

PROJECT STAGE-I REPORT ON

**Credit Card Fraud Detection Using Data Science And
Machine Learning**

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT FOR THE AWARD OF THE DEGREE



**BACHELOR OF ENGINEERING
IN
INFORMATION TECHNOLOGY
BY**

Mr. Omesh Anil Satpute

Exam No: B190428521

Mr. Vishal Thange

Exam No: B190428518

Under The Guidance of

Prof. S. D. Bhopale



Sinhgad Institutes

DEPARTMENT OF INFORMATION TECHNOLOGY

STES's

SINHGAD INSTITUTE OF TECHNOLOGY,

LONAVALA

Lonavala, 410401

2023-24



**SINHGAD INSTITUTE OF TECHNOLOGY, LONAVALA
DEPARTMENT OF INFORMATION TECHNOLOGY**

CERTIFICATE

This is to certify that the Project Entitled

Alzheimer's Disease Detection using Machine Learning

Submitted by

Mr. Omesh Anil Satpute
Mr. Vishal Thange

Exam No: B190428521
Exam No: B190428518

is a bonafide work carried out by them under the supervision of Prof. S. D. Bhopale and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University for the award of the Degree of Bachelor of Engineering (Information Technology)

This project report has not been earlier submitted to any other Institute or University for the award of any degree or diploma.

Prof. S. D. Bhopale
Internal Guide
Dept. of Information Technology

Dr. R. V. Babar
H.O.D
Dept. of Information Technology

External Examiner
Date:

Dr. M. S. Gaikwad
Principal
Sinhgad Institute of Technology, Lonavala

Place:
Date:

Abstract

Credit card transactions have become common place today and so is the frauds associated with it. One of the most common modus operandi to carry out fraud is to obtain the card information illegally and use it to make online purchases. For credit card companies and merchants, it is infeasible to detect these fraudulent transactions among thousands of normal transactions. If sufficient data is collected and made available, machine learning algorithms can be applied to solve this problem. In this work, popular supervised and unsupervised machine learning algorithms have been applied to detect credit card frauds in a highly imbalanced dataset. It was found that unsupervised machine learning algorithms can handle the skewness and give best classification results. In today's economic scenario, credit card use has become extremely commonplace. These cards allow the user to make payments of large sums of money without the need to carry large sums of cash. They have revolutionized the way of making cashless payments and made making any sort of payments convenient for the buyer. This electronic form of payment is extremely useful but comes with its own set of risks. With the increasing number of users, credit card frauds are also increasing at a similar pace. The credit card information of a particular individual can be collected illegally and can be used for fraudulent transactions. Some Machine Learning Algorithms can be applied to collect data to tackle this problem. This paper presents a comparison of some established supervised learning algorithms to differentiate between genuine and fraudulent transactions.

Acknowledgment

It gives us great pleasure in presenting the preliminary project report on '**Credit Card Fraud Detection Using Data Science And Machine Learning**'.

I would like to take this opportunity to thank my internal guide **Prof. S. D. Bhopale** for giving me all the help and guidance I needed. I am really grateful to them for their kind support. Their valuable suggestions were very helpful.

I am also grateful to **Dr. R. V. Babar**, Head of Department, Information Technology, Sinhgad Institute of Technology, for his indispensable support, suggestions.

In the end our special thanks to **Dr. M. S. Gaikwad**, Principal, Sinhgad Institute of Technology, STES Campus, Lonavala, for his constant support throughout this project.

Thank you all for your encouragement!

Omesh Anil Satpute

Vishal Thange

(B.E. Information Technology)

INDEX

1	Introduction	1
1.1	Overview	2
1.1.1	Motivation	2
1.1.2	Objective	2
2	Literature Survey	3
2.1	Study Of Research Paper	4
3	Problem Statement	13
3.1	Problem Statement	14
4	Project Requirement	15
4.1	EXTERNAL INTERFACE REQUIREMENT	16
4.1.1	User Interface	16
4.1.2	Hardware Interfaces:	16
4.1.3	Software Interfaces.....	16
4.2	NONFUNCTIONAL REQUIREMENT	17
4.2.1	Performance Requirements	17
4.2.2	Safety Requirement	17
4.2.3	Software Quality Attributes	17
5	System Analysis	19
5.1	System Architecture	20
5.1.1	Module	20
5.1.2	Data Flow Diagram	21
5.2	UML DIAGRAMS	24
6	Software Information	28
7	Project Plan	33

7.1	System Implementation Plan	34
8	Results	35
8.1	Results	36
9	Conclusion	37
9.1	Conclusion	38
10	References	39

List of Figures

5.1	System Architecture	20
5.2	Data Flow (0) diagram	22
5.3	Data Flow (1) diagram	23
5.4	Data Flow (2) diagram	23
5.5	Class Diagram	24
5.6	Use case Diagram.....	25
5.7	Activity Diagram.....	26
5.8	Sequence Diagram	27
7.1	Project Timeline Chart.....	34
8.1	MR Image Input given to the system.....	36
8.2	Processed MR Image.....	36

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Credit Card Fraud can be defined as a case where a person uses someone else's credit card for personal reasons while the owner and the card issuing authorities are unaware of the fact that the card is being used. Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid objectionable behaviour, which consist of fraud, intrusion, and defaulting. Due to increase the rise of E- Commerce, there has been a tremendous use of credit cards for online shopping which led to High amount of frauds related to credit cards. In the era of digitalization, the need to identify credit card frauds is necessary. Fraud detection involves monitoring and analyzing the behavior of various users in order to estimate detect or avoid undesirable behavior. In order to identify credit card fraud detection effectively, we need to understand the various technologies, algorithms and types involved in detecting credit card frauds. Algorithm can differentiate transactions which are fraudulent or not. Find fraud, they need to passed dataset and knowledge of fraudulent transaction. They analyze the dataset and classify all transactions.

1.1.1 Motivation

Credit card fraud occurs when an unauthorized person gains access to your information and uses it to make purchases. To design and assess a new technique that effectively addresses credit card frauds. Credit card fraud primarily occurs due to the use of the internet to make payments and transfer funds. The trend for online money transfers accelerated rapidly after the Covid-19 pandemic, which led to a rise in credit card use and, thus, credit card fraud online.

1.1.2 Objective

The main objective of this thesis is to perform predictive analysis on credit card transaction dataset using machine learning techniques and detect the fraudulent transactions from the given dataset. The focus is to identify if a transaction comes under normal class or fraudulent class using predictive models. Different sampling techniques will be implemented to tackle the class imbalance problem and series of machine learning algorithms like logistic regression, random forest and xgboost will be implemented on the dataset, and the results will be reported.

CHAPTER 2

LITERATURE SURVEY

2.1 STUDY OF RESEARCH PAPER

1. Paper Name: Performance Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection

Author: Sangeeta Mittal, Shivani Tyagi

Abstract: Credit card transactions have become common place today and so is the frauds associated with it. One of the most common modus operandi to carry out fraud is to obtain the card information illegally and use it to make online purchases. For credit card companies and merchants, it is in-feasible to detect these fraudulent transactions among thousands of normal transactions. If sufficient data is collected and made available, machine learning algorithms can be applied to solve this problem. In this work, popular supervised and unsupervised machine learning algorithms have been applied to detect credit card frauds in a highly imbalanced dataset. It was found that unsupervised machine learning algorithms can handle the skewness and give best classification results.

2. Paper Name: Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison

Author: Samidha Khatri, Aishwarya Arora, Arun Prakash Agrawal

Abstract: — In today's economic scenario, credit card use has become extremely commonplace. These cards allow the user to make payments of large sums of money without the need to carry large sums of cash. They have revolutionized the way of making cashless payments and made making any sort of payments convenient for the buyer. This electronic form of payment is extremely useful but comes with its own set of risks. With the increasing number of users, credit card frauds are also increasing at a similar pace. The credit card information of a particular individual can be collected illegally and can be used for fraudulent transactions. Some Machine Learning Algorithms can be applied to collect data to tackle this problem. This paper presents a comparison of some established supervised learning algorithms to differentiate between genuine and fraudulent transactions.

3. Paper Name: Comparative Evaluation of Credit Card Fraud Detection Using Machine Learning Techniques

Author: Olawale Adepoju, Julius Wosowei, Shiwani lawte.

Abstract: Credit card fraud is a serious and growing problem with the increase in e-commerce and online transactions in this modern era. With this identity theft and loss of money, such mischievous practices can affect millions of people around the world. Criminal activity is a rising threat to the financial sector with-reaching implications. Information extraction seemed to have assumed a basic job in recognition of online payment fraud, fraud detection efficiency in credit card purchases is significantly affected by the data set measuring strategy, the choice of variable and the detection techniques used.

This publication inspects execution of, Support Vector Machine, Naive Bayes, Logistic Regression and K-Nearest Neighbor on exceptionally distorted data on credit card fraud. The execution of these techniques is assessed dependent on accuracy, sensitivity, precision, specificity. The outcomes show an ideal accuracy for logistic regression, Naive Bayes, knearest neighbor and Support vector machine classifiers are 99.07%, 95.98%, 96.91%, and 97.53% respectively. The relative outcomes demonstrate that logistic regression performs superior to other algorithms.

4.Paper Name: Credit Card Fraud Detection System

Author: V. Filippov, L. Mukhanov, B. Shchukin.

Abstract: The use of credit cards is prevalent in modern day society. But it is obvious that the number of credit card fraud cases is constantly increasing in spite of the chip cards worldwide integration and existing protection systems. This is why the problem of fraud detection is very important now. In this paper the general description of the developed fraud detection system and comparisons between models based on using of artificial intelligence are given. In the last section of this paper the results of evaluative testing and corresponding conclusions are considered.

5.Paper Name: Credit Card Fraud Detection using Machine Learning and Deep Learning Techniques

Author: Mohammed Azhan, Shazli Meraj.

Abstract: In general, fraudulent activities are always intended to cause financial detriment to the second party. With the aggrandizement of digital money in various countries, the fraudulent activities will be even more increased. Credit card companies

and Banks lose billions to such fraudulent activities every year, where it accounts to a huge part of their revenue and affects the jobs of various employees. The proposed research work discusses more about the different fraudulent activities associated with credit cards. While all of them cannot be dealt simultaneously, this research work discusses how Machine Learning and Neural Networks can be used to determine the potential fraudsters by referring to their previous mistakes and details of previous fraudsters. Machine Learning algorithms such as Multinomial Naive Bayes, Random Forest Regression, Logistic Regression, Support Vector Machine and a basic Neural Network are also used.

6. Paper Name: Research on Credit Card Fraud Detection Model Based on Distance Sum

Author: Wen-Fang YU, Na Wang, Yasmin.

Abstract: Along with increasing credit cards and growing trade volume in China, credit card fraud rises sharply. How to enhance the detection and prevention of credit card fraud becomes the focus of risk control of banks. This paper proposes a credit card fraud detection model using outlier detection based on distance sum according to the infrequency and unconventionality of fraud in credit card transaction data, applying outlier mining into credit card fraud detection. Experiments show that this model is feasible and accurate in detecting credit card fraud.

7. Paper Name: Credit Card Fraud Detection - Machine Learning methods

Author: Dejan Varmedja, Mirjana Karanovic, Srdjan Sladojevic, Marko Arsenovic, Andras Anderla

Abstract: Credit card fraud refers to the physical loss of credit card or loss of sensitive credit card information. Many machine learning algorithms can be used for detection. This research shows several algorithms that can be used for classifying transactions as fraud or genuine one. Credit Card Fraud Detection dataset was used in the research. Because the dataset was highly imbalanced, SMOTE technique was used for oversampling.

CHAPTER 3

PROBLEM STATEMENT

3.1 PROBLEM STATEMENT

The Credit Card Fraud Detection Problem includes modeling past credit card transactions with the knowledge of the ones that turned out to be a fraud. This model is then used to identify whether a new transaction is fraudulent or not. Credit card frauds are increasing heavily because of fraud financial loss is increasing drastically. Every year due to fraud Billions of amounts lost. To analyze the fraud there is lack of research. Many machine learning algorithms are implemented to detect real world credit card fraud. KNN and K means clustering etc. are applied.

CHAPTER 4

PROJECT REQUIREMENT

4.1 EXTERNAL INTERFACE REQUIREMENT

4.1.1 User Interface

Web based application detection

4.1.2 Hardware Interfaces:

RAM: 8 GB

As we are using Machine Learning Algorithm and Various High Level Libraries
Laptop

RAM minimum required is 8 GB.

Hard Disk: 40 GB

Data Set of CT Scan images is to be used hence minimum 40 GB Hard Disk memory
is required.

Processor: Intel i5 Processor

Pycharm IDE that Integrated Development Environment is to be used and data
loading should be fast hence Fast Processor is required

IDE: Pycharm

Best Integrated Development Environment as it gives possible suggestions at the
time of typing code snippets that makes typing feasible and fast.

Coding Language: Python Version 3.5

Highly specified Programming Language for Machine Learning because of
availability of High Performance Libraries.

Operating System: Windows 10

Latest Operating System that supports all type of installation and development
Environment

4.1.3 Software Interfaces

Operating System: Windows 10

IDE: Pycharm, Spyder
Programming Language: Python

4.2 NOFUNCTIONAL REQUIREMENT

4.2.1 Performance Requirements

The performance of the functions and every module must be well. The overall performance of the software will enable the users to work efficiently. Performance of encryption of data should be fast. Performance of the providing virtual environment should be fast

- The application is designed in modules where errors can be detected and indexed easily. This makes it easier to install and update new functionality if required.

4.2.2 Safety Requirement

The application is designed in modules where errors can be detected and fixed easily. This makes it easier to install and update new functionality if required.

4.2.3 Software Quality Attributes

Our software has many quality attributes that are given below:

Adaptability: This software is adaptable by all users.

Availability: This software is freely available to all users. The availability of the software is easy for everyone.

CHAPTER 5

SYSTEM ANALYSIS

5.1 SYSTEM ARCHITECTURE

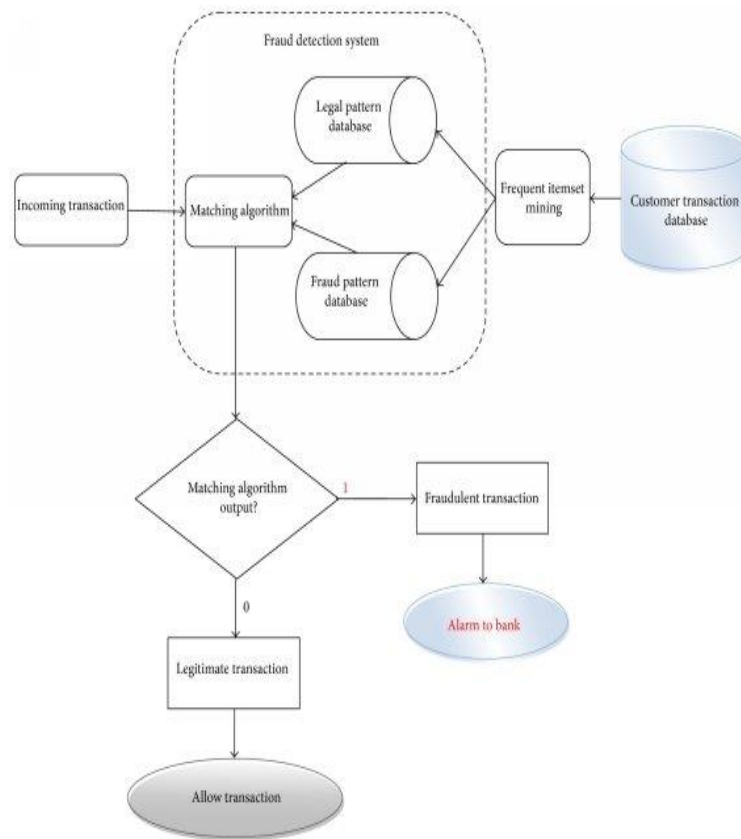


Figure 5.1: System Architecture

5.1.1 Module

- In this module, the user has to upload the MRI Images of the Brain into the system.
- The system then trains the dataset with the image fed to it.
- In the next step, the image is processed and converted to a greyscale image.
- Further, feature extraction takes place on the image where the detection of parameters from the fetal image is done. This is used for testing purposes

5.1.2 Data Flow Diagram

In Data Flow Diagram, we Show that flow of data in our system in DFD0 we show that base DFD in which rectangle present input as well as output and circle show our system, In DFD1 we show actual input and actual output of system input of our system is text or image and output is rumor detected likewise in DFD 2 we present operation of user as well as admin.

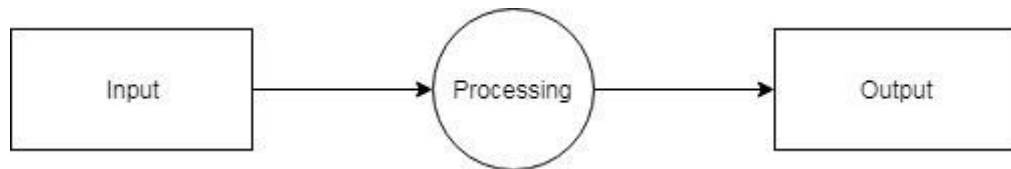


Figure 5.2: Data Flow (0) diagram

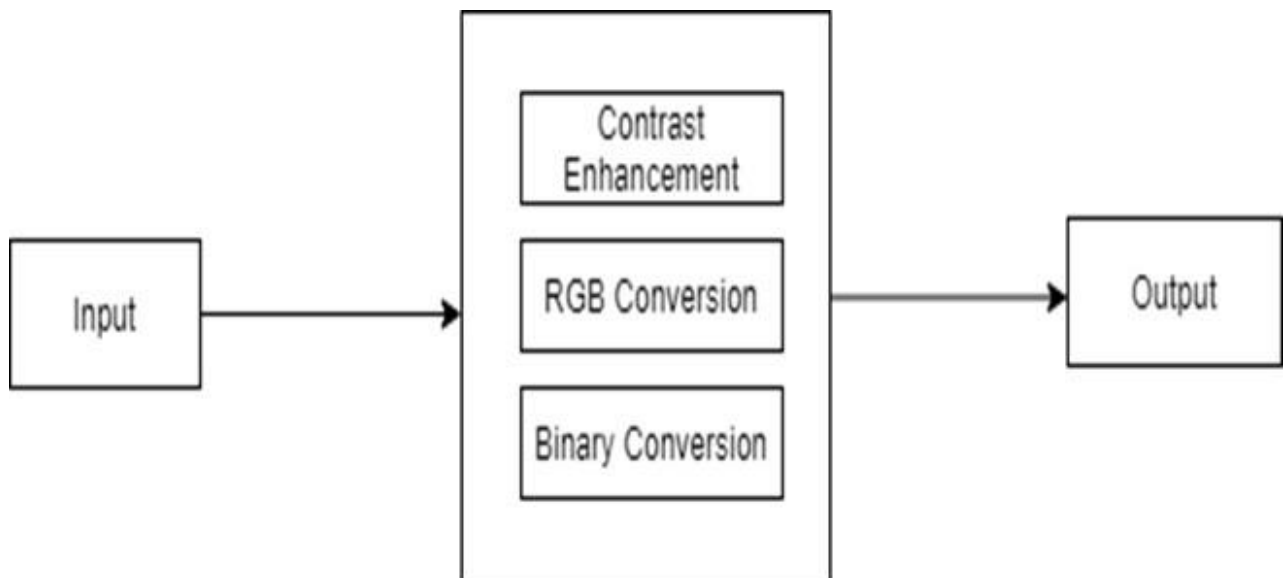


Figure 5.3: Data Flow (1) diagram

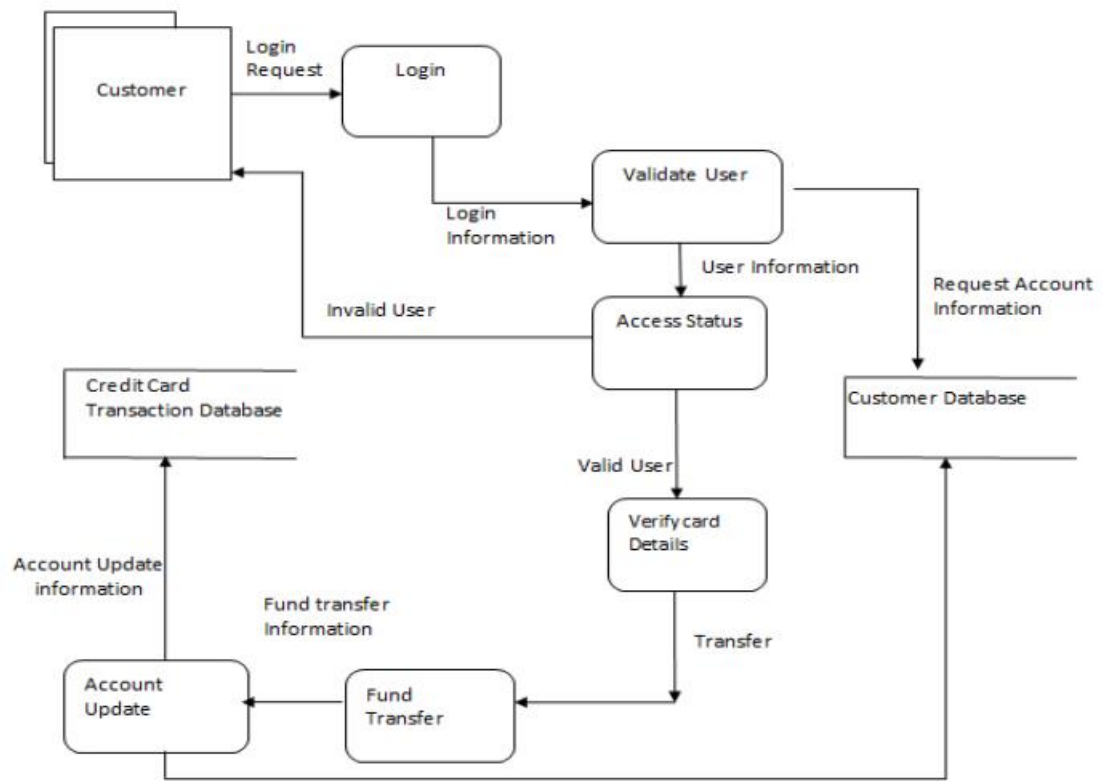


Fig 3:- DFD diagram [25]

Figure 5.4: Data Flow (2) diagram

5.2 UML DIAGRAMS

Unified Modeling Language is a standard language for writing software blueprints. The UML may be used to visualize, specify, construct and document the artifacts of a software intensive system. UML is process independent, although optimally it should be used in process that is use case driven, architecture-centric, Iterative, and incremental. The Number of UML Diagram is available.

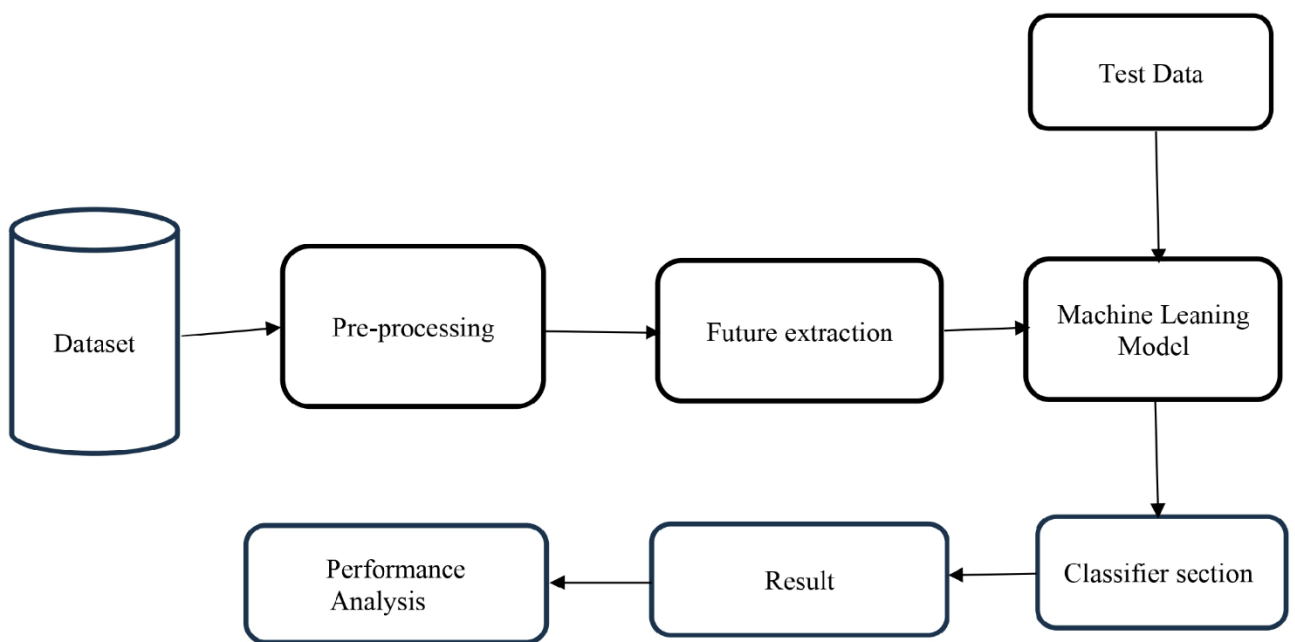


Fig. 1 System Architecture

Figure 5.5: Class Diagram

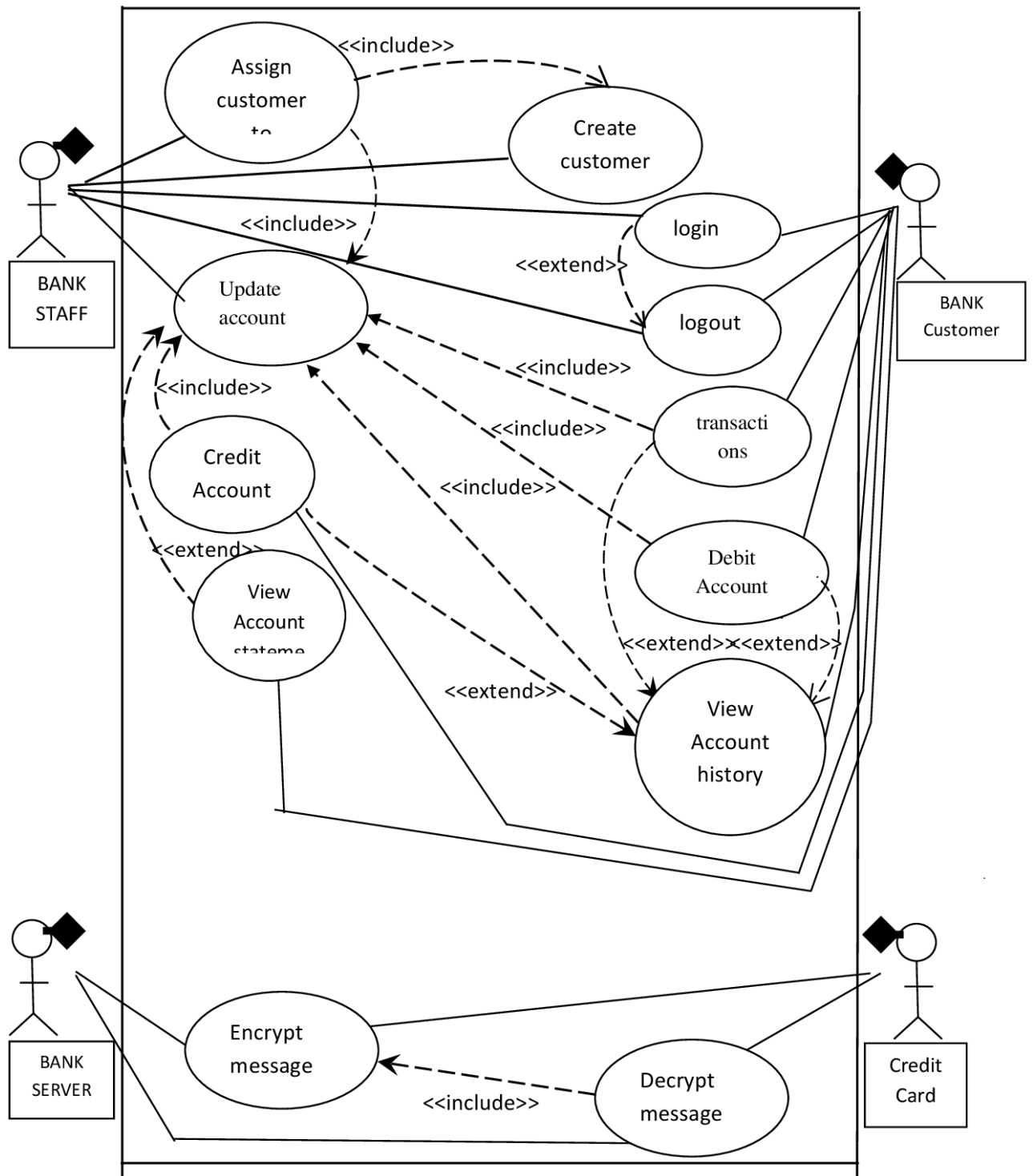


Figure 5.6: Use case Diagram

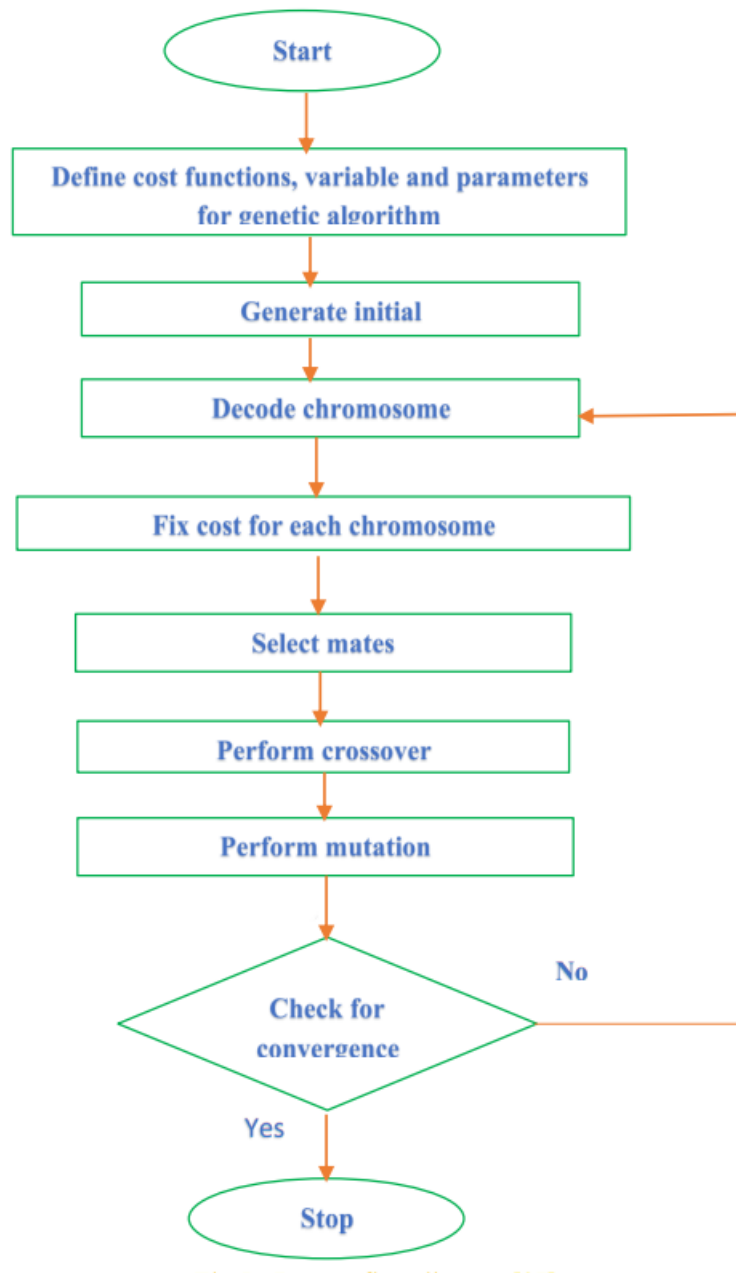


Figure 5.7: Activity Diagram

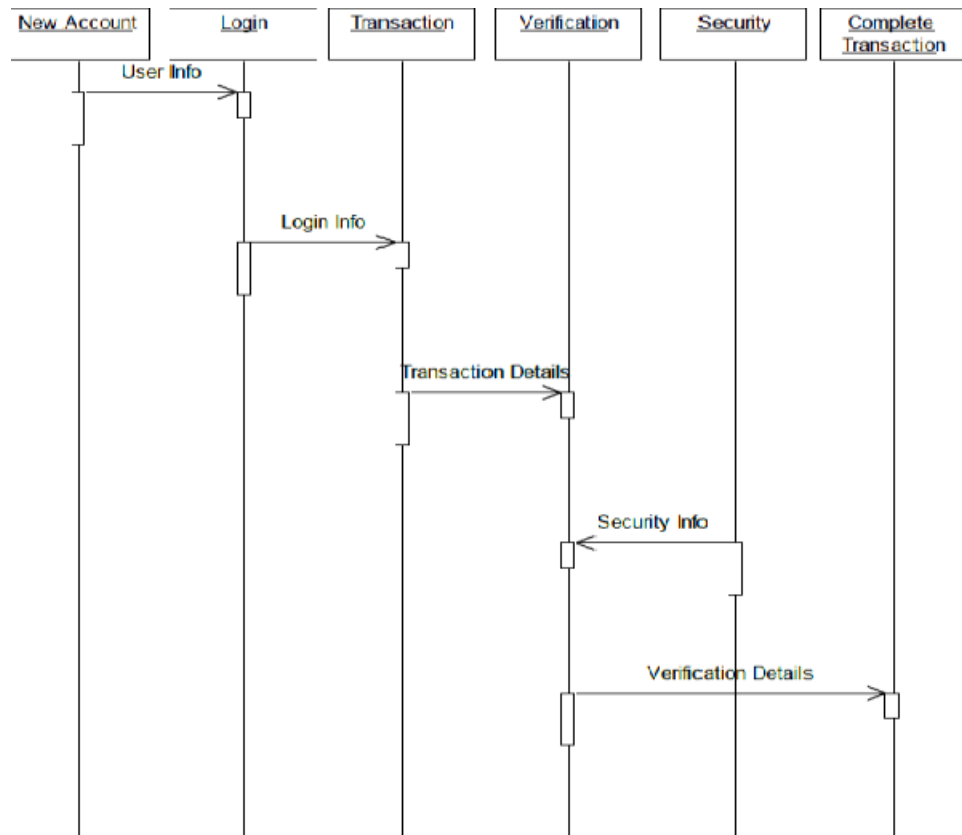


Figure 5.8: Sequence Diagram

CHAPTER 6

SOFTWARE INFORMATION

Python:

Python is an interpreted, high-level and general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented, and functional programming. Python is often described as a “batteries included” language due to its comprehensive standard library.

Anaconda: Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. The distribution includes data science packages suitable for Windows, Linux, and macOS. It is developed and maintained by Anaconda, Inc., which was founded by Peter Wang and Travis Oliphant in 2012. As an Anaconda, Inc. product, it is also known as Anaconda Distribution or Anaconda Individual Edition, while other products from the company are Anaconda Team Edition and Anaconda Enterprise Edition, both of which are not free.

CHAPTER 7

PROJECT PLAN

In this chapter we are going to have an overview about how much time does it took to complete each task like- Preliminary Survey Introduction and Problem Statement, Literature Survey, Project Statement, Software Requirement and Specification, System Design, Partial Report Submission, Architecture Design, Implementation, Deployment, Testing, Paper Publish, Report Submission and etcetera. This chapter also gives focus on stakeholder list which gives information about project type, customer of the proposed system, user and project member who developed the system.

7.1 SYSTEM IMPLEMENTATION PLAN

The System Implementation plan table, shows the overall schedule of tasks compilation and time duration required for each task.

Sr. No.	Name/Title	Start Date	End Date
1	Preliminary Survey	18 July 2023	22 July 2023
2	Introduction and Problem Statement	23 July 2023	05 August 2023
3	Literature Survey	06 August 2023	29 August 2023
4	Project Statement	30 August 2023	02 September 2023
5	Software Requirement And Specification	03 September 2023	19 September 2023
6	System Design	21 September 2023	05 October 2023
7	Partial Report Submission		
8	Architecture Design		
9	Implementation		
10	Deployment		
11	Testing		
12	Paper Publish		
13	Report Submission		

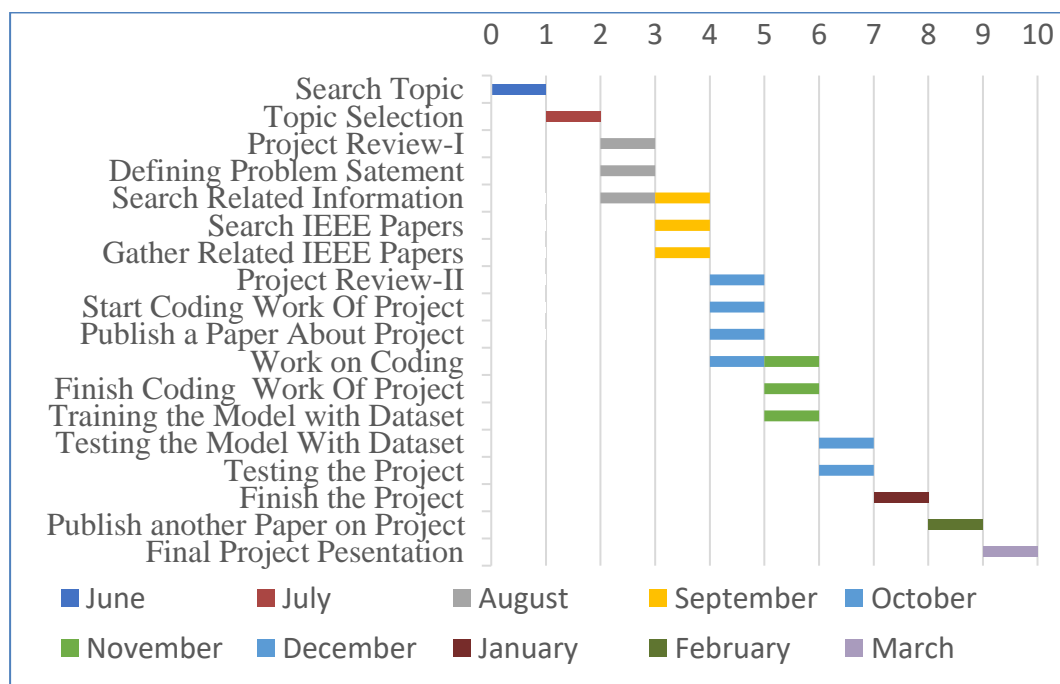


Figure 7.1: Project Timeline Chart

CHAPTER 8

RESULTS

8.1 RESULTS

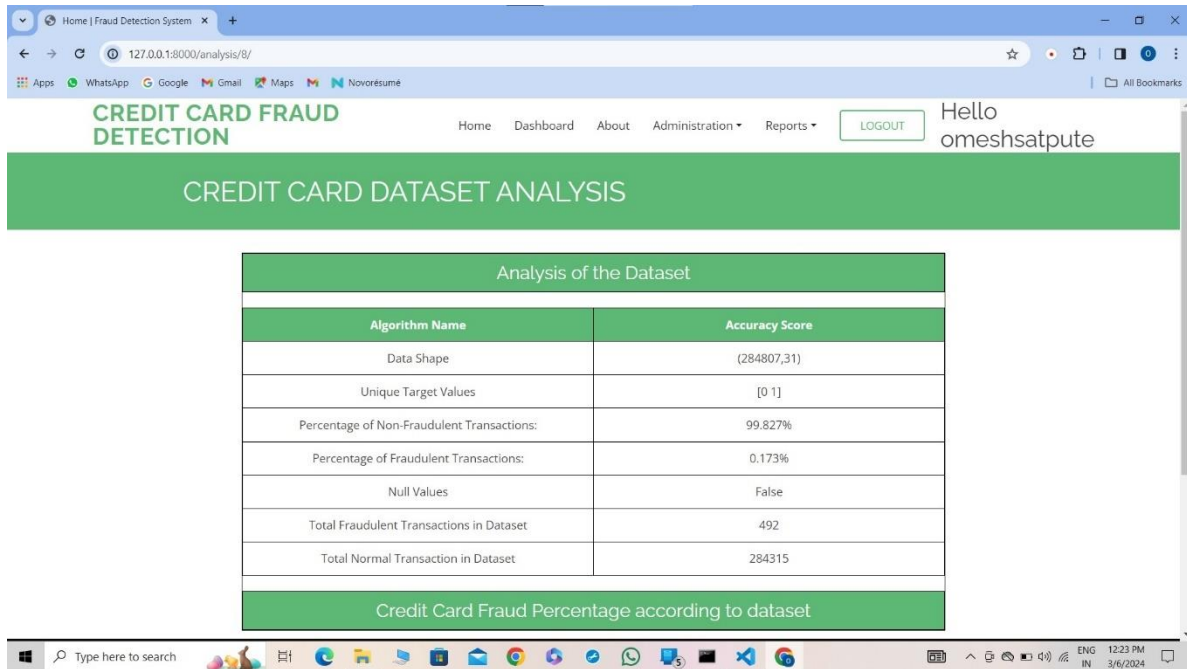


Figure 8.1: Single CSV fraud analysis

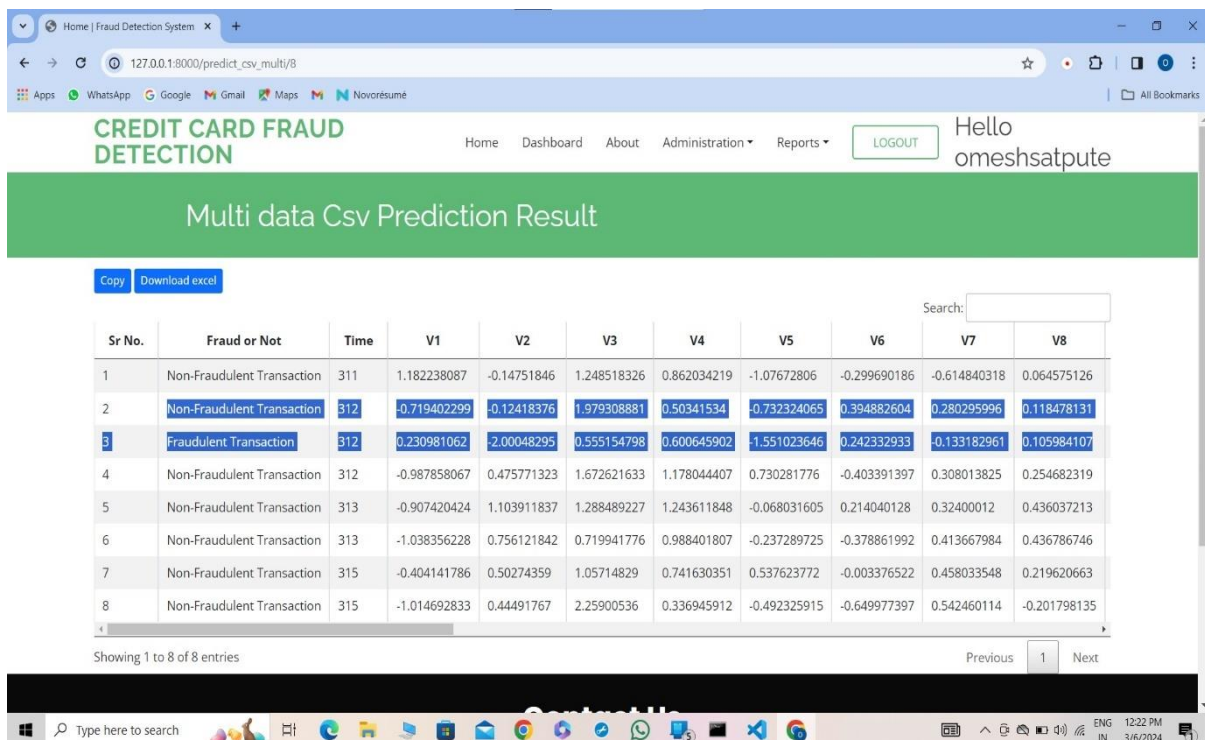


Figure 8.2: Multiple CSV fraud analysis

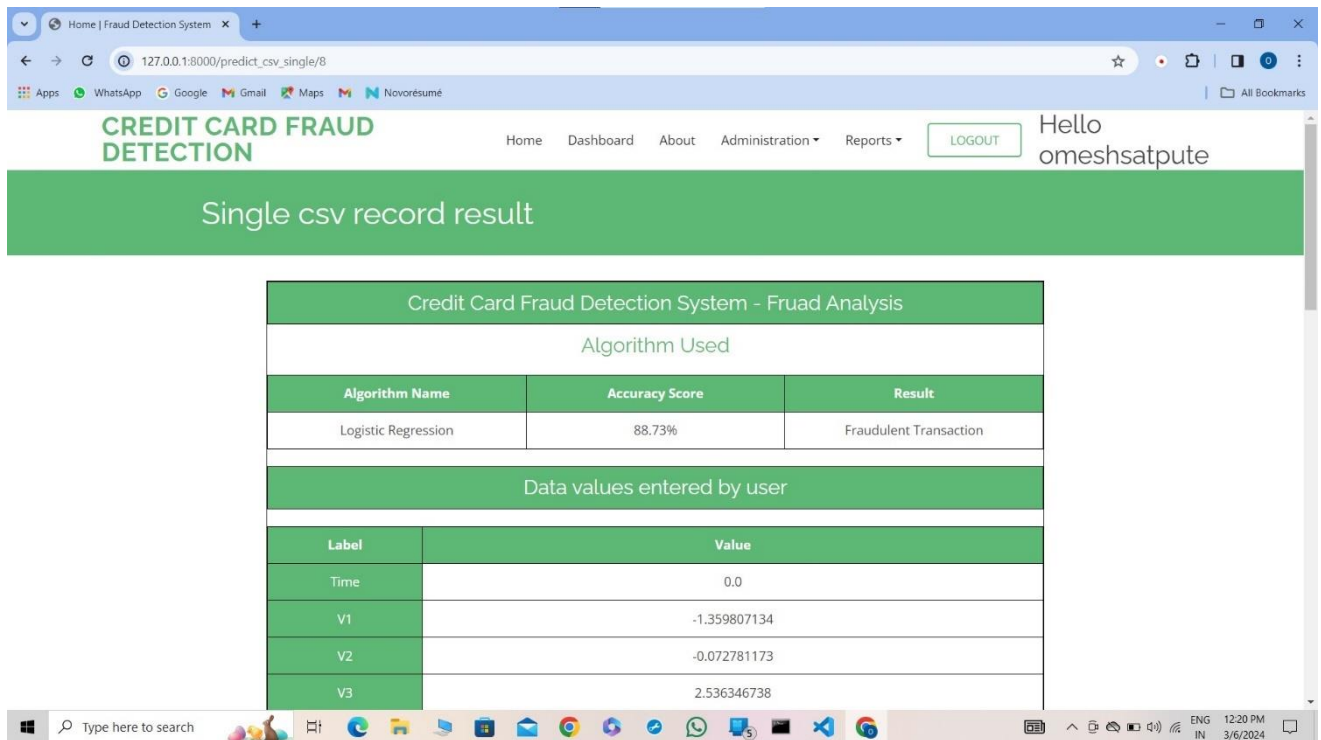


Figure 8.3: Final Analysis

CHAPTER 9

CONCLUSION

9.1 CONCLUSION

In this work, an effort has been made to study the 3D brain MR image slices for AD diagnosis. All the three different views of slices (Axial, Sagittal, and Coronal) of gray matter and the white matter has been used for this study. Based on several observations slice number 51 has been chosen and used for further analysis. The first-order statistical feature has been extracted from each slice.

REFERENCES

- [1] A. Association et al., “2017 Alzheimer’s disease facts and figures,” *Alzheimer’s Dementia*, vol. 13, no. 4, pp. 325–373, 2017.
- [2] S. Li, O. Okonkwo, M. Albert, and M.-C. Wang, “Variation in variables that predict progression from MCI to AD dementia over duration of follow-up,” *American Journal of Alzheimer’s Disease (Columbia, Mo.)*, vol. 2, no. 1, pp. 12–28, 2013.
- [3] R. Roberts and D. S. Knopman, “Classification and epidemiology of MCI,” *Clinics in Geriatric Medicine*, vol. 29, no. 4, pp. 753–772, 2013.
- [4] N. Fox, R. Black, S. Gilman, M. Rossor, S. Griffith, L. Jenkins, M. Koller et al., “Effects of A immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease,” *Neurology*, vol. 64, no. 9, pp. 1563–1572, 2005.
- [5] G. B. Frisoni, N. C. Fox, C. R. Jack Jr, P. Scheltens, and P. M. Thompson, “The clinical use of structural MRI in Alzheimer disease,” *Nature Reviews Neurology*, vol. 6, no. 2, pp. 67–77, 2010.
- [6] C. R. Jack, R. C. Petersen, Y. C. Xu, P. C. O’Brien, G. E. Smith, R. J. Ivnik, B. F. Boeve, S. C. Waring, E. G. Tangalos, and E. Kokmen, “Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment,” *Neurology*, vol. 52, no. 7, pp. 1397–1397, 1999.
- [7] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehericy, M.-O. Habert, M. Chupin, H. Benali, O. Colliot, A. D. N. Initiative et al., “Automatic classification of patients with Alzheimer’s disease from structural MRI: a comparison of ten methods using the ADNI database,” *Neuroimage*, vol. 56, no. 2, pp. 766–781, 2011.
- [8] F. Falahati, E. Westman, and A. Simmons, “Multivariate data analysis and machine learning in Alzheimer’s disease with a focus on structural magnetic resonance imaging,” *Journal of Alzheimer’s Disease*, vol. 41, no. 3, pp. 685–708, 2014.
- [9] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, J. Tohka, A. D. N. Initiative et al., “Machine learning framework for early MRI-based Alzheimer’s conversion prediction in mci subjects,” *Neuroimage*, vol. 104, pp. 398–412, 2015.
- [10] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, “Early diagnosis of Alzheimer’s disease with deep learning,” in *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*. IEEE, 2014, pp. 1015–1018.
- [11] Dubois B, Padovani A, Scheltens P, Rossi A, Dell’Agnello G. Timely diagnosis

