CREDIT CARD FRAUD DETECTION

- *Shaman Chandini*
B.TECH CSE
(2019 - 2023)
3rd year Mini project

INTRODUCTION

As cases of credit card fraud transactions have been increased dramatically in recent years, the amount of card fraud has risen simultaneously.

In our project it was mainly focused on "credit card fraud detection" in the real world initially.

We will collect the credit card dataset and then will provide the user credit card queries for testing the dataset.

MOTIVATION

Millions and billions of people use credit card for payment in both online and offline transactions, due to this countless transactions occur per minute everywhere on the planet.

The reason behind fraud is negligence of the user, when a third person steals the most important information about credit card and user details easily fraud can be achieved.

To detect what type of fraud occurs during a transaction, we need to face several challenges. Fetching that among all the transactions that occurred and which one is real could be the task.

ABSTRACT :

The Usage of cards for online and regular purchases is increasing and so is the fraud related to it.

A large number of fraud transactions are made everyday.

Deep learning presents a solution to the problem of credit card fraud detection make optimal use of their historic customer data as well as real time transaction details that recorded at the time of transaction.

Problem Statement :

In case of existing system the fraud is detected after the fraud is done and so the card holder faced a lot of trouble before the investigation finished.In this proposed model,

We are going to classify fraudulent and legitimate credit card transactions accurate results and to help us to get awareness about the any financially sensitive data by comparing the accuracy obtained.

REQUIREMENT ANALYSIS

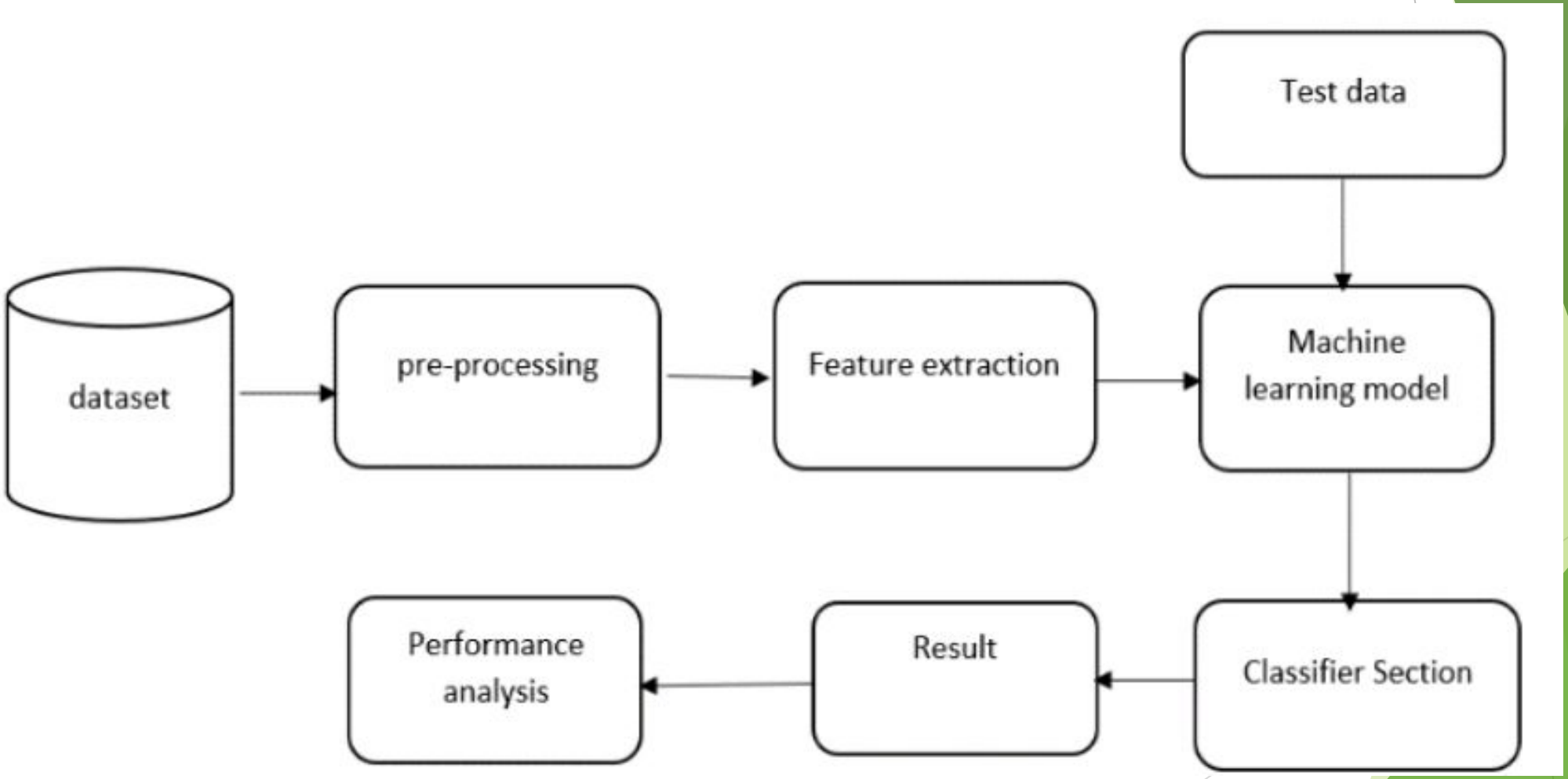
Functional Requirements:

- User details
- Account details
- Credit card details
- Transaction details
- User Accounts to control the access and maintain security

Non-functional Requirements:

- Performance
- Peak workload performance
- Maintainability

SYSTEM ARCHITECTURE

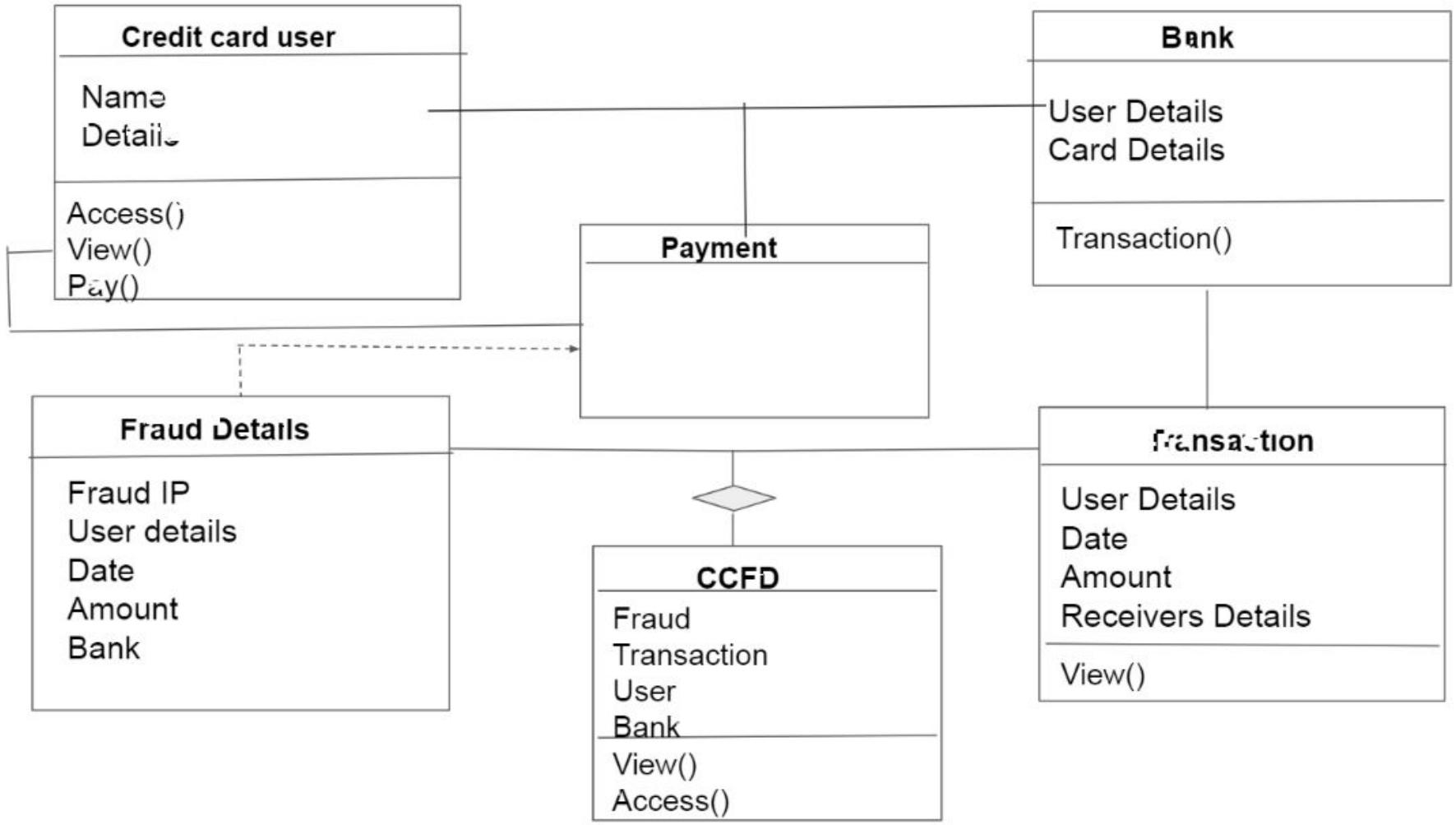


SRS

S.No.	Requirements	Requirement no.	Essential/Desirable	Description
1	Loading of dataset	RS1	Essential	Data set should be loaded without any errors.
2	Data Cleaning	RS2	Essential	Data should be cleaned properly
3	Features Selection	RS3	Essential	Feature selection is Performed using heatmap.
4	Splitting of dataset into train and test data	RS4	Essential	According to the given ratio the dataset should be split into train and test data
5	RFA and Decision tree Algorithms are called for given dataset	RS5	Essential	Each of the model applied should work without any errors
6	Analysis	RS6	Essential	Algorithms are compared.

UML Diagrams

Class Diagram:



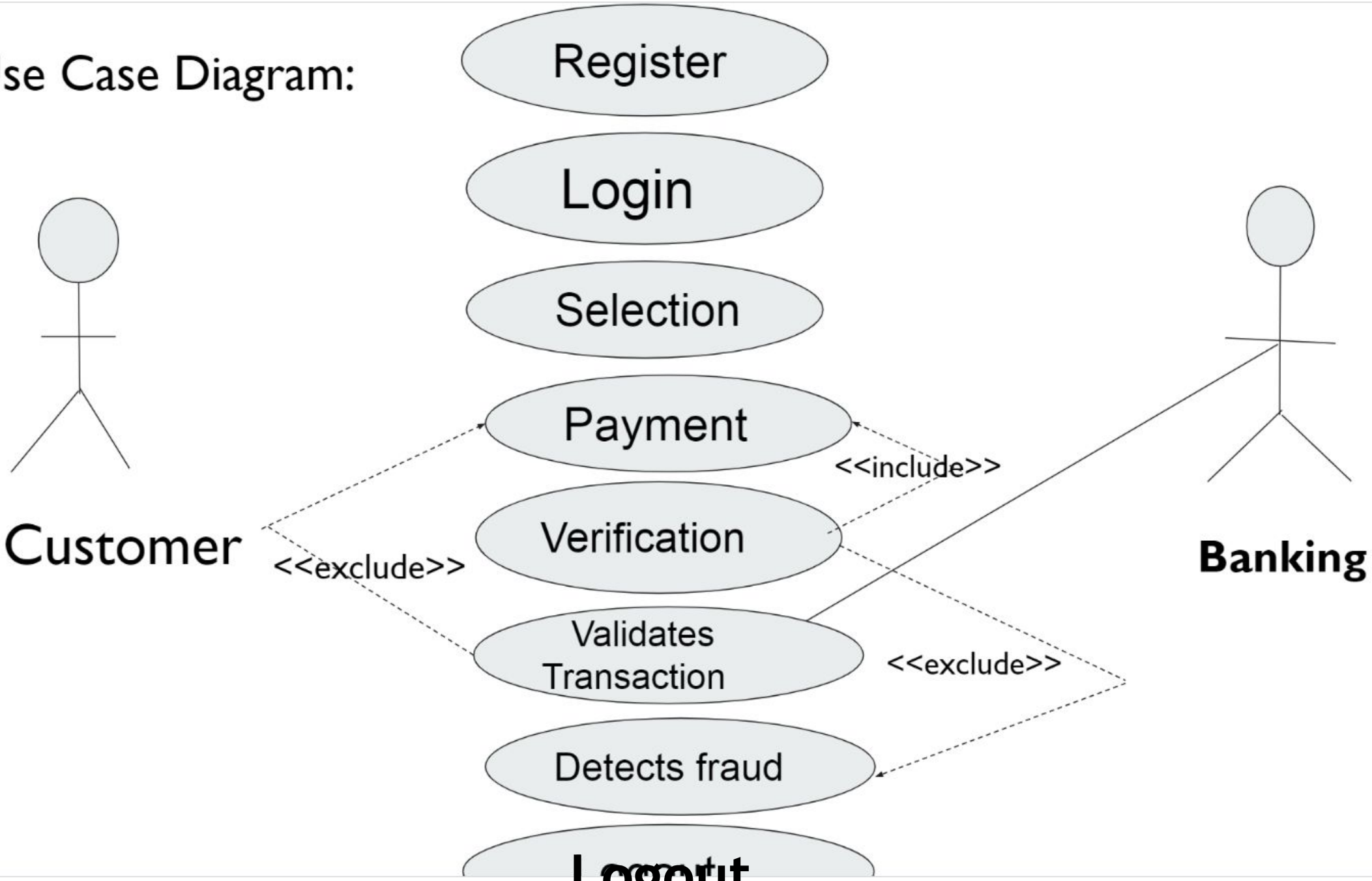
Use Case

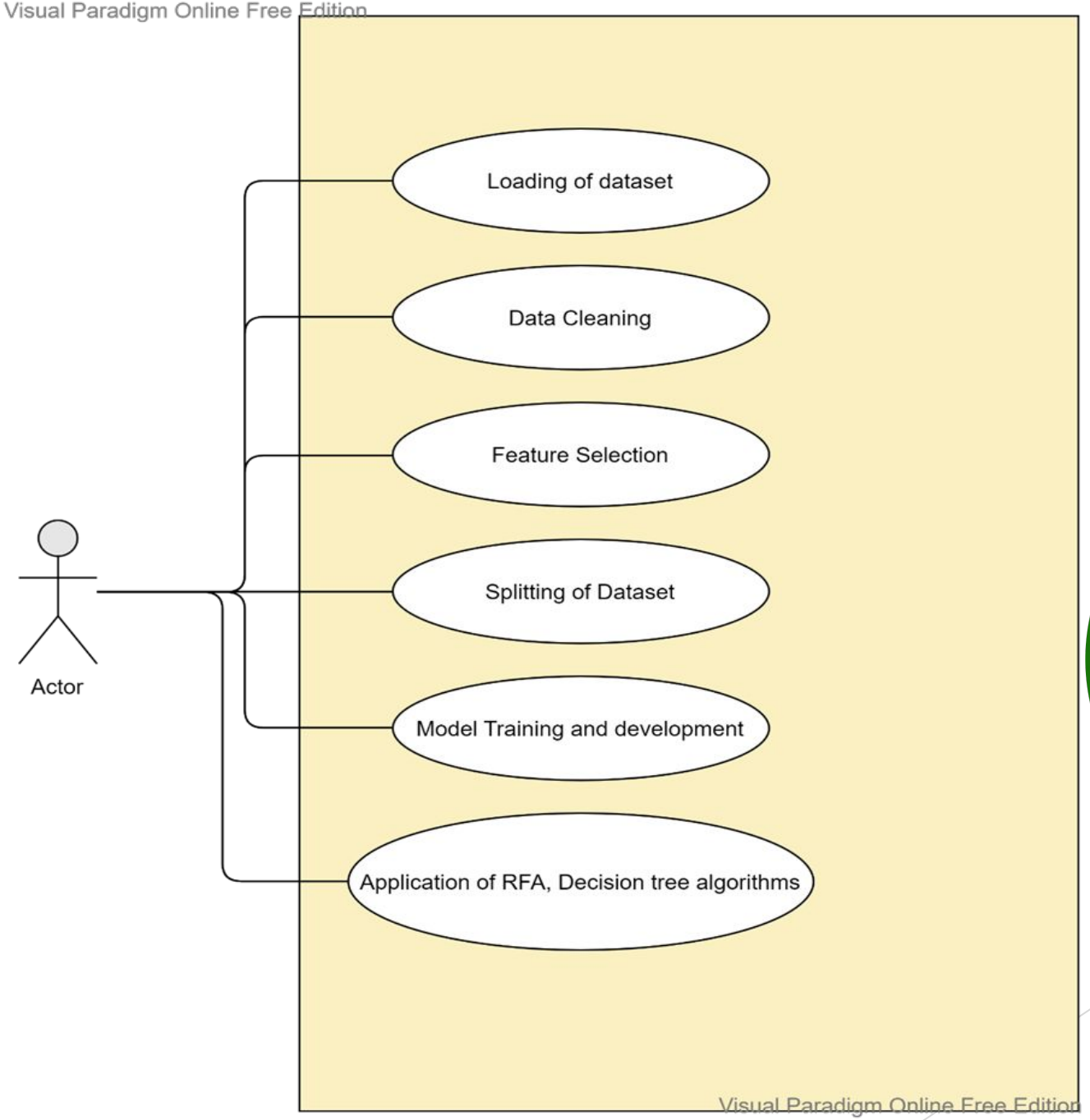
The purpose of use case diagram is to capture the dynamic aspect of a system. But this definition is too generic to describe the purpose.

Use case diagrams are used to gather the requirements are mostly design requirements.

So, when a system is analysed to gather its functionalities use cases are prepared and actors are identified.

Use Case Diagram:



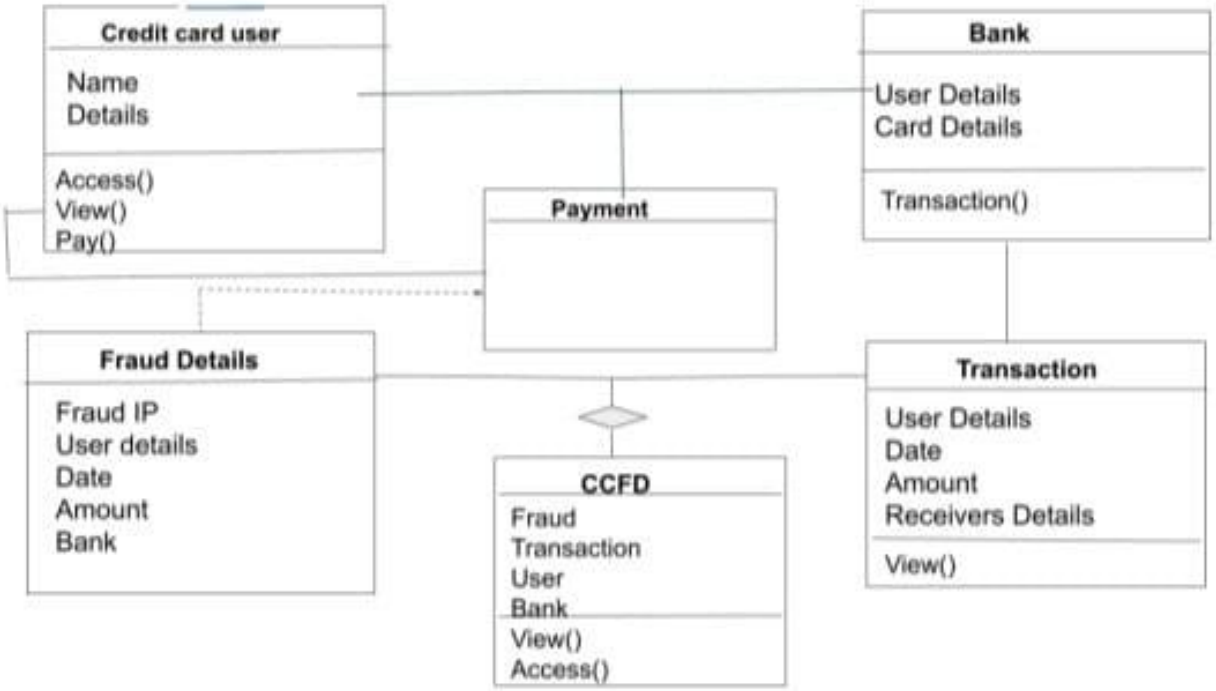


Class Diagram:

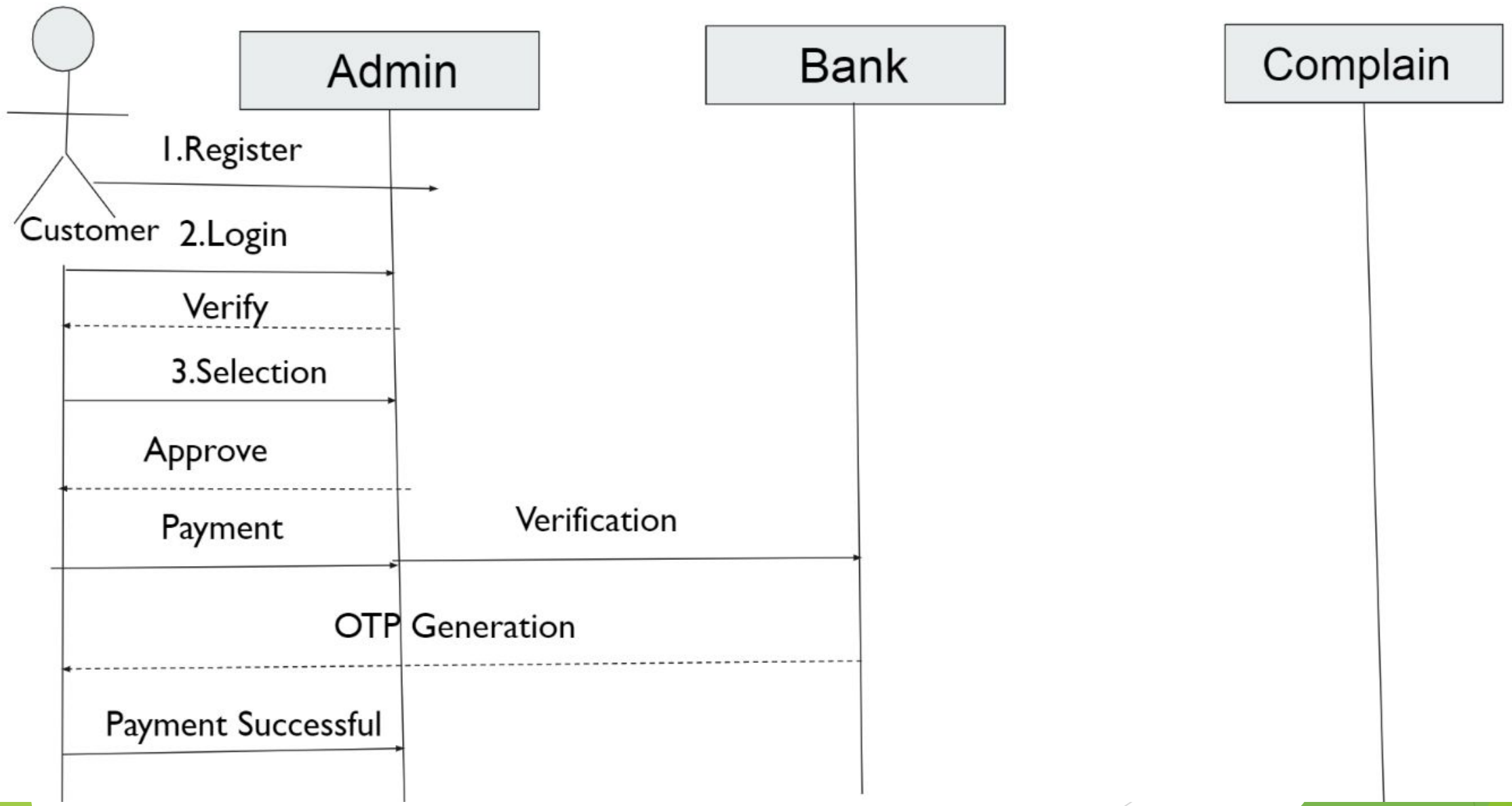
The class diagram depicts a class attributes and operations, as well as the system limitations.

Because class diagrams are the only UML diagrams that can be directly mapped with object-oriented languages, they are frequently utilized in the modelling of object-oriented systems.

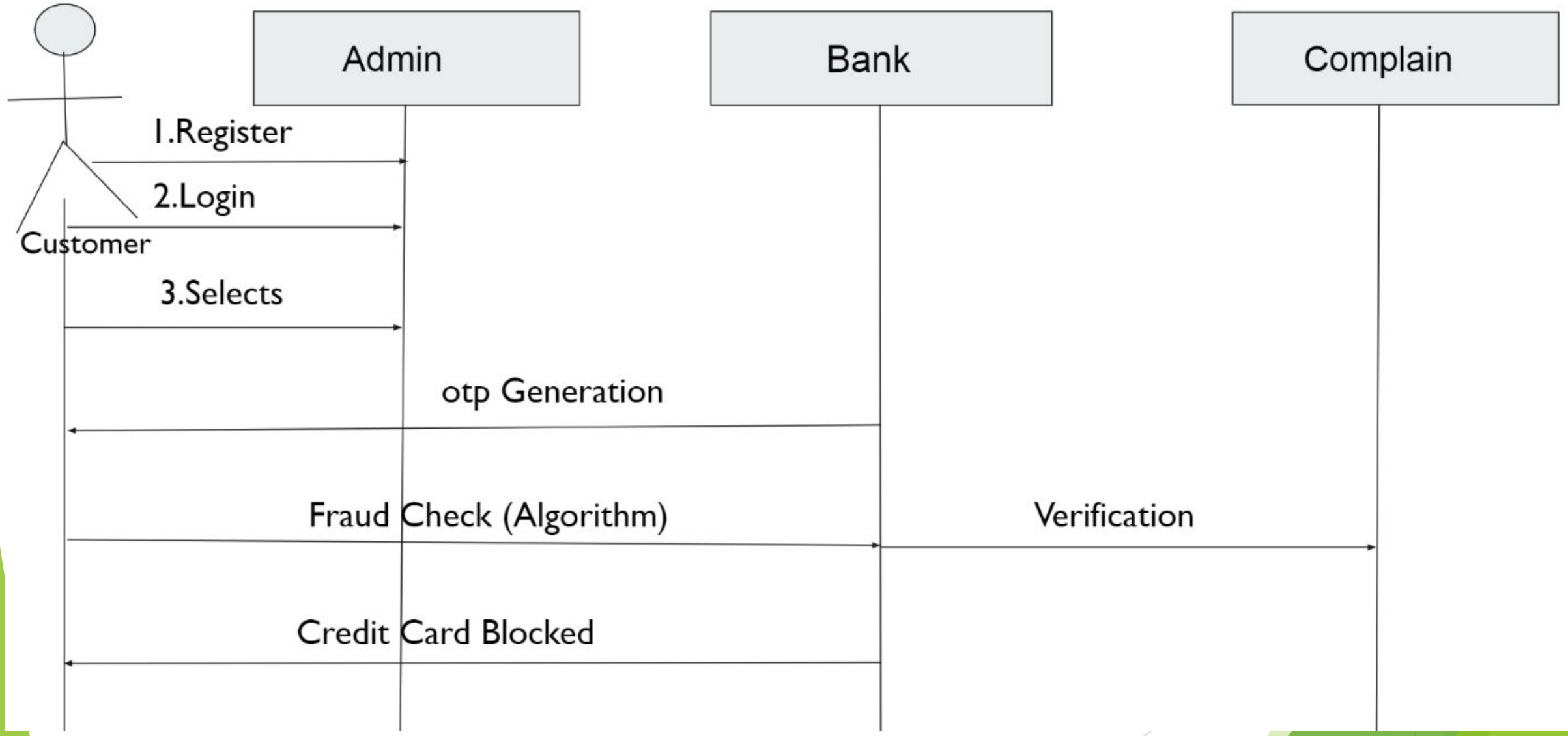
A collection of classes, interfaces, relationships, and collaborations are depicted in the class diagram.


Credit card user
Name
Details
Access()
View()
Pay()
Bank
User Details
Card Details
Transaction()
Payment
Fraud Details
Fraud IP
User details
Date
Amount
Bank
CCFD
Fraud
Transaction
User
Bank
View()
Access()
Transaction
User Details
Date
Amount
Receivers Details
View()

Sequence Diagram: Module 1



Module 2


Admin
Bank
Complain
1.Register
2.Login
Customer
3.Selects
otp Generation
Fraud Check (Algorithm)
Verification
Credit Card Blocked

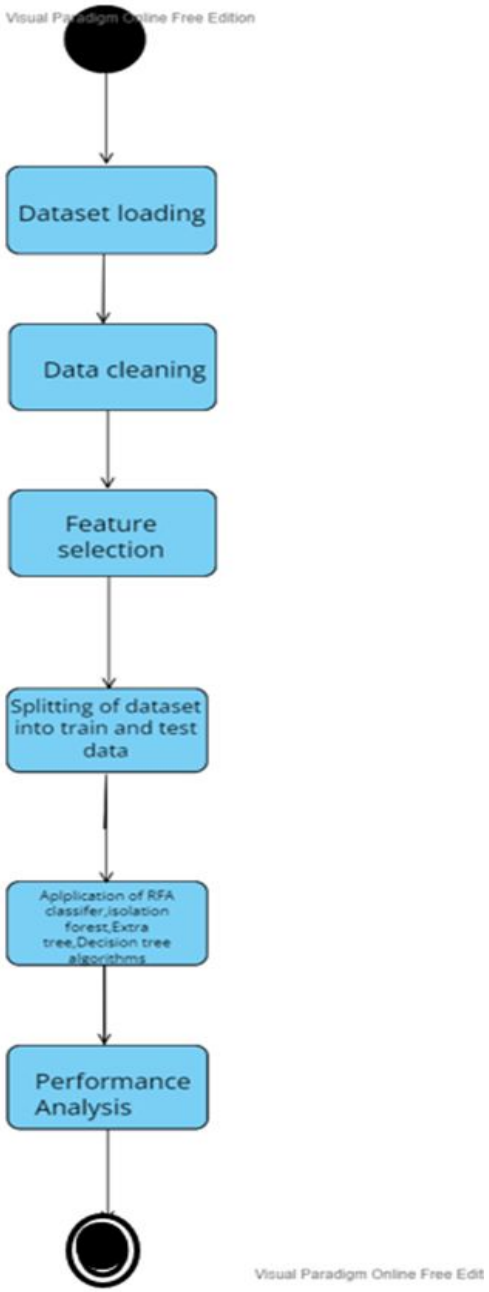
State Chart Diagram:

A State chart diagram describes a state machine. State machine can be defined as a machine which defines different states of an object and these states are controlled by external or internal events.

Following are the main purposes of using State chart diagrams −

- To model the dynamic aspect of a system.
- To model the life time of a reactive system.
- To describe different states of an object during its life time.
- Define a state machine to model the states of an object.

State Chart Diagram:



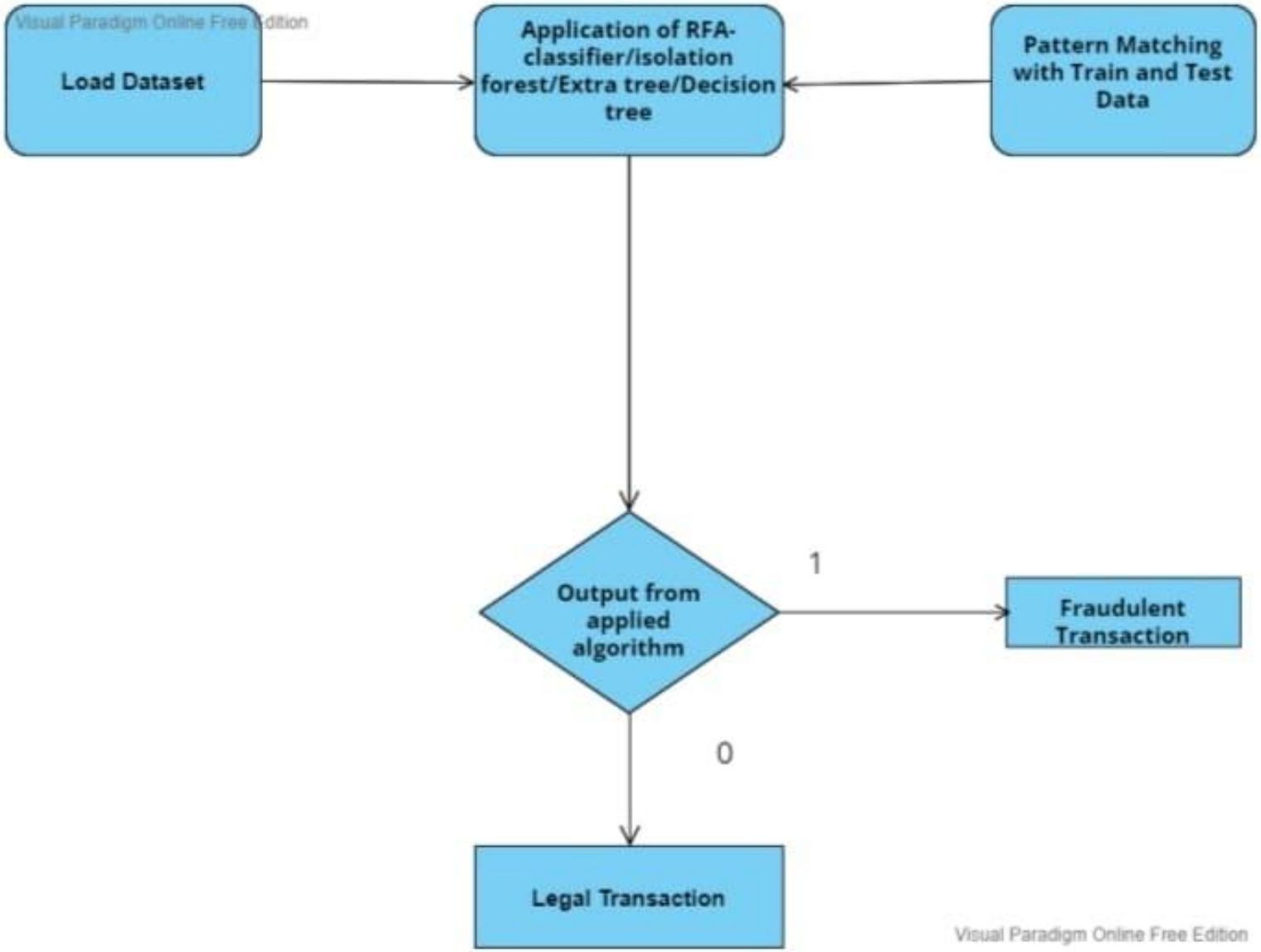
Activity Diagram:

An activity diagram describes the behaviour of a system in terms of activities.

Activities are modelling elements that represent the execution of a set of operations.

The completion of these operations triggers a transition to another activity.

Activity diagrams are similar to flowchart diagrams in that they can be used to represent control flow


Visual Paradigm Online Free Edition
Load Dataset
Application of RFA-classifier/isolation forest/Extra tree/Decision tree
Pattern Matching with Train and Test Data
Output from applied algorithm
1
Fraudulent Transaction
0
Legal Transaction
Visual Paradigm Online Free Edition

Algorithm

Random forest is a supervised algorithm based on ensemble learning. The random forest algorithm is not biased and depends on multiple trees where each tree is trained separately based on the data.

The random forest algorithm works in a similar way and uses multiple algorithm i.e multiple decision trees.
It is a very stable algorithm.

Using this random forest algorithm we have extracted the accurate percentage of detection of fraud from the given dataset.

Test Cases

S. No	REQUIREMENTS	REQ No.	ESSENTIAL/ DESIRABLE	DESCRIPTION	EXPECTED OUTPUT	ACTUAL OUTPUT	RESULT
1.	Loading of dataset	RS1	ESSENTIAL	Dataset is downloaded from kaggle and then loaded	The data set should be loaded without any errors	The data set is loaded without any errors	SUCCESS
2.	Data Cleaning	RS2	ESSENTIAL	Dataset is checked for null values; if any null values are found they should be handled.	Data should be cleaned and should contain no null values.	No null values found	SUCCESS
3.	Splitting of the dataset into train and test data.	RS4	ESSENTIAL	Data set is split into train and test data in the ratio 8:2.	According to the given ratio the dataset should be split into train and test data.	The given dataset is split into train and test data in 8:2 ratio	SUCCESS
4.	RFA and decision tree algorithms are called for given dataset	RS5	ESSENTIAL	RFA, Decision tree are applied.	Each of the model applied should work without any errors.	Results are classified into fraud and nonfraud transaction	SUCCESS
5.	Evaluation and comparison of models	RS6	ESSENTIAL	Accuracy, precision, Recall and F1-score are measured.	All the evaluation properties are measured properly.	All the properties are measured and compared	SUCCESS

Conclusion

Here, a comparative study is made on different tree-based algorithms on a European credit card transactions dataset of size which was taken from Kaggle website.

On raw and pre-processed data, there are three four techniques such as Random Forest Classifier and decision tree are applied in Python.

Based on certain parameters like precision, accuracy, F1-score, recall and performances of these techniques are evaluated. It is seen through the achieved results that in comparison to these four techniques, the performance of Random Forest classifier is better.