

Credit Card Fraud Detection

A Project Report

Submitted in partial fulfilment of the
Requirements for the award of the Degree of

MASTER OF SCIENCE (INFORMATION TECHNOLOGY)

By

Pankaj Jayprakash Yadav

Seat Number-211



Under the esteemed guidance of

Prof. Rujuta Sawant

DEPARTMENT OF INFORMATION TECHNOLOGY

Ramanand Arya D.A.V College

(Autonomous)

OF

ARTS, SCIENCE & COMMERCE

BHANDUP, MUMBAI – 400 042 MAHARASHTRA

AY 2021 – 2022

RAMANAND ARYA D.A.V
COLLEGE
OF ARTS, SCIENCE & COMMERCE
(AUTONOMOUS)
BHANDUP, MUMBAI, MAHARASHTRA – 400 042
DEPARTMENT OF INFORMATION TECHNOLOGY

CERTIFICATE

This is to certify that the project entitled, "**CREDIT CARD FRUAD DETECTION Application** ", is bonafied work of **PANKAJ JAYPRAKASH YADAV** bearing Seat No: **211** submitted in partial fulfilment of the requirements for the award of degree of **MASTERS OF SCIENCE** in **INFORMATION TECHNOLOGY** from University of Mumbai

Internal Guide

Coordinator

External Examiner

Date:

College Seal

ACKNOWLEDGEMENT

I express sincere gratitude of Professor Rujuta Sawant on my project guide for making available, the facilities required for the completion of the project.

I greatly indebted to my project guide for his valuable and timely guidance along with his wholehearted cooperation, which has played a key role in the successful completion of the project.

A vote of thanks to all staff members of Ramanand Arya D.A.V College, who have directly or indirectly, contributed towards the completion of the project.

I would also like to thank the non-teaching staff members of the college for allowing me to use the laboratory facilities for as much time as required.

DECLARATION

I hereby declare that the project entitled, “**Credit Card Fraud Detection**” done at **Ramanand Arya D.A.V College**, has not been any case duplicated to submit to any other university for the award of any degree. To the best of my knowledge other than me, no one has submitted to any other university.

The project is done in partial fulfilment of the requirements for the award of degree of **MASTER OF SCIENCE (INFORMATION TECHNOLOGY)** to be submitted as final semester project as part of our curriculum.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Introduction

1.2 Objective and scope

CHAPTER 2

2.1 Identification of need

2.2 Preliminary Investigation

2.3 Feasibility Study

2.3.1 Economic Study

2.3.2 Legal Feasibility

2.3.3 Operational Feasibility

2.4 Project Planning

2.5 Project Scheduling

2.6 Software Requirement Specification

2.6.1 Overall Description

2.6.2 User class and Characteristics

2.6.3 Operating Environment

2.6.4 Design Constraints

2.6.5 Software Engineering Paradigm

CHAPTER 3

3.1 Algorithm Details

3.2 Review of Literature

CHAPTER 4: SYSTEM DESIGN

4.1 Block Diagram

4.1.1 Class Diagram

4.1.2 Use Case Diagram

4.1.3 System Architecture Diagram

4.1.4 Control Flow Diagram

4.2 Data Dictionary

Credit Card Fraud Detection

Abstract

Financial fraud is a growing problem with long term consequences in the financial industry and while many techniques have been discovered to solve this problem faced by various companies, data mining has been successfully applied to finance databases to automate analysis of huge volumes of complex data. Data mining has also played a salient role in the detection of credit card fraud in online transactions.

Fraud detection in credit card is a data mining problem. It becomes challenging due to two major reasons – first, the profiles of normal and fraudulent behavior change frequently and second, the credit card fraud data sets are highly skewed. This project investigates and checks the performance of SVC (Support Vector Classifier) on highly skewed credit card fraud data.

Chapter 1

1.Introduction

Financial fraud is a growing concern with far reaching consequences in the government, corporate organizations, finance industry. In today's world, high dependency on internet technology has enjoyed increased credit card transactions, but, credit card fraud has also accelerated as online and offline transaction. As credit card transactions become a widespread mode of payment, focus has been given to recent computational methodologies to handle the credit card fraud problem. There are many fraud detection solutions and software which prevent frauds in businesses such as credit card, retail, e-commerce, insurance, and industries.

Data mining technique is one notable and popular methods used in solving credit fraud detection problem. It is impossible to be sheer certain about the true intention and rightfulness behind an application or transaction. In reality, to seek out possible evidences of fraud from the available data, using mathematical algorithms is the best effective option. Fraud detection in credit card is truly the process of identifying those transactions that are fraudulent into two classes of legit class and fraud class transactions. Several techniques are designed and implemented to solve credit card fraud detection such SVC.

Credit card transaction datasets are rarely available, highly imbalanced and skewed. Optimal feature (variables) selection for the models, suitable metric is most important part of data mining to evaluate performance of techniques on skewed credit card fraud data. A number of challenges are associated with credit card detection, namely fraudulent behavior profile is dynamic, that is fraudulent transactions tend to look like legitimate ones. Credit card fraud detection performance is greatly affected by type of sampling approach used, selection of variables and detection technique used. In the end, conclusions about results of classifier evaluative testing are made and collated.

1.1 Objective and Scope of the Project

- Credit card fraud detection is a very popular but also a difficult problem to solve.
- Firstly, due to issue of having only a limited amount of data, credit card makes it challenging to match a pattern for dataset.
- Secondly, there can be many entries in dataset with truncations of fraudsters which also will fit a pattern of legitimate behavior. Also, the problem has many constraints.
- Firstly, data sets are not easily accessible for public and the results of researches are often hidden and censored, making the results inaccessible and due to this it is challenging to benchmarking for the models built.
- Secondly, the improvement of methods is more difficult by the fact that the security concern imposes a limitation to exchange of ideas and methods in fraud detection, and especially in credit card fraud detection.
- Lastly, the data sets are continuously evolving and changing making the profiles of normal and fraudulent behaviors always different, that is, a legit transaction in the past may be a fraud in present or vice versa.
- Create a model based on the patterns derived in the previous step.
- We would need to predict whether a Credit Card transaction would be fraud or not by using the information provided by the customers.
- This is possible because of the previous record of various customers to where the card details were stored. On the basis of that we trained our data using machine learning algorithm to predict the result.

Chapter 2

System Analysis

Preliminary Investigation:-

Difficulties of Credit Card Fraud Detection

Fraud detection systems are prone to several difficulties and challenges enumerated below. An effective fraud detection technique should have abilities to address these difficulties in order to achieve best performance.

- **Imbalanced data:** The credit card fraud detection data has imbalanced nature. It means that very small percentages of all credit card transactions are fraudulent. This causes the detection of fraud transactions very difficult and imprecise.
- **Different misclassification importance:** In fraud detection task, different misclassification errors have different importance. Misclassification of a normal transaction as fraud is not as harmful as detecting a fraud transaction as normal. Because in the first case the mistake in classification will be identified in further investigations.
- **Overlapping data:** many transactions may be considered fraudulent, while actually they are normal (false positive) and reversely, a fraudulent transaction may also seem to be legitimate(false negative). Hence obtaining low rate of false positive and false negative is a key challenge of fraud detection systems.
- **Lack of adaptability:** classification algorithms are usually faced with the problem of detecting new types of normal or fraudulent patterns. The supervised and unsupervised fraud detection systems are inefficient in detecting new patterns of normal and fraud behaviors, respectively.
- **Fraud detection cost:** The system should take into account both the cost of fraudulent behavior that is detected and the cost of preventing it. For example, no revenue is obtained by stopping a fraudulent transaction of a few dollars.

2.2 Feasibility Study:

A quality information system costs money, time and requires a consistent step by step well thought out approach in its development. The correct application of information technology is a critical success factor for the optimal functioning and competitive position of an organization, therefore, the integration of social needs, the resources and the possibilities offered by technology in an organization is required. System design focuses on the conceptual and logical design of information system looking at the function and data models. The physical design on the other hand looks at design in terms of hard and software, databases and interfaces. System realization, which has three components and these are implementation, installation and operation. The implementation looks at building and testing the system and integrating the separately developed systems among other objectives.

2.3.1 Technical Feasibility:-

The proposed system is technically feasible as it does not incur any changes in the configuration of the existing computer system for being compatible with the system developed. It does not require additional paid professional or cost of training for working on the system as no professional knowledge is required to work on this system. Hence the solution is technically feasible. Technical evolution must also access whether the organization has expertise to use it.

2.3.2 Economical Feasibility

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus, the developed system is well within the budget and this was achieved, because most of the technologies used are freely available.

2.3.3 Legal Feasibility Study

Credit card fraud detection uses freely available development tools, and provide the system as an open-source system. The machine learning libraries that are used in this system are free open-source libraries

2.3.4 Operational Feasibility Study

The operational scope of the system is verified under operational feasibility. The proposed system will have enough operational reach, which ensures the security of the information. Hence, operational feasibility of the proposed system is found to be high. This project involves the general user-friendly windows environment. Graphical user Interface, has been designed to give the user a nice look and feel. Operational feasibility ensures that the project is successfully implemented. The project can be used by the users with basic internet knowledge. Hence, we conclude that this project is operationally feasible.

2.3 Project Planning

The project planning began in the month of December with the requirement gathering and scope definition already done in the month of November.

It was estimated to be a 6-month long plan with only one person responsible for development, testing etc.

The software used are open source and readily available, hence there was no dependency on buying a subscription plan.

Some additional features had to be removed from the final version presented due to time constraints. These additional features were the provision to send a message to the customer whose credit and debit card transaction has been Tampered/detected as fraud transaction so that the system could be used in Banking Sector .

Had this feature been included the scope of the project would have expanded to incorporate the needs of banking industry as well as catering to corporate needs.

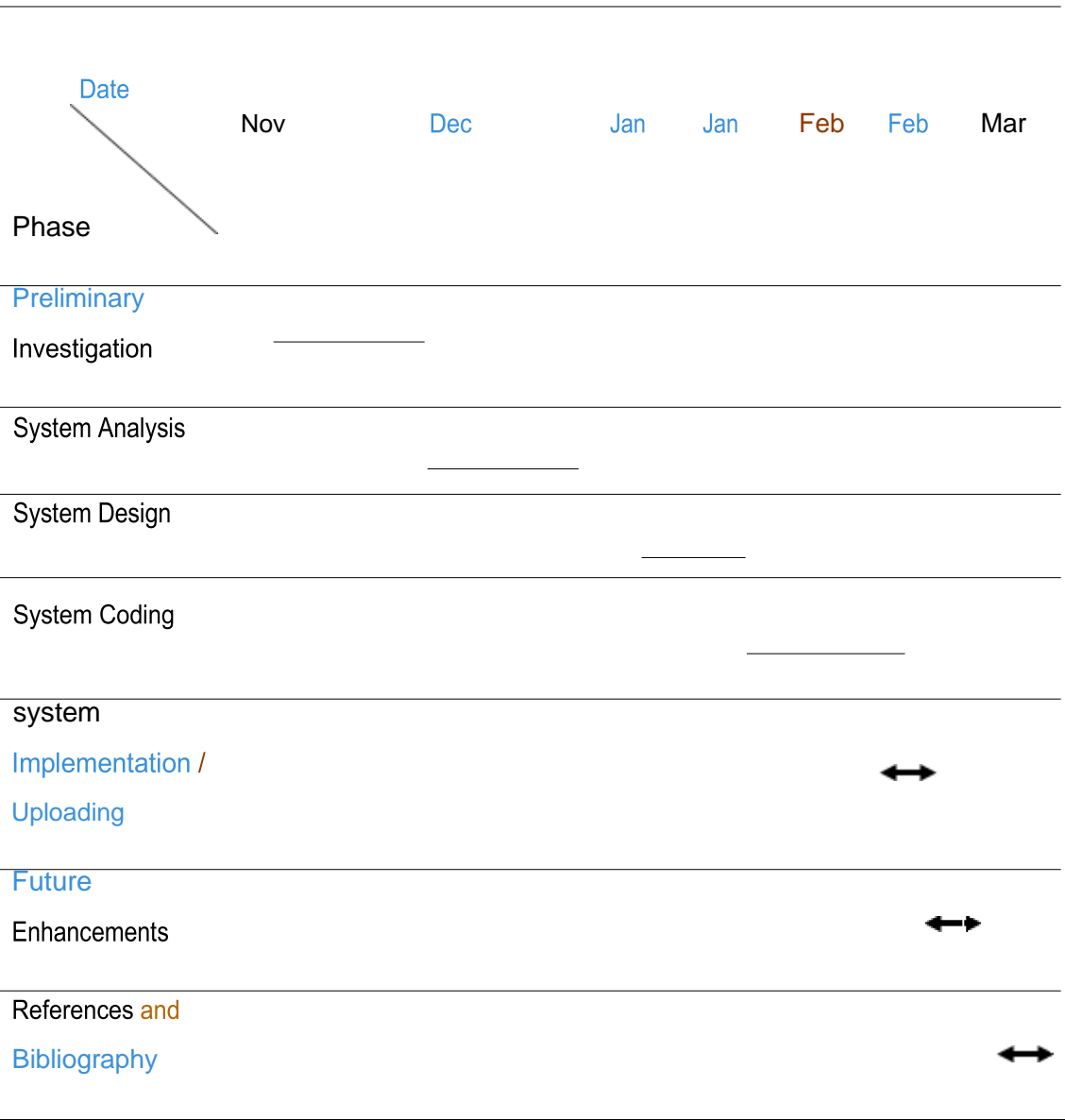
2.4 Project Scheduling

Planning and scheduling are an important part of project development. Planning makes the work more easily to handle and scheduling makes how the project will go, how much time it will take to complete the project. Planning and Scheduling is just an estimation which will play important role in project development

First of all, first we have to select the topic of the project. To developing the Credit Card Detection web application first I have to finding the resources and then what different types of features I will want to add in Credit Card Detection web application and also a GUI which can see to nice and people with easily interact with it make them to understand the application more easily.

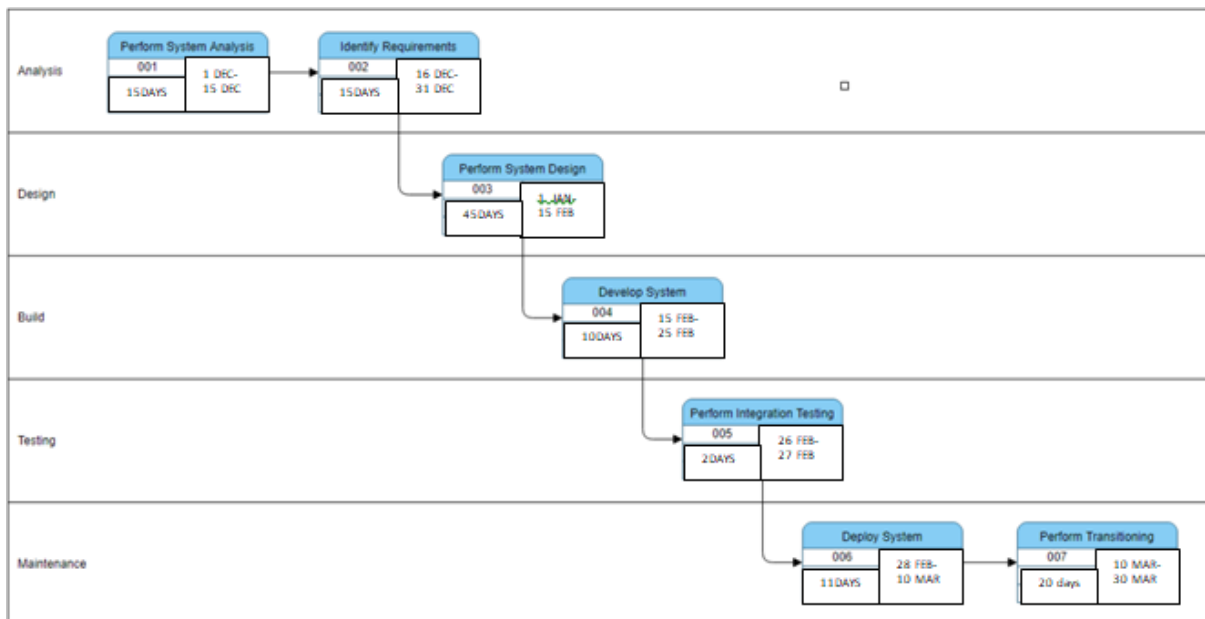
2.5.1 Gantt Chart

This chart is a mathematical chart that is used to show the progress of specific tasks in a graphical manner, the graph used is the horizontal bar graph that shows different segments of different parts of the project. The numerical values entered are in percentage format for the percentage of task completed along with the schedule of start of task and its duration and actual start. The Gantt chart below shows the tasks and their duration and percentage completed accordingly for example, tasks like project topic exploration and project topic selection are 100% completed and that too within the duration allotted to them similarly, some tasks like UML diagrams were completed but went ahead of actual duration allotted. All the tasks mentioned in the chart below are a particular stage of project development.



2.5.2 Pert chart

A PERT chart is a statistical tool used in project management that analyses and tracks activities and milestones by mapping them out on a timeline. A PERT chart is event-oriented: it's comprised of numbered nodes, directional arrows. Activities are represented by these arrows, and nodes are the milestones associated with them. The raw data produced by PERT can be imported into project scheduling software, which helps you manage the whole project.



2.5 Software Requirement Specification

- Python:- It is an open-source free tool used for coding the machine learning model.
- Django Framework:- This Framework has been used to build a web model for machine learning model.
- Bootstrap:- It is a free and open-source CSS framework which used to design we application.
- Anaconda Framework:- This the framework where we can use install and use many tools. For our machine learning model, we have used VS Code for designing and developing machine learning model.

2.6.1 Overall Description

This project is used for credit card fraud detection web application and to detect whether the transaction is fraud or not. This is going to decide by machine learning algorithm. Along with that we are uploading respective document. So that information of customer credit card transactions can verify by the admin department. This is overall concept of project. With help of machine learning algorithm, we achieve our goal.

2.6.2 User class and characteristics

- In this system the admin can verify the credit card fraud transactions.
- The Admin can upload the required file and then can submit to know the prediction of the algorithm.

2.6.3 Operating Environment

- Front End

- VS Code
- Bootstrap
- Django Framework
- Anaconda Framework

- Back End

We have used SQLite database for storing the username and password of the user/admin.

- Operating System

- Windows Operating system and Linux Operating System only

- Tools Used

- Visual Studio code
- Anaconda Framework
- Python
- Jupyter Notebook

2.6.4 Design Constraints

➤ Software Constraints

- Using this system is fairly simple and intuitive.
- A user familiar with basic computer skills should be able to understand all functionalities provided by the system.

➤ **Hardware Constraints**

- The solution is functional on a Windows Operating system since the UI is made using Django Framework and Bootstrap .
- It cannot be used on a mobile device or on Mac OS

2.6 Software Engineering Paradigm

The paradigm used here is a of waterfall and agile model, thus making this a hybrid model of engineering. All the requirements were gathered prior to the planning phase could begin. Similarly, the development began based on the plan made. However, there were certain features such as the feature to sending the email to the user if there is any fraud transaction. Had this functionality been implemented, the scope of the project would have expanded to incorporate the banking industry as well. Unfortunately, owing to time constraints the feature could not be added.

2.7 Cost estimation of project

This estimation is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus, the developed system is well within the budget and this was achieved, because most of the technologies used are freely available.

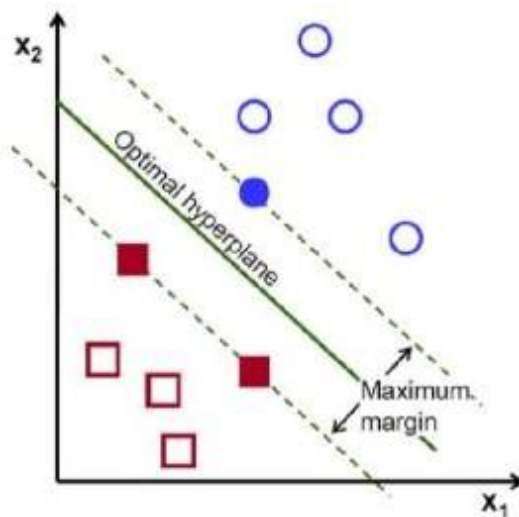
Chapter 3

3.1 Algorithm details

Ability of system to automatically learn and improve from experience without being explicitly programmed is called machine learning and it focuses on the development of computer programs that can access data and use it to learn by themselves. And classifier can be stated as an algorithm that is used to implement classification especially in concrete implementation, it also refers to a mathematical function implemented by algorithm that will map input data into category. It is an instance of supervised learning i.e., where training set of correctly identified observations is available.

SVM Model (Support Vector Machine)

SVM is a one of the popular machine learning algorithms for regression, classification. It is a supervised learning algorithm that analyses data used for classification and regression. SVM modeling involves two steps, firstly to train a data set and to obtain a model & then, to use this model to predict information of a testing data set. A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane where SVM model represents the training data points as points in space and then mapping is done so that the points which are of different classes are divided by a gap that is as wide as possible. Mapping is done in to the same space for new data points and then predicted on which side of the gap they fall.



In SVM algorithm, plotting is done as each data item is taken as a point in n-dimensional space where n is number of features, with the value of each feature being the value of a particular coordinate. Then, classification is performed by locating the hyperplane that separates the two classes very well.

The steps that we are going to implement are listed below

Step 1: Read the Dataset.

Step 2: Random Sampling is done on the data set to make it balanced.

Step 3: Divide the dataset into two parts i.e., Train dataset and Test dataset.

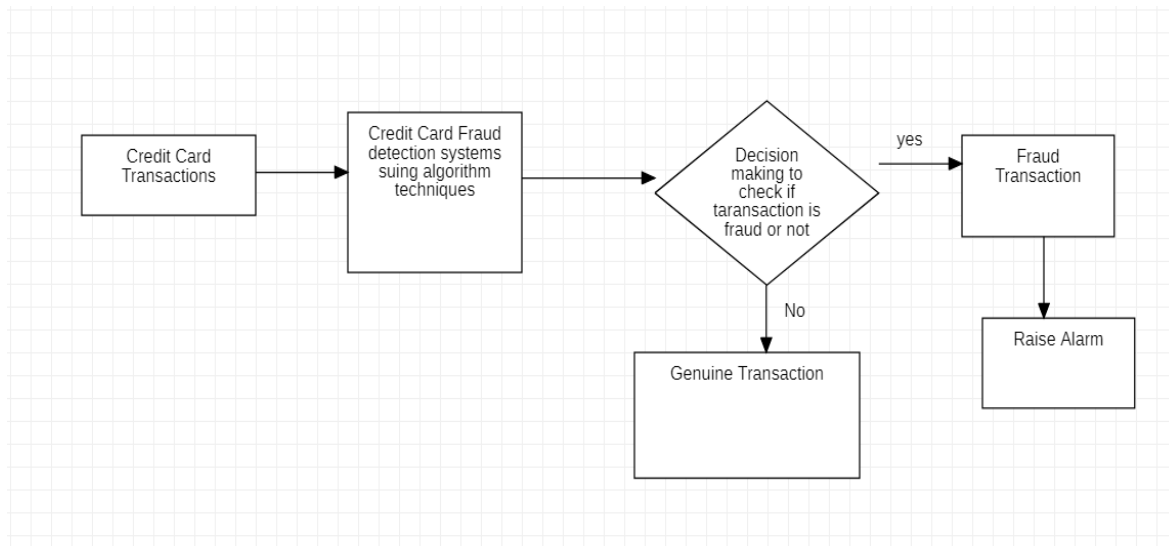
Step 4: Training SVC model with Train dataset.

Step 5: Accuracy and performance metrics has been calculated to know the efficiency for Support Vector Machine algorithm

Chapter 4

Design Models

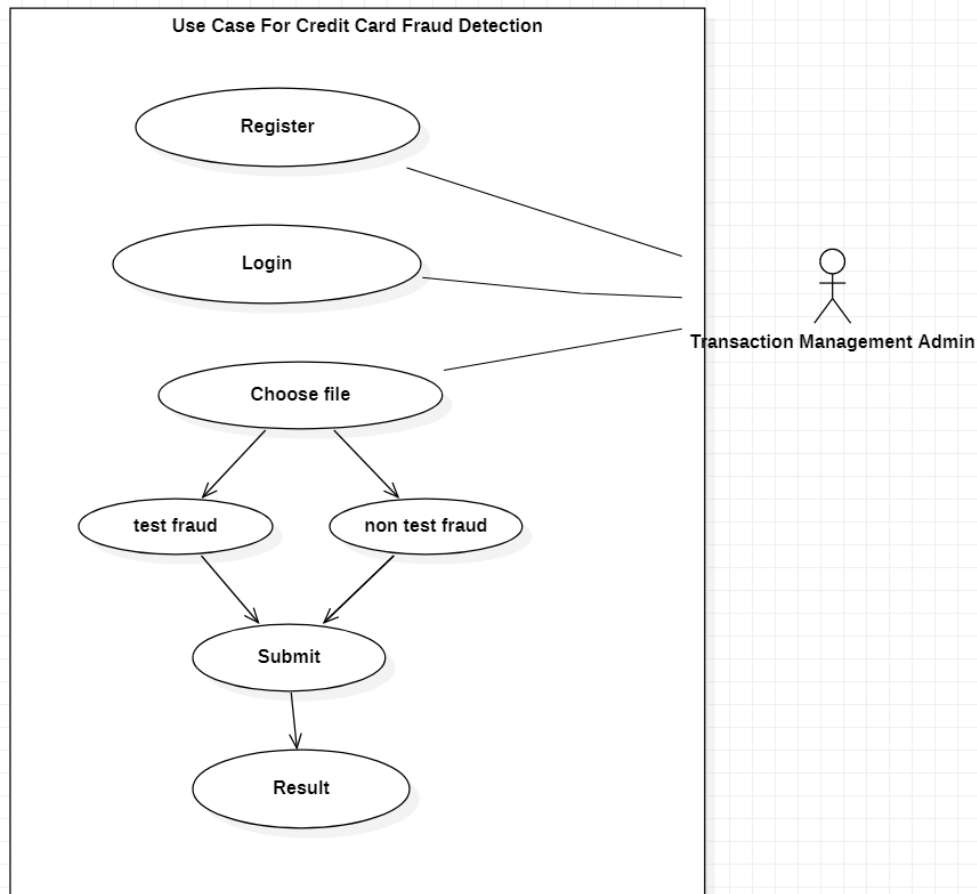
4.1 Block Diagram



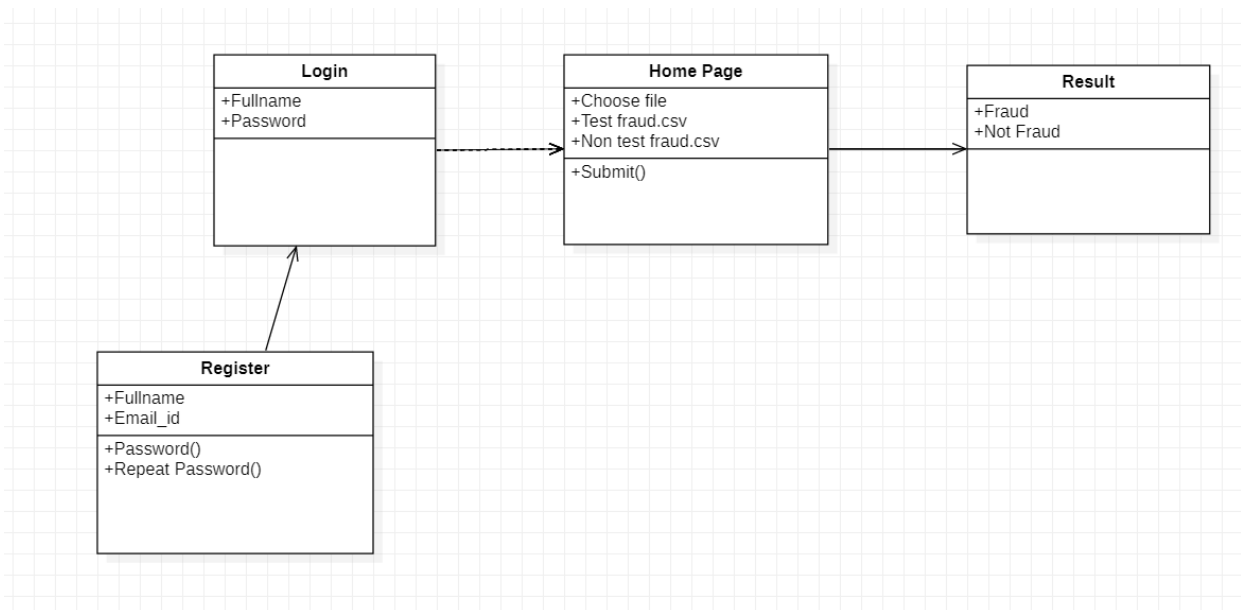
A Block diagram is a type of diagram that represents a workflow or process. In this diagram explains the step-by-step approach to Detect any Fraud Transaction of Credit Card.

4.2 Use Case Diagram

Use Case diagram is used for how the bank management will perform task on this web application. It is outline for bank management people that how system will detect any fraud transactions.

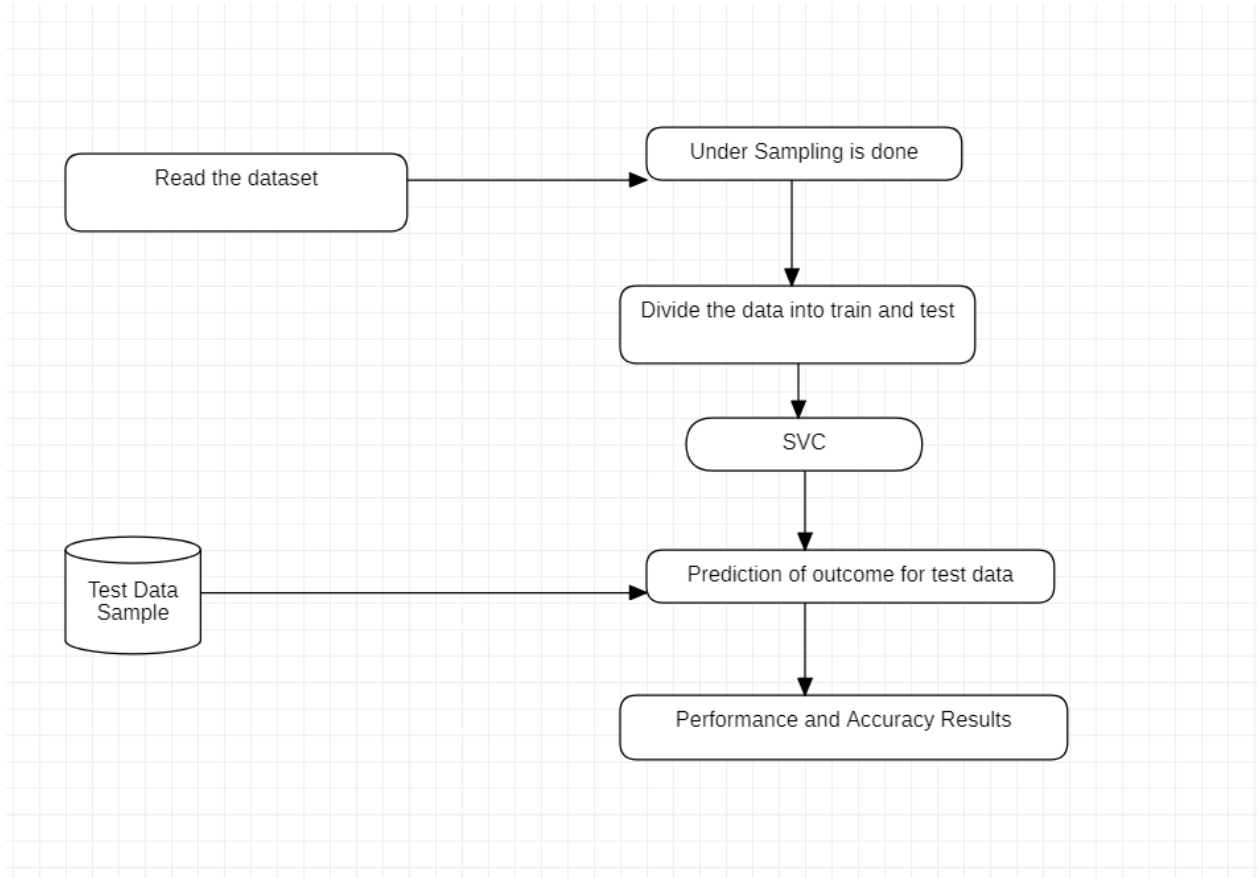


4.3 Class Diagram



Class diagrams show classes and subclasses with related attributes and operations within them, which are performed in Credit Card Fraud Detection. With the help of the above Class diagram, we can construct executable code of the software application. The above classes describe the attributes and operations of a class and also the constraints imposed on the system.

4.4 System flow



A system architecture is the conceptual model that defines the structure, behavior, and more views of a system. In the above diagram we can see that how the machine learning model is trained for prediction using the algorithm

Data Dictionary

Field Name	Data Type	Description	Constraint
Full Name	Integer	Unique ID	Primary Key
Email_id	Integer	Unique ID	Primary Key
Password	Integer	Unique ID	Primary Key
Repeat Password	Integer	Unique ID	Primary Key

