number, expiration date, secure code, etc. Such purchases are normally done on the Internet or over telephone. To commit fraud in these types of purchases, a person simply needs to know the card details. The mode of payment for online purchase is mostly done by credit card. The details of credit card should be kept private. To secure credit card privacy, the details should not be leaked. Different ways to steal credit card details are phishing websites, steal/lost credit cards, counterfeit credit cards, theft of card details, intercepted cards etc. For security purpose, the above things should be avoided. In online fraud, the transaction is made remotely and only the card's details are needed. The simple way to detect this type of fraud is to analyse the spending patterns on every card and to figure out any variation to the "usual" spending patterns.

Fraud detection by analysing the existing data purchase of cardholder is the best way to reduce the rate of successful credit card frauds. As the data sets are not available and also the results are not disclosed to the public. The fraud cases should be detected from the available data sets known as the logged data and user behaviour. At present, fraud detection has been implemented by a number of methods such as data mining, statistics, and artificial intelligence. Machine learning has emerged as a powerful tool in this domain, leveraging advanced algorithms to analyse vast datasets and identify anomalous patterns indicative of fraudulent behaviour. By employing predictive modelling and anomaly detection techniques, machine learning algorithms can learn from historical data, adapt to evolving fraud tactics, and enhance the accuracy of fraud detection systems. This proactive approach enables banks to stay ahead of fraudsters, minimize financial losses, and ensure a secure and trustworthy financial environment for their clients.

Chapter-2

LITERATURE SURVEY

2.1 LITERATURES

Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective Authors: SamanehSorournejad, Zahra Zojaji, Amir Hassan Monadjemi.

In this paper, after investigating difficulties of credit card fraud detection, we seek to review the state of the art in credit card fraud detection techniques, datasets and evaluation criteria. In this section, the authors, Samaneh Sorournejad, Zahra Zojaji, and Amir Hassan Monadjemi, delve into a comprehensive exploration of credit card fraud detection techniques with a specific focus on both data and technique-oriented perspectives. The primary objective is to address the challenges associated with credit card fraud detection and present a critical review of the current state of the art, encompassing techniques, datasets, and evaluation criteria.

Disadvantages are

→ Lack of standard metrics.

Detection of credit card fraud: State of art Authors: Imane Sadgali, Nawal Sael, Faouzia Benabbau

In this paper, we propose a state of the art on various techniques of credit card fraud detection. The purpose of this study is to give a review of implemented techniques for credit card fraud detection, analyses their incomes and limitless, and synthesize the finding in order to identify the techniques and methods that give the best results so far. In this research paper authored by Imane Sadgali, Nawal Sael, and Faouzia Benabbau, the primary focus is on providing a comprehensive overview of various techniques employed in the realm of credit card fraud detection. The authors aim to conduct a state-of-the-art analysis, offering a critical review of implemented methods, analyzing their strengths and limitations, and synthesizing findings to identify the most effective techniques and methods in the current landscape.

→ Lack of adaptability.

Credit card fraud detection using machine learning algorithm Authors: Vaishnavi Nath Dornadulaa, Geetha S.

The main aim of the paper is to design and develop a novel fraud detection method for Streaming Transaction Data, with an objective, to analyse the past transaction details of the customers and extract the behavioural patterns. Authored by Vaishnavi Nath Dornadulaa and Geetha S., this research paper is centered around the design and development of a novel fraud detection method specifically tailored for Streaming Transaction Data. The primary objective is to leverage machine learning algorithms to analyze the historical transaction details of customers and extract meaningful behavioral patterns that can aid in the identification of fraudulent activities.

→ Imbalanced Data.

Fraudulent Transaction Detection in Credit Card by Applying Ensemble Machine Learning techniques Authors: Debachudamani Prusti, Santanu Kumar Rath

In this study, the application of various classification models is proposed by implementing machine learning techniques to find out the accuracy and other performance parameters to identify the fraudulent transaction. This research, authored by Debachudamani Prusti and Santanu Kumar Rath, focuses on the application of ensemble machine learning techniques for the detection of fraudulent transactions in credit cards. The study proposes the utilization of various classification models within the framework of machine learning to assess their accuracy and other performance parameters, ultimately aiming to enhance the identification of fraudulent transactions.

→ Overlapping data.

Detection of Credit Card Fraud Transactions using Machine Learning Algorithms and Neural Networks Authors: Deepti Dighe, Sneha Patil, Shrikant Kokate

Credit card fraud resulting from misuse of the system is defined as theft or misuse of one's credit card information which is used for personal gains without the permission of the card holder. To detect such frauds, it is important to check the usage patterns of a user over the past transactions. Comparing the usage pattern and current transaction, we can classify it as either fraud or a legitimate transaction. Authored by Deepti Dighe, Sneha Patil, and Shrikant Kokate, this research paper addresses the crucial issue of credit card fraud detection. The authors emphasize the definition of credit card fraud resulting from the misuse of the system, specifically the theft or unauthorized use of credit card information for personal gains. The

central premise of the study revolves around the importance of analyzing user transaction patterns to distinguish between legitimate and fraudulent transactions.

→ Different misclassification importance.

Credit card fraud detection using machine learning algorithms and cyber security Authors: JiatongShen

As they have the same accuracy the time factor is considered to choose the best algorithm. By considering the time factor they concluded that the Adaboost algorithm works well to detect credit card fraud. Authored by Jiatong Shen, this research paper investigates the intersection of credit card fraud detection, machine learning algorithms, and cybersecurity. The study not only focuses on achieving high accuracy in fraud detection but also emphasizes the importance of considering the time factor when selecting the most suitable algorithm. The conclusion drawn is that, among various algorithms, the Adaboost algorithm stands out as particularly effective in the timely detection of credit card fraud.

→ Accuracy is not perfect.

Chapter-3

PROBLEM STATEMENT AND SOLUTION STRATEGY

3.1 PROBLEM STATEMENT

In the dynamic landscape of banking, the persistent rise in fraudulent activities, ranging from identity theft to intricate financial scams, poses a formidable challenge. Traditional methods of fraud detection have become increasingly inadequate in addressing the evolving strategies employed by fraudsters. This inadequacy has resulted in substantial financial losses and a decline in customer confidence. Recognizing this pressing issue, there is an urgent imperative to implement cutting-edge solutions, notably harnessing the power of machine learning techniques. The primary challenge lies in developing models that can dynamically adapt to emerging fraud patterns, efficiently process vast and diverse datasets, and deliver real-time identification of suspicious transactions. As a result, there is an urgent need to implement advanced solutions, particularly leveraging machine learning techniques, to bolster the effectiveness of fraud detection in the banking sector. The challenge lies in developing models capable of dynamically adapting to new fraud patterns, handling massive and diverse datasets, and providing real-time identification of suspicious transactions to mitigate risks and fortify the security of financial systems.

3.2 EXISTING SYSTEM

→ Since the credit card fraud detection system is a highly researched field, there are many different algorithms and techniques for performing the credit card fraud detection system.

→ One of the earliest systems is CCFD system using Markov model. Some other various existing algorithms used in the credit cards fraud detection system includes Cost sensitive decision tree (CSDT).

→ Credit card fraud detection (CCFD) is also proposed by using neural networks. The existing credit card fraud detection system using neural network follows the whale swarm optimization algorithm to obtain an incentive value.

3.3 LIMITATIONS

If the time interval is too short, then Markov models are inappropriate because the individual displacements are not random, but rather are deterministically related in time. This example suggests that Markov models are generally inappropriate over sufficiently short time intervals.

3.4 SOLUTION STRATEGY

A robust solution strategy for counterfeiting detection in banking using machine learning involves a multifaceted approach that combines advanced analytics, anomaly detection, and real-time monitoring. Firstly, a comprehensive dataset comprising historical transaction data, customer profiles, and other relevant information should be compiled. Feature engineering plays a crucial role in extracting meaningful patterns from this data, enhancing the model's ability to discern normal from fraudulent activities.

Supervised machine learning algorithms, such as logistic regression or random forests, can be trained on labeled datasets to identify patterns associated with known fraud cases. Additionally, unsupervised learning techniques, like clustering and autoencoders, prove beneficial in detecting anomalies and previously unseen fraud patterns. The model's performance can be further refined through continuous learning, incorporating new data and adapting to evolving fraud tactics.

Real-time monitoring is essential for prompt detection and response. Implementing a scoring system that assigns risk scores to transactions based on their deviation from normal behavior enables immediate identification of suspicious activities. This can be complemented by leveraging technologies like Natural Language Processing (NLP) to analyze textual data, such as transaction descriptions or customer communications, for additional context.

Collaboration with industry peers for shared threat intelligence and leveraging external data sources, such as device fingerprints or geolocation information, enhances the model's predictive power. Finally, incorporating explain ability features into the model ensures transparency, aiding investigators in understanding the reasoning behind flagged transactions. Regular model audits, updates, and continuous improvement based on feedback and emerging fraud trends solidify the system's effectiveness in combating fraudulent activities in the dynamic landscape of banking.

Chapter-4

PROPOSED SYSTEM

The proposed system for credit card fraud detection utilizes machine learning algorithms to enhance the accuracy and efficiency of fraud detection processes. Leveraging advanced data analytics, the system analyses a wide array of features and patterns associated with credit card transactions to identify anomalies indicative of fraudulent activity. This includes transaction amount, frequency, location, and user behaviour, among other parameters. Machine learning models, such as Random Forest, SVM or Logistic Regression, are trained on historical data to learn and adapt to evolving fraud patterns.

In the event that a transaction is flagged as fraudulent by our system's fraud detection mechanisms, an automated process is initiated to promptly notify the customer via email. This notification serves as a critical communication channel to alert the customer about the suspicious activity detected on their account. By promptly informing the customer, we aim to mitigate any potential financial losses and reassure them of our commitment to their security and protection. Real-time monitoring of transactions allows the system to promptly flag suspicious activities, triggering alerts for further investigation by fraud detection teams. Additionally, the system employs adaptive learning mechanisms, continually updating its models based on new data to stay resilient against emerging fraud tactics. The integration of machine learning in credit card fraud detection not only improves accuracy but also enables faster response times, ultimately enhancing the security of financial transactions for cardholders and financial institutions alike.

4.1 PROPOSED MODEL

Logistic Regression is a widely employed machine learning algorithm for fraud detection in the banking sector. This algorithm is particularly suitable for binary classification problems, making it an apt choice for distinguishing between legitimate and fraudulent transactions. In the context of fraud detection, logistic regression leverages historical transaction data to learn patterns and relationships between various features, such as transaction amount, location, time, and user behavior. By analyzing these features, the algorithm calculates probabilities and assigns a likelihood score to each transaction. Accuracy of training dataset: 94.95611109160493 Accuracy of test dataset: 90.95611109160493 Precision:

77.01149425287356 Recall: 93.05555555555556 F1-Score: 0.8427672955974842 AUC score: 96.51019999609383.

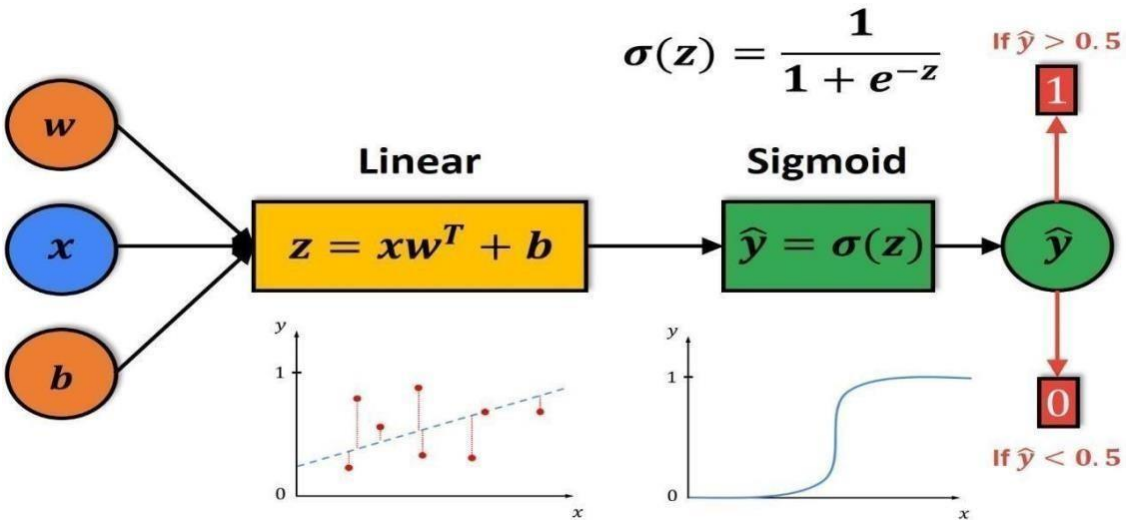

Fig: 4.1 Logistic Regression Model Representation

4.2 PROJECT MODULES

Project Modules Entire project is divided into 3 modules as follows: Training the model using the Machine Learning algorithm Logistic Regression,

Module 1: Data Gathering and Data Preprocessing

a. A proper dataset is searched among various available ones and finalized with the dataset.

b. The dataset must be pre-processed to train the model.

c. In the preprocessing phase, the dataset is cleaned and any redundant values, noisy data and null values are removed.

d. The Pre-processed data is provided as input to the module.

Module 2: Training the model

a. The Pre-processed data is split into training and testing datasets in the 80:20 ratio to avoid the problems of over-fitting and under-fitting.

b. A model is trained using the training dataset with algorithm Logistic Regression.

c. The trained models are trained with the testing data and results are visualized using bar graphs, scatter plots.

d. The accuracy rates of the algorithm are calculated using different params like F1 score, Precision, Recall. The results are then displayed using various data visualization tools for analysis purpose.

e. This algorithm has provided the better accuracy rate compared to remaining algorithms is taken as final prediction model.

Module 3: Final Prediction model integrated with front end

a. The algorithm which has provided better accuracy rate has considered as the final prediction model.

b. The model thus made is integrated with front end (Flask).

4.3 PERFORMANCE ANALYSIS

Performance metrics: The basic performance measures derived from the confusion matrix. The confusion matrix is a 2 by 2 matrix table contains four outcomes produced by the binary classifier. Various measures such as sensitivity, specificity, accuracy and error rate are derived from the confusion matrix.

Accuracy: Accuracy is calculated as the total number of two correct predictions(A+B) divided by the total number of the dataset(C+D). It is calculated as (1-error rate).

Accuracy=A+B/C+D

Whereas, A=True Positive B=True Negative C=Positive D=Negative

Error rate: Error rate is calculated as the total number of two incorrect predictions(F+E) divided by the total number of the dataset(C+D).

Error rate=F+E/C+D

Whereas, E=False Positive F=False Negative C=Positive D=Negative

Sensitivity: Sensitivity is calculated as the number of correct positive predictions(A) divided by the total number of positives(C).

Sensitivity=A/C

Specificity: Specificity is calculated as the number of correct negative predictions(B) divided by the total number of negatives(D).

Specificity=B/D

Precision: The ratio of true positive predictions to the total predicted positives.Calculation:

True Positives

True Positives + False Positives

True Positives + False Positives

True Positives

F1 Score: The harmonic mean of precision and recall, providing a balance between the two metrics.

Area Under the Receiver Operating Characteristic (ROC) Curve (AUC-ROC): A measure of the model's ability to distinguish between classes. It plots the true positive rate against the false positive rate.

Interpretation: A higher AUC-ROC indicates better discrimination between classes.

Confusion Matrix: A matrix that summarizes the number of true positives, true negatives, false positives, and false negatives.

Cross-Validation: Using techniques like k-fold cross-validation to assess the model's performance on different subsets of the data.

Learning Curve Analysis: Examining how the model's performance changes as the size of the training data increases.

Chapter-5

SYSTEM REQUIREMENTS ANALYSIS AND SPECIFICATIONS

5.1 HARDWARE REQUIREMENTS:

- Processor : Intel i3 or higher, with a clock speed of 2.1 GHz or above
- Memory : Recommended minimum of 4 GB RAM or higher
- Storage : A minimum storage of 30 GB HDD or more

5.2 SOFTWARE REQUIREMENTS:

- Operating System : Windows 10 or Linux based system
- Language : Python 3.6 or above, HTML,CSS
- Python Libraries : Pandas, NumPy, Matplotlib, SMTP
- IDE : VS CODE
- Technologies Used : Flask

Chapter-6

SYSTEM DESIGN

6.1 ARCHITECTURE DIAGRAM

Our Project main purpose is to making Credit Card Fraud Detection awaring to people from credit card online frauds. the main point of credit card fraud detection system is necessary to safe our transactions & security. This model is then used to identify whether a new transaction is fraudulent or not. Our aim here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications.

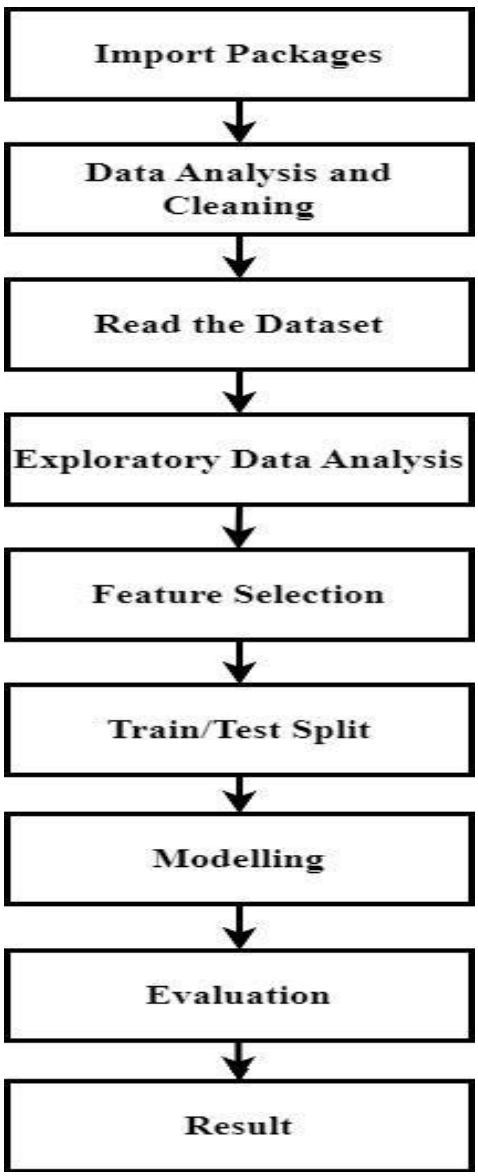

Fig: 6.1 Architecture Diagram

6.2 USE-CASE DIAGRAM

In UML, use-case diagrams model the behaviour of a system and help to capture the requirements of the system. Use-case diagrams describe the high-level functions and scope of a system. These diagrams also identify the interactions between the system and its actors. The use cases and actors in use-case diagrams describe what the system does and how the actors use it, but not how the system operates internally. Use-case diagrams illustrate and define the context and requirements of either an entire system or the important parts of the system.

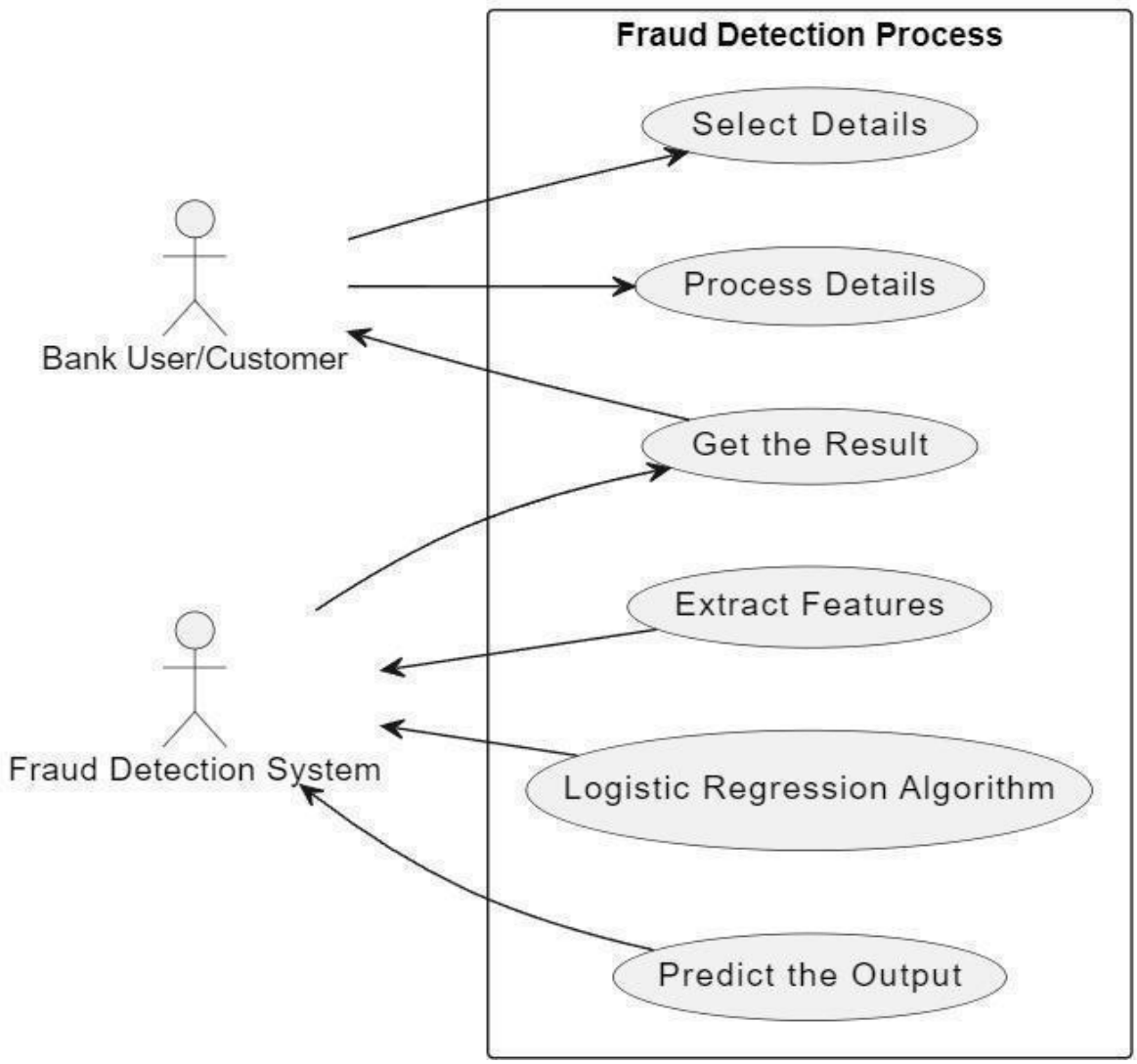

Fig: 6.2 Use Case Diagram

6.3 SEQUENCE DIAGRAM

A sequence diagram is a fundamental tool in system design, providing a graphical representation of the dynamic interactions between various components or objects. It showcases the flow of messages within the system, often termed as an "event diagram," enabling stakeholders to envision multiple scenarios. At its core, the diagram illustrates communication between lifelines, each representing an object or component, through time-ordered events. In UML notation, lifelines are depicted as vertical bars, while message flow is indicated by dotted lines extending horizontally. This representation not only captures the sequence of events but also highlights the timing and dependencies between them. Additionally, sequence diagrams accommodate iterations and branching, offering a detailed depiction of system behavior under different conditions. Overall, sequence diagrams serve as a collaborative tool, facilitating communication and decision-making among stakeholders, developers, and designers throughout the software development process.

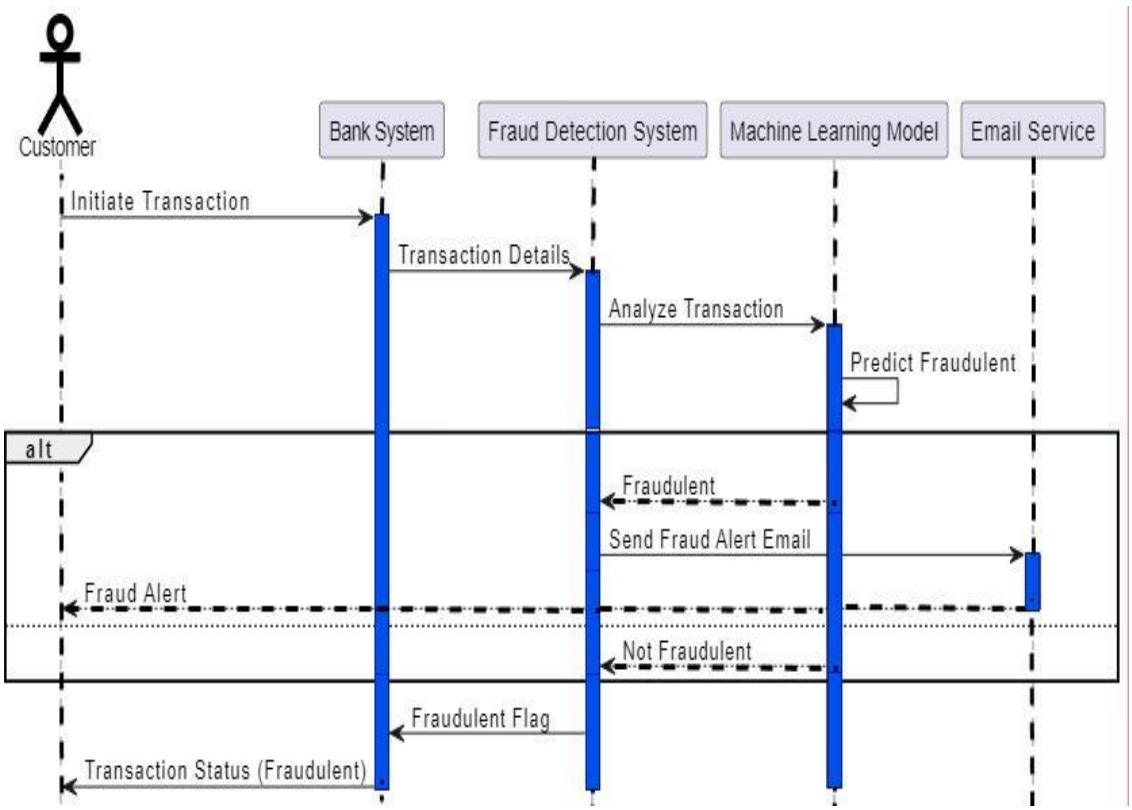

Fig: 6.3 Sequence Diagram

6.4 DATAFLOW DIAGRAM

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It can be manual, automated, or a combination of both. It shows how data enters and leaves the system, what changes the information, and where data is stored. Ensemble methods such as random forests and gradient boosting combine multiple classifiers to improve accuracy and robustness. The objective of a DFD is to show the scope and boundaries of a system as a whole. It may be used as a communication tool between a system analyst and any person who plays a part in the order that acts as a starting point for redesigning a system. The DFD is also called as a data flow graph or bubble chart. This are often used as a first step towards redesigning a system.

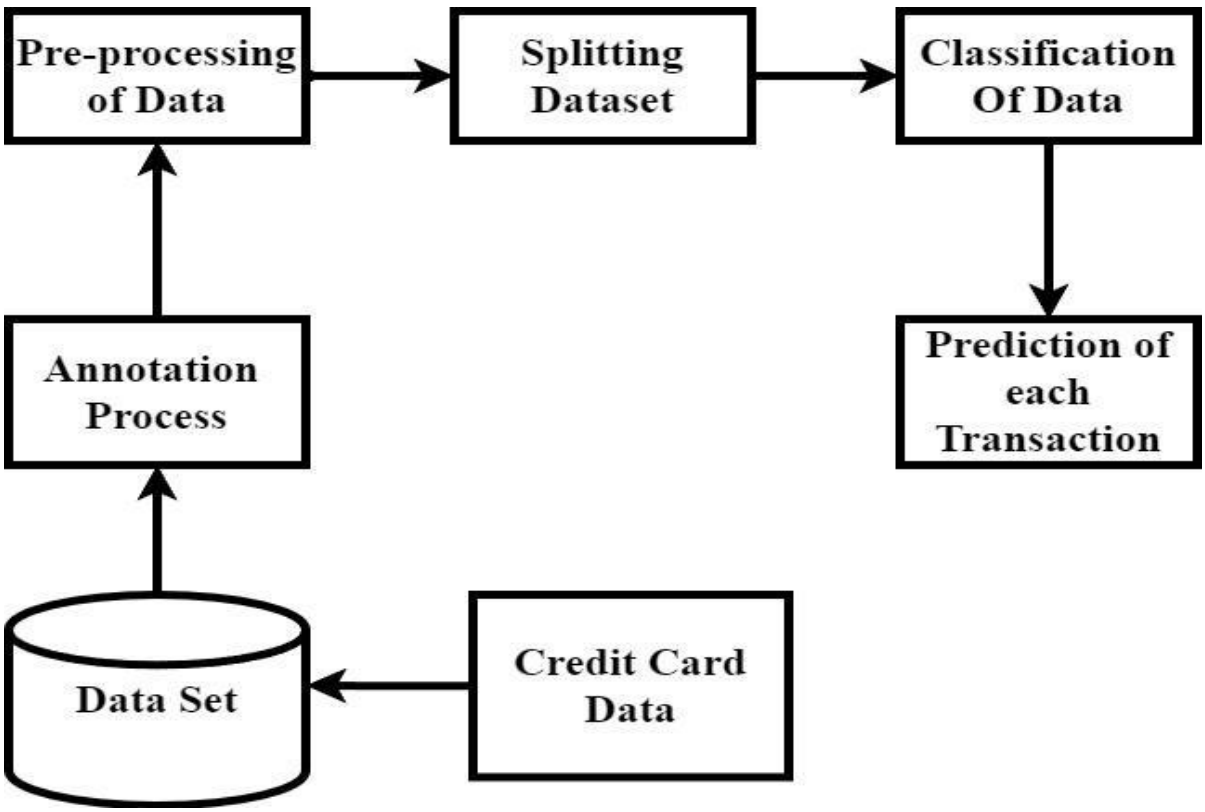

Fig: 6.4 Data Flow Diagram

Chapter-7

SYSTEM IMPLEMENTATION AND TESTING

PRE-PROCESSING:

- Data cleaning: Handle missing values, outliers, and inconsistencies in the dataset to ensure data quality.
- Feature engineering: Create new features and transform existing ones to provide relevant information for the classification model.
- Normalization and standardization: Scale numerical features like transaction amount toa similar range to prevent certain features from dominating the model.
- Handling imbalanced data: Address the issue of having more legitimate transactionsthan fraudulent ones by using techniques like oversampling or under sampling.
- Feature selection: Identify the most significant features that contribute to the classification of fraudulent transactions.

CLASSIFICATION:

- In this Project classification is done using Logistic regression which is basically a supervised classification algorithm. In a classification problem, the target variable (or output), y, can take only discrete values for given set of features (or inputs), X.
- Contrary to popular belief, logistic regression is a regression model. The model builds a regression model to predict the probability that a given data entry belongs to the category numbered as “1”(Legitimate) and “0”(Fraudulent).
- Model evaluation involves metrics like accuracy, precision, recall, and F1-score to assess the classifier's performance. Ensemble methods such as random forests and gradient boosting combine multiple classifiers to improve accuracy and robustness.
- Within the realm of classification, algorithms can be broadly categorized into binary classification, where the model predicts between two classes, and multi-class classification, where there are multiple possible outcomes.

GUI DESIGN:

- Develop the user interface using HTML/CSS to create a visually appealing and intuitive design.

- Include input fields for users to enter transaction details such as amount and time, as well as buttons to submit queries.
- Provide visual feedback, such as loading indicators or success/error messages, to enhance user experience.
- User-friendly features such as customizable filters, search functionality, and drill-down capabilities enhance usability and efficiency.
- GUI elements include dashboards displaying real-time transaction data, alerts for suspicious activities, and options for manual review and investigation.

FLASK BACKEND:

- Use Flask, a lightweight Python web framework, to build the backend server responsible for handling requests from the frontend.
- Implement routes to process incoming data from the GUI and integrate the logistic regression model for fraud detection.
- Flask is a lightweight and versatile web framework for building web applications and APIs in Python. It provides the tools and libraries necessary to create web services with minimal overhead and a simple, yet powerful, architecture.
- Flask is known for its simplicity, flexibility, and ease of use, making it an excellent choice for both beginners and experienced developers alike.

PREDICTION:

- After creating a logistic regression model, SVM model and KNN model it should be trained using the dataset. After that by evaluating all the modules the accuracy can be found then concluding by comparing the accuracy of the model with high performance
- Utilize the trained logistic regression model to predict whether the transaction is fraudulent or legitimate based on the provided features.
- Continuous monitoring and retraining of predictive models ensure they remain effective in detecting evolving fraud schemes.
- Time-series forecasting techniques anticipate trends and anomalies in transaction behavior, enabling proactive risk management
- Prediction is the final stage of system implementation, where the trained classification model is deployed to predict the likelihood of fraud for new transactions in real-time.
- The prediction results are integrated into the banking system's workflow, enabling timely detection and prevention of fraudulent activities.

- Continuous monitoring and updating of the model are essential to maintain its accuracy and effectiveness over time.

SMTP:

- SMTP (Simple Mail Transfer Protocol) serves as the backbone for sending emails in most digital communication systems. Integrating SMTP into our project enables the seamless transmission of emails.
- This connection facilitates the transfer of email messages from our application to the recipient's mail server. Python offers the smtplib module, simplifying SMTP integration with its intuitive API.
- Through this module, we can craft email messages and dispatch them via the designated SMTP server, empowering our project with the capability to deliver emails efficiently and reliably.
- we can compose email messages within our application and dispatch them via the designated SMTP server. This capability empowers our project to deliver emails efficiently and reliably, ensuring effective communication with our users or clients.

FEATURES/ATTRIBURES:

In our project, we've integrated various features as input, including numerical values such as time, account number, credit card number, transaction date, credit limit, transaction category, transaction ID, card expiry date, CVV, account type, and transaction amount. These inputs collectively form a comprehensive dataset that aids in analyzing and processing financial transactions efficiently.

- Time: The timestamp indicating the time at which a transaction occurred, providing temporal context to the data.
- Account Number: Unique identifier associated with the account involved in the transaction, facilitating tracking and identification.
- Credit Card Number: Unique numerical code assigned to the credit card used for the transaction, ensuring secure payment processing.
- Transaction Date: Date on which the transaction took place, contributing to chronological analysis and reporting.
- Credit Limit: Maximum amount of credit available to the account holder, influencing transaction approval and credit risk assessment.

- Transaction Category: Classification of the transaction based on its nature or purpose (e.g., retail, dining, entertainment), aiding in expenditure analysis and budgeting.
- Transaction ID: Unique identifier assigned to each transaction, enabling traceability and audit trail.
- Card Expiry Date: Date indicating the expiration of the credit card, essential for card validation and renewal.
- CVV (Card Verification Value): Security code associated with the credit card, used for authentication during online transactions.
- Account Type: Classification of the account (e.g., savings, checking, credit card), influencing transaction processing and fee structures.
- Transaction Amount: The numerical value representing the monetary value of the transaction, essential for financial analysis, budgeting, and fraud detection.

TESTING:

- Testing the project involves validating the functionality and accuracy of variousfeatures incorporated into our system, including numerical inputs like time, account number, credit card details, transaction dates, credit limits, transaction categories, IDs, card expiry dates, CVV, account types, and transaction amounts.
- This involves verifying the correct handling of timestamps for temporal relevance, validating the uniqueness and integrity of account and transaction identifiers, and confirming the accurate processing of credit card information while maintaining data privacy and security.
- In testing credit card transactions, we assess the system's ability to validate card expiry dates and CVV codes accurately, ensuring adherence to payment security standards. Moreover, we scrutinize the processing of transaction categories and amounts, verifying that transactions are categorized correctly and that the system accurately handlesvarying transaction values and types.
- Throughout the testing process, we prioritize robustness and accuracy, rigorously evaluating the system's performance under different scenarios and edge cases. By conducting thorough testing, we aim to identify and rectify any potential issues or vulnerabilities, thereby enhancing the reliability, security, and efficiency of our financial transaction system.

Test case	Input	Test description	Output
1	Click on Check Transactions button	Navigates into the form and we can enter the details	Predicts the whether transaction is Fraudulent or Legitimate.
2	Click on Send Mail	Mail sent to the recipient if the transaction fraudulent	Displays on the screen as Mail has been sent.
3	Click on Bank rules and regulations	The objective of this test is to verify the functionality of the "Bank Rules and Regulations" button within the application.	Rules and Regulations of the Bank are displayed on the window.
4	Click on Data set graph	Graph is plotted blw the fraudulent and legitimate values in the dataset.	Graph is displayed.

Fig: 7.1 Testing Cases

Chapter 8

RESULTS AND DISCUSSION

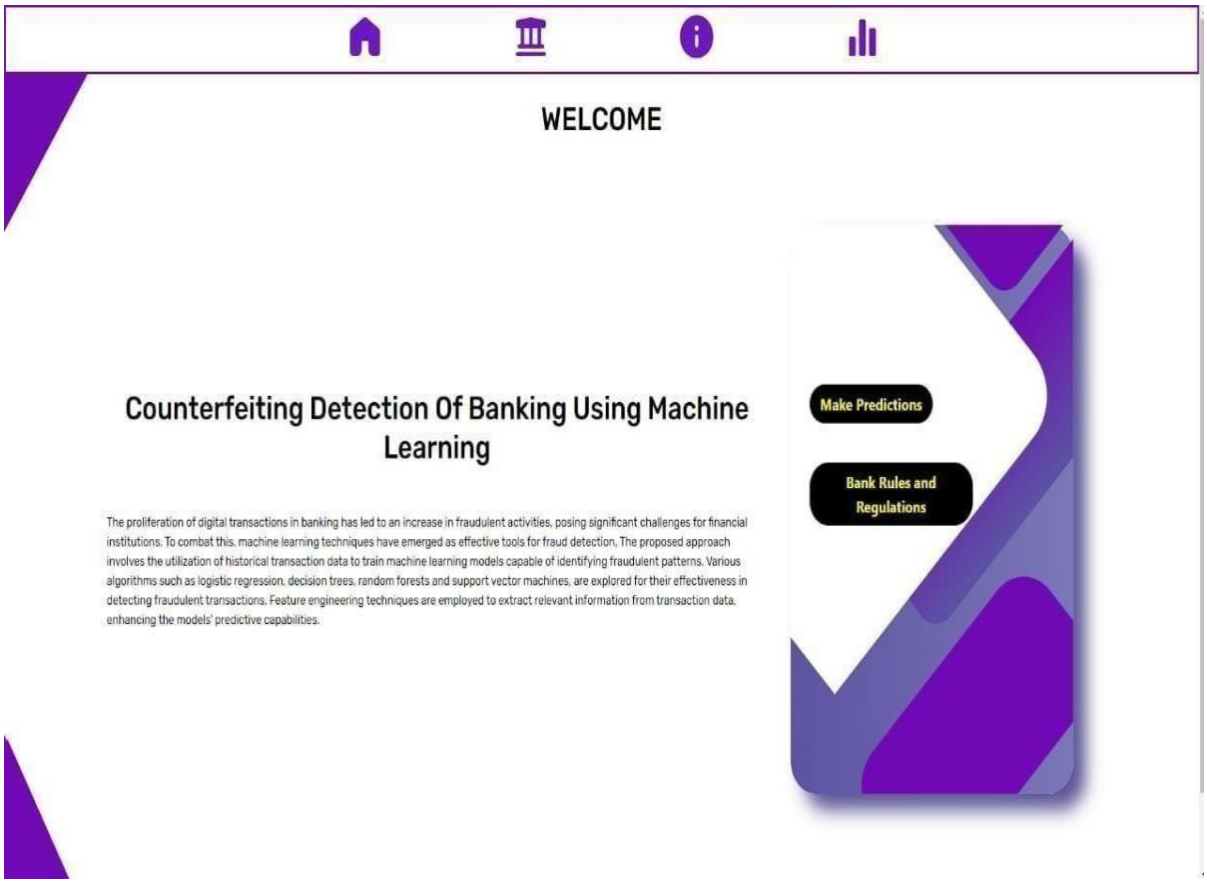

Fig: 8.1 Main Page

Here there is one Navbar which consists of the Home, Banking Application and About us sections. It consists of two buttons namely Check transaction and Bank Rules and Regulations.

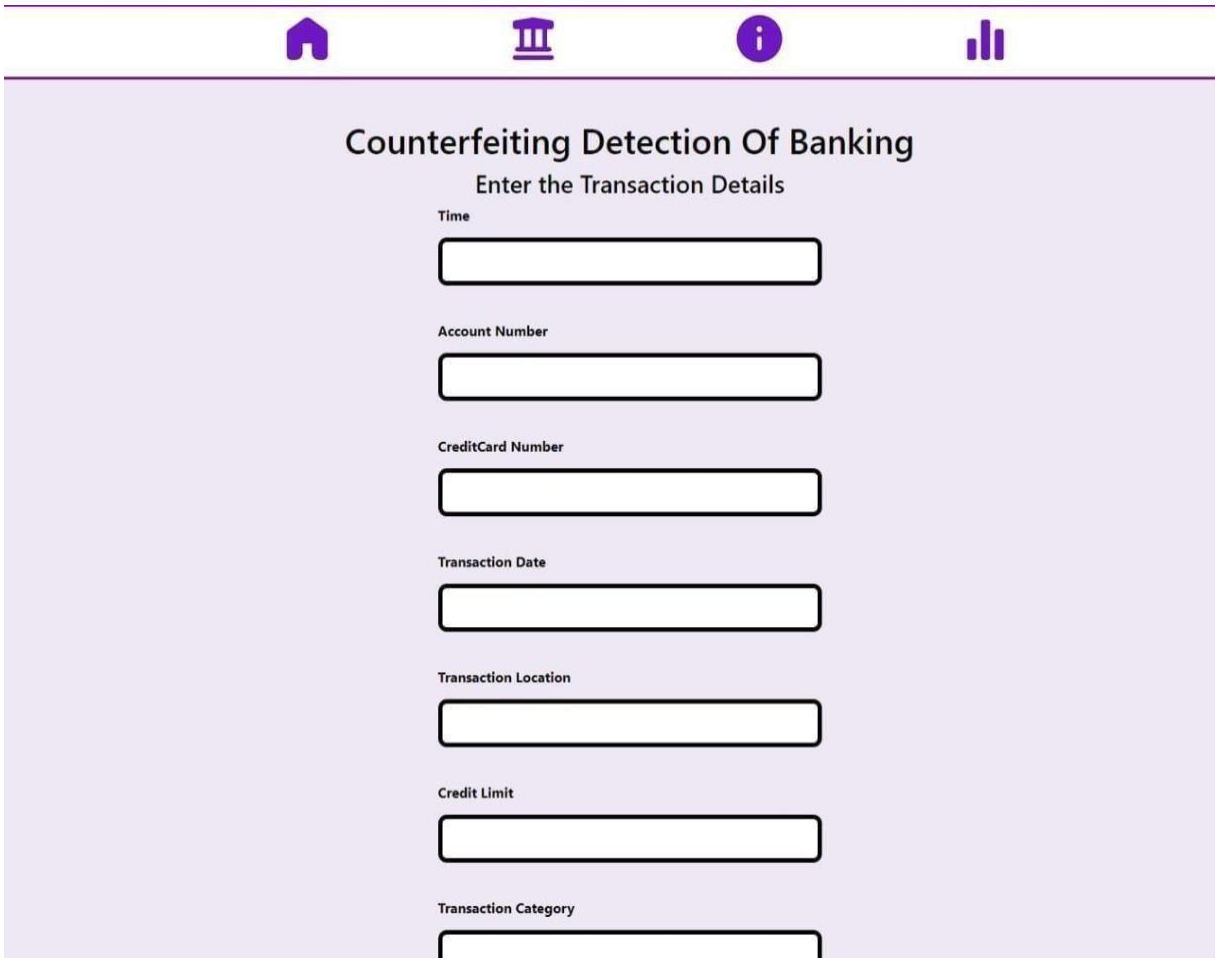

Fig: 8.2 Transaction Details Form

As we integrate various features into our project, we capture a range of numerical inputs essential for processing financial transactions effectively. These inputs encompass crucial details such as the transaction time, account number, credit card number, transaction date, credit limit, transaction category, transaction ID, card expiry date, CVV, account type, and transaction amount. Each of these numerical values plays a distinct role in facilitating the analysis and management of financial data. For instance, the timestamp associated with each transaction provides temporal context, while the account number and credit card number serve as unique identifiers for tracking and validation purposes.

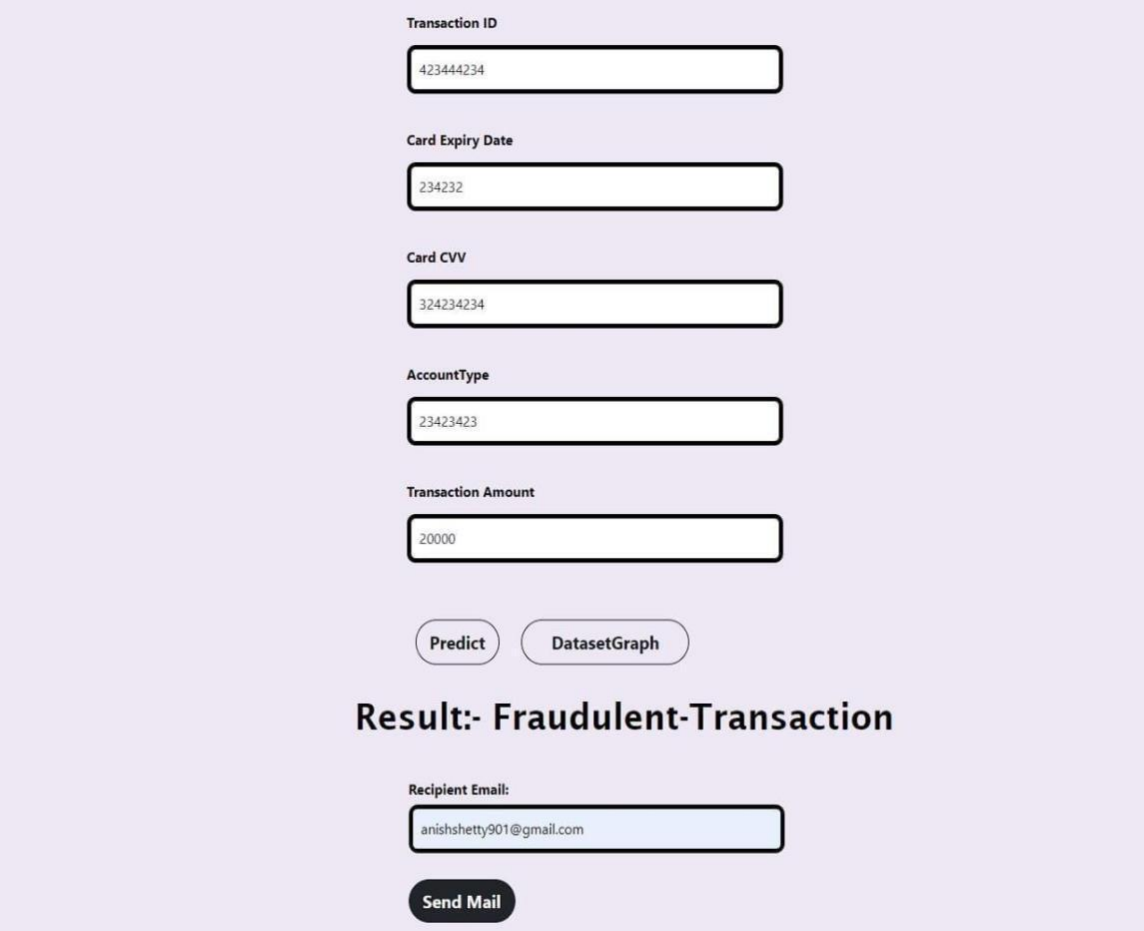

Fig: 8.3 Send mail

In the event that a transaction is flagged as fraudulent by our system's fraud detection mechanisms, an automated process is initiated to promptly notify the customer via email. This notification serves as a critical communication channel to alert the customer about the suspicious activity detected on their account. By promptly informing the customer, we aim to mitigate any potential financial losses and reassure them of our commitment to their security and protection.

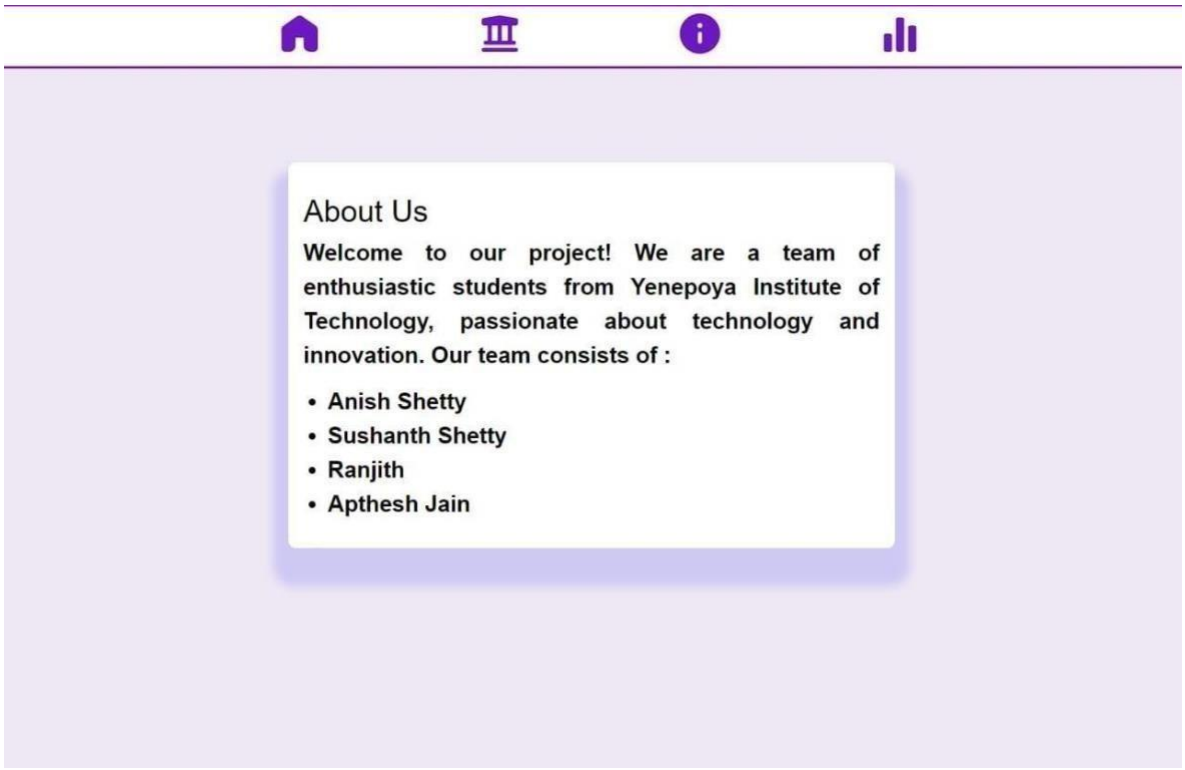

Fig: 8.4 About Us Section

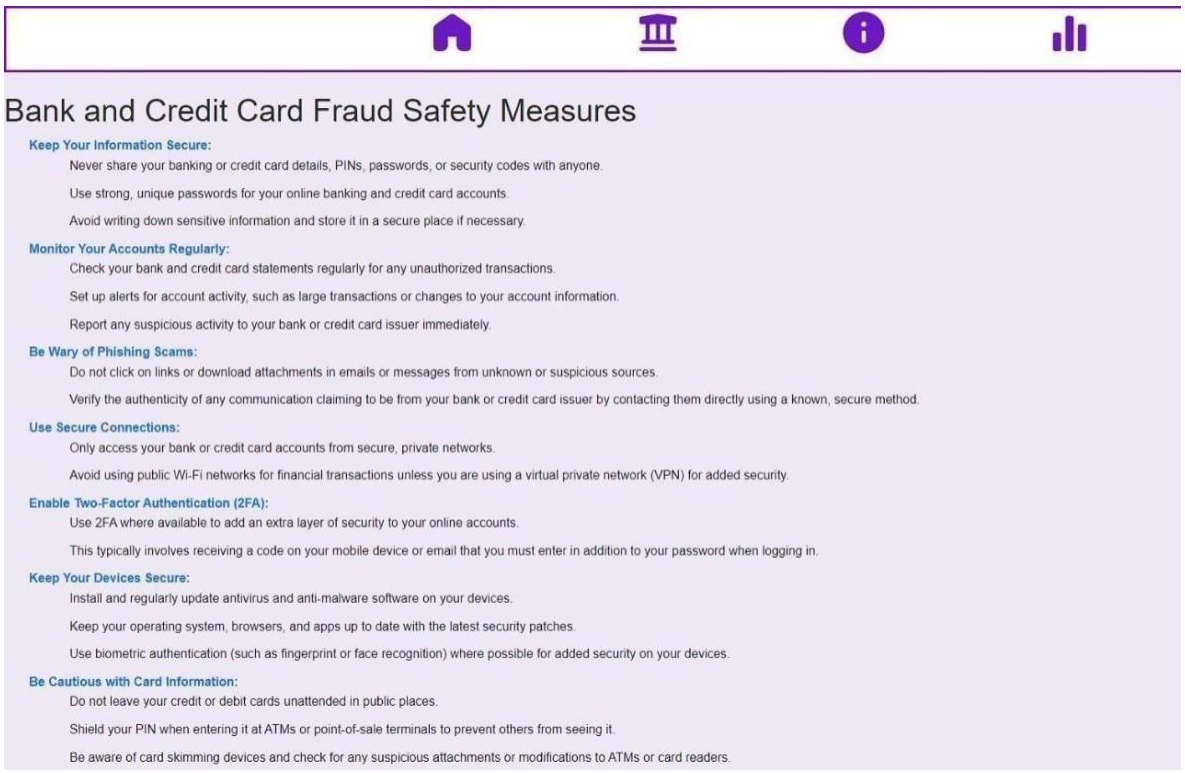

Fig: 8.5 Bank rules and regulations Section

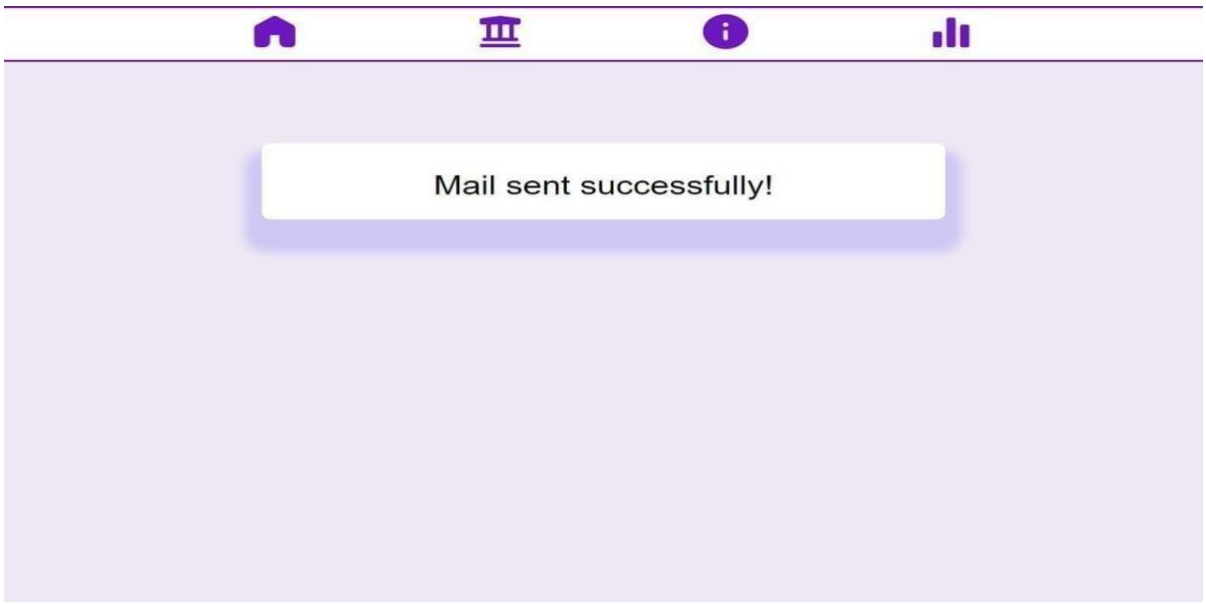

Fig: 8.6 Mail Confirmation

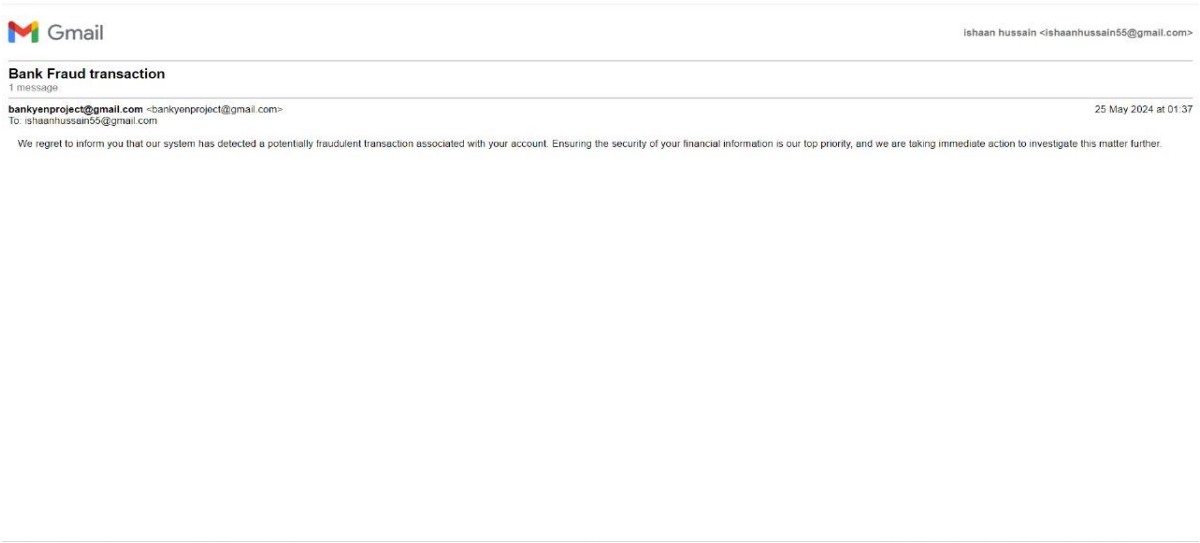

Fig : 8.7 Mail Received

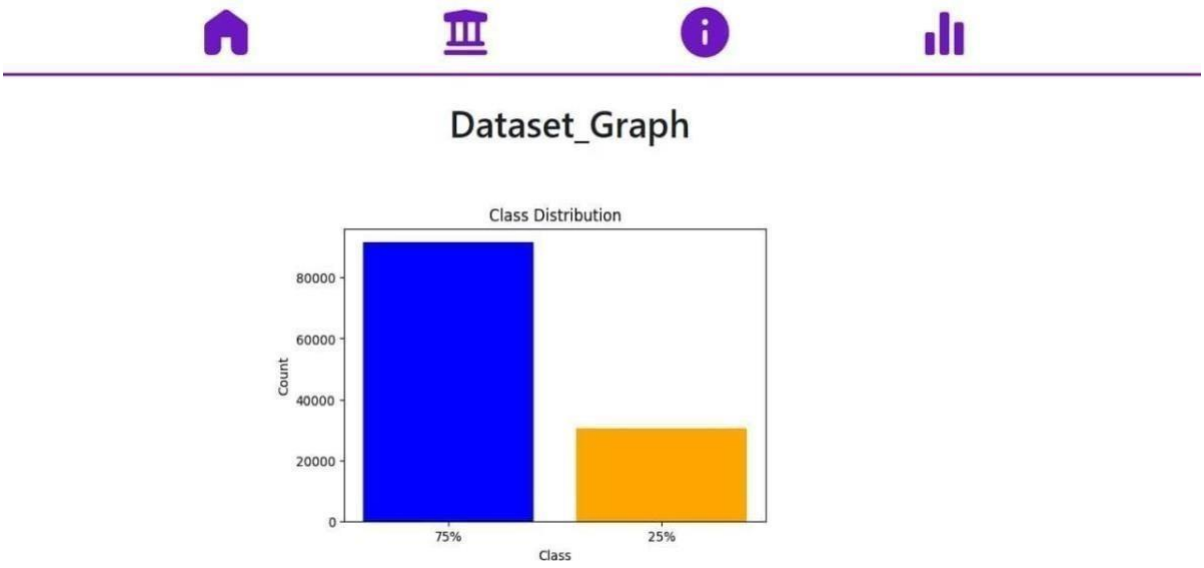

Fig : 8.8 Dataset Graph

CONCLUSION AND FUTURE SCOPE

In conclusion, the implementation of fraud transaction detection in banking using machine learning, specifically leveraging the logistic regression algorithm and Flask in the backend, marks a significant step forward in enhancing security and trust within financial institutions. By harnessing the power of machine learning, we have developed a robust system capable of effectively identifying fraudulent transactions, thereby safeguarding the interests of both the bank and its customers.Through the utilization of logistic regression, we have successfully constructed a predictive model that analyzes various features associated with each transaction, such as time, account number, transaction amount, and more. This model has been trained on historical data, allowing it to learn patterns indicative of fraudulent behavior and make accurate predictions in real-time. Additionally, by deploying Flask in the backend, we have created a seamless and efficient system for handling incoming transaction data, processing it through the predictive model, and delivering actionable insights to stakeholders.

In summary, by leveraging machine learning techniques such as logistic regression and Flask in the backend, we have developed a robust fraud transaction detection system that enhances security, mitigates risks, and protects the interests of both banks and customers. Continuously evolving and improving upon this system through the incorporation of advanced algorithms, enhanced feature engineering, real-time monitoring, and collaborative efforts will ensure its effectiveness in combating fraud in the ever-changing landscape of banking and finance.

- **Integration of Advanced Machine Learning Algorithms:** While logistic regression provides a solid foundation, exploring and integrating more advanced machine learning algorithms
- **Real-Time Mobile Application Alerts:** Develop a mobile app that sends instant notifications to users when suspicious transactions are detected on their accounts, allowing them to take immediate action.
- **Integration with External Data Sources:** Integrating external data sources, such as public databases or third-party fraud databases, can provide additional context and information to enhance the accuracy of fraud detection.

REFERENCES

[1]. Andrew. Y. Ng, Michael. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes", Advances in neural information processing systems, vol. 2, pp. 841-848, 2002.

[2]. A. Shen, R. Tong, Y. Deng, "Application of classification models on credit card fraud detection", Service Systems and Service Management 2007 International Conference, pp. 1-4, 2007.

[3]. A. C. Bahnsen, A. Stojanovic, D. Aouada, B. Ottersten, "Cost sensitive credit card fraud detection using Bayes minimum risk", Machine Learning and Applications (ICMLA). 2013 12th International Conference, vol. 1, pp. 333-338, 2013.

[4]. B.Meena, I.S.L.Sarwani, S.V.S.S.Lakshmi, ” Web Service mining and its techniques in Web Mining” IJAEGT,Volume 2,Issue 1 , Page No.385-389.

[5]. F. N. Ogwueleka, "Data Mining Application in Credit Card Fraud Detection System", Journal of Engineering Science and Technology, vol. 6, no. 3, pp. 311-322, 2011.

[6]. G. Singh, R. Gupta, A. Rastogi, M. D. S. Chandel, A. Riyaz, "A Machine Learning Approach for Detection of Fraud based on SVM", International Journal of Scientific Engineering and Technology, vol. 1, no. 3, pp. 194198, 2012, ISSN ISSN: 2277-1581.

[7]. K. Chaudhary, B. Mallick, "Credit Card Fraud: The study of its impact and detection techniques", International Journal of Computer Science and Network (IJCSN), vol. 1, no. 4, pp. 31-35, 2012, ISSN ISSN: 2277-5420.

[8]. M. J. Islam, Q. M. J. Wu, M. Ahmadi, M. A. Sid- Ahmed, "Investigating the Performance of Naive-Bayes Classifiers and KNearestNeighbor Classifiers", IEEE International Conference on Convergence Information Technology, pp. 1541-1546, 2007.

[9]. R. Wheeler, S. Aitken, "Multiple algorithms for fraud detection" in Knowledge-Based Systems, Elsevier, vol. 13, no. 2, pp. 93-99, 2000.

[10]. S. Patil, H. Somavanshi, J. Gaikwad, A. Deshmane, R. Badgujar, "Credit Card Fraud Detection Using Decision Tree Induction Algorithm", International Journal of Computer Science and Mobile Computing (IJCSMC), vol. 4, no. 4, pp. 92-95, 2015, ISSN ISSN: 2320-088X.

[11]. S. Maes, K. Tuyls, B. Vanschoenwinkel, B. Manderick, "Credit card fraud detection using Bayesian and neural networks", Proceedings of the 1st international naiso congress on neuro fuzzy technologies, pp. 261- 270, 2002.

[12]. S. Bhattacharyya, S. Jha, K. Tharakunnel, J. C. Westland, "Data mining for credit card fraud: A comparative study", Decision Support Systems, vol. 50, no. 3, pp. 602-613, 2011. [13]. Y. Sahin, E. Duman, "Detecting credit card fraud by ANN and logistic regression", Innovations in Intelligent Systems and Applications (INISTA) 2011 International Symposium, pp. 315-319, 2011.

[14]. Selvani Deepthi Kavila, LAKSHMI S.V.S.S., RAJESH B " Automated EssayScoring using Feature Extraction Method " IJCER , volume 7, issue 4(L), Page No. 12161-12165.

[15]. S.V.S.S. Lakshmi, K.S.Deepthi,Ch.Suresh "Text Summarization basing on Font and Cue-phrase.