VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi-590018, Karnataka

Phase-1
Project work Report
on

"Detect Transaction Anomalies in Credit Card System using Machine Learning"

Submittedin partial fulfilment of the requirements for the award of the degree of Bachelor of Engineering in Computer Science & Engineering

Submitted by

USN	Name
1BI19CS011	AKASH JAIN
1BI19CS015	AMAN ADITYA PANDEY
1BI19CS026	ARINDAM DUTTA
1BI19CS141	SHIVASHANKAR TADAKI

Under the Guidance of
Dr. MAYA B S

Assistant Professor
Dept of CS&E, BIT

Bengaluru-560004

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

BANGALORE INSTITUTE OF TECHNOLOGY

K.R. Road, V.V. Pura, Bengaluru-560 004

2022-23

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi-590018, Karnataka

BANGALORE INSTITUTE OF TECHNOLOGY

Bengaluru-560 004

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Certificate

This is to certify that the project work entitled "**Detect Transaction Anomalies in Credit Card System using Machine Learning**" carried out by

USN	Name
1BI19CS011	AKASH JAIN
1BI19CS015	AMAN ADITYA PANDEY
1BI19CS026	ARINDAM DUTTA
1BI19CS141	SHIVASHANKAR TADAKI

a Bonafede students of VII semester B.E. for the partial fulfilment of the requirements for the Bachelor's Degree in Computer Science & Engineering of the **VISVESVARAYA TECHNOLOGICAL UNIVERSITY** during the academic year 2020-21. 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 said degree.

Internal Guide	Peer Reviewer	HOD, CSE
Dr. MAYA B S		Dr. J Girija
Assistant Professor		Professor and Head

ACKNOWLEDGEMENT

The knowledge & satisfaction that accompany the successful completion of any task would be incomplete without mention of people who made it possible, whose guidance and encouragement crowned my effort with success. We would like to thank all and acknowledge the help we have received to carry out this final year project.

We would like to convey our sincere thanks to our college the **Bangalore Institute of Technology, Dr. Aswath M U, Principal,** for being kind enough to provide an opportunity and platform to complete and present our final year project "**Detect Transaction Anomalies in Credit Card System using Machine Learning".**

We would also like to thank **Dr. Girija J,** Professor and Head of the Department for Computer Science and Engineering, Bangalore Institute of Technology, for her constant encouragement and making us believe in ourselves and ultimately present our final year project "**Detect Transaction Anomalies in Credit Card System using Machine Learning".**

We would also like to express our gratitude to the project coordinators, **Dr. Gunavathi HS & Dr Mamatha V**, Assistant Professor for coordinating the project activities.

We are most humbled to mention the enthusiastic influence provided during the development and ideation phase by our guide **Dr. Maya BS**, Assistant Professor, Department of Computer Science & Engineering for her ideas, time to time suggestions, and co-operation shown during the venture and help make our final year project a success.

AKASH JAIN (1BI19CS011)

AMAN ADITYA PANDEY(1BI19CS015)

ARINDAM DUTTA (1BI19CS026)

SHIVASHANKAR TADAKI (1BI19CS141)

ABSTRACT

Credit card fraud detection is presently the most frequently occurring problem in the present world. This is due to the rise in both online transactions and e-commerce platforms. Credit card fraud generally happens when the card was stolen for any of the unauthorized purposes or even when the fraudster uses the credit card information for his use. In the present world, we are facing a lot of credit card problems. To detect the fraudulent activities the credit card fraud detection system was introduced. This project aims to focus mainly on machine learning algorithms. The algorithms used are random forest algorithm and the Adaboost algorithm. The results of the two algorithms are based on accuracy, precision, recall, and F1-score. The ROC curve is plotted based on the confusion matrix. The Random Forest and the Adaboost algorithms are compared and the algorithm that has the greatest accuracy, precision, recall , and F1-score is considered as the best algorithm that is used to detect the fraud.

Digital transactions can take place over the phone or on the internet. For executing a transaction, very basic information is required such as expiry date, card number, card verification number etc. Cardholders provide this information through phone or the internet. Fraudsters apply several techniques and attempt to steal the credit card information of the customers so that they can use it for doing fraudulent transactions. It is a very serious, and costly problem for financial service providers. Billions of dollars are subject to fraudulent transactions every year. The fraudulent transaction is an issue of concern for all the credit card providers or by expansion for all the financial systems that provide the facilities for online transactions to their customers. It is usually the result of someone stealing the credit card information of the customers which also impact the brand value of the credit card service providers and the merchants.

Fraud detection involves monitoring the activities of populations of users to estimate, perceive or avoid objectionable behavior, which consist of fraud, intrusion, and defaulting. This is a very relevant problem that demands the attention of communities such as machine learning and data science where the solution to this problem can be automated. This problem is particularly challenging from the perspective of learning, as it is characterized by various factors such as class imbalance. The number of valid transactions far outnumber fraudulent ones. Also, the transaction patterns often change their statistical properties over the course of time.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

INRODUCTION

1.1 Overview

- Despite the promising progress made in detecting anomalies in day-to-day transactions, identifying frauds remains a challenging task due to semantic gap between the various predefined fraud detection models and diversities in implementing them.
- The hybrid approach made it possible to train generative models for the system to maintain the privacy as the transaction data is kept in a decentralized manner using the concept of federated learning.
- This is a very relevant problem that demands the attention of communities such as machine learning and data science where the solution to this problem can be automated. This problem is particularly challenging from the perspective of learning, as it is characterized by various factors such as class imbalance.
- Digital transactions can take place over the phone or on the internet. For executing a transaction, very basic information is required such as expiry date, card number, card verification number etc. So accidental disclosure of any one of them leads to serious disruptions

Fig 1.1. Credit Card Fraud Detection

1.2 Objectives

- Design and implement the technique for detecting anomalies in the credit card system using the existing algorithms.
- Customizing the existing bank authentication system. Implement the message authentication system for successful transaction to the user.
- Ability to identify new customer behavior patterns and adapt to changes.
- Unlike rule-based systems, algorithms are to be aligned with a constantly changing environment and financial conditions

1.3 Purpose, Scope, and Applicability

1.3.1 Purpose

- Fraud detection costs huge money loss to different financial companies and consumers so it have become essential for banks and financial institutions to minimize their losses.

- With digital crime and online fraud of all kinds on the rise, it's more important than ever for organizations to take firm and clear steps to prevent payment card fraud through advanced technology and strong security measures.

1.4 Scope

- The scope of this project is to classify the transactions are fraudulent or not,
    - Obtain factual and accurate information that will lead to an appropriate credit decision.
    - Predict fragility based on the transaction amount, location, and other transaction related data.

1.5 Applications

- Customer Service
- Banking
- Payment Gateways
- Merchandise Payments

Chapter 2

LITERATURE SURVEY

2.1 Review of Machine Learning Approach on Credit Card Fraud Detection [1]

Proposed Idea:

- The idea is utilizing the real-time datasets to train the model in a privacy-preserving manner.
- A Federated learning(decentralized) framework with ANN can enhance the capability of the ML model to detect fraudulent transactions

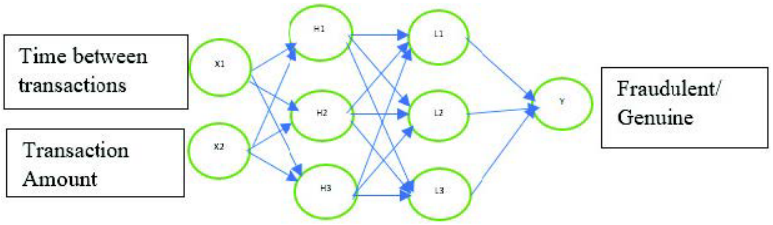

Fig 2.1. Flowchart of the process of detection of CC fraud detection

Conclusion: The efficient method to identify an affected person so as to save the users is proposed in this paper . Also, Privacy of customers of the banks are preserved.

Drawback: Although data is not shared centrally, even the trained model will be going to learn patterns that can be possibly decoded by hackers.

Therefore, while keeping the limitations in place, there still needs to be work done for gaining the confidence of banks and financial institutes to adopt this technology.

2.2 Credit Card Fraud Detection using Machine Learning Algorithm [2]

Proposed Idea:

- Customers are grouped based on their transactions and extract behavioral patterns to develop a profile for every cardholder.
- Then different classifiers are applied on three different groups (low ,medium, high) ,later rating scores are generated for every type of classifier.
- This dynamic changes in parameters lead the system to adapt to new cardholder's transaction behaviors timely. Followed by a feedback mechanism to solve the problem of concept drift.

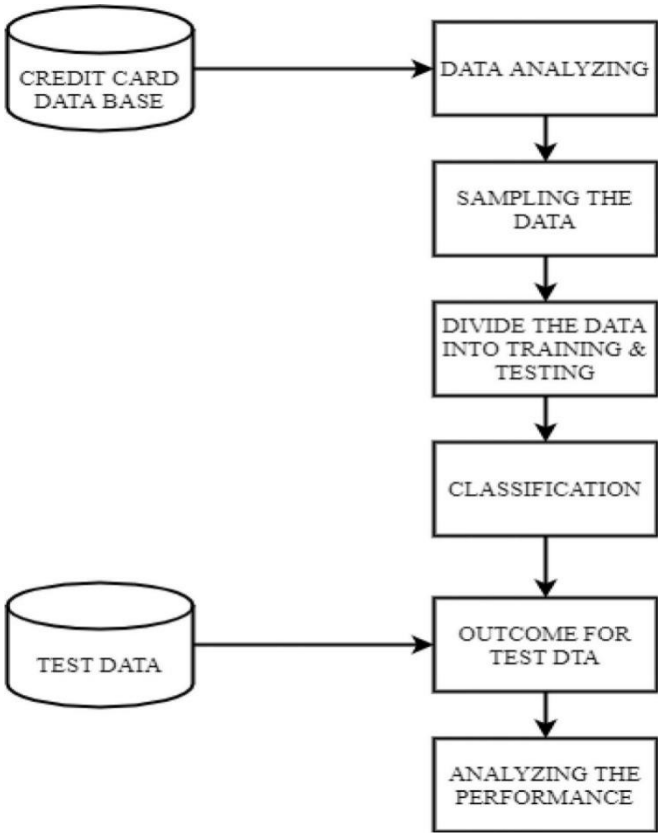

Fig: 2.2. Credit card framework for working of the model

Conclusion: It tackles the problem of concept drift.

2.3 Implementing Machine Learning in detecting transaction anomalies [3]

Proposed Idea:

- The main aim of this paper is to classify the transactions that have both the fraud and non-fraud transactions in the dataset using algorithms like that the Random Forest and the Adaboost algorithms. Then these two algorithms are compared to choose the algorithm that best detects the credit card fraud transactions includes many steps from gathering dataset to deploying model and performing analysis based on results.
- In this model we take the Kaggle dataset and pre-processing is to be done for the dataset. We split the data into the training data and the testing data.
- We use the training data to prepare the Random Forest and the Adaboost models. Then we develop both the models. Finally, the accuracy, precision, recall, and F1-score is calculated for the models.

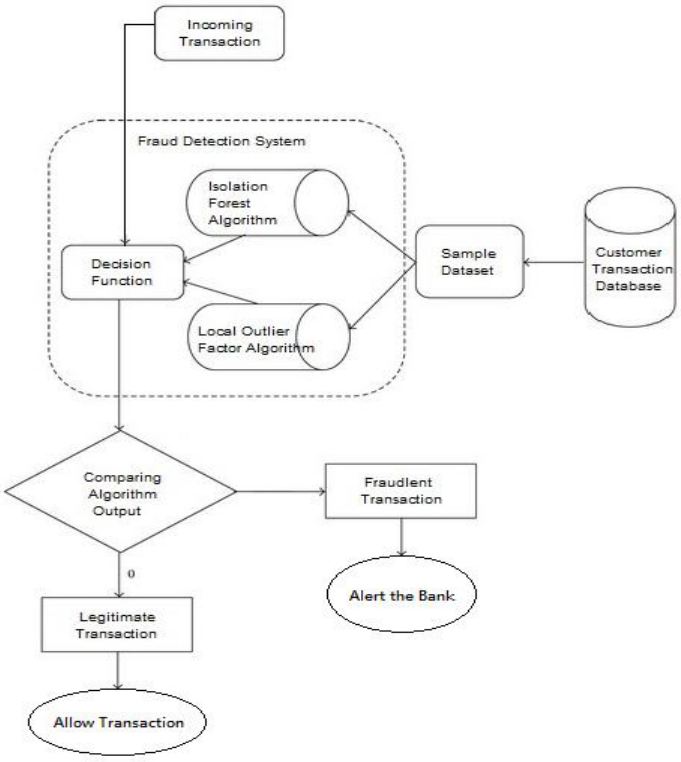

Fig 2.3. Work Flowchart of CC fraud detection data

Main Drawback: To achieve good accuracy and low computation time they selected just 10 features from the dataset.

2.4 GA algorithm for feature selection and enhancement [4]

Proposed Idea:

- we implement a feature selection algorithm that is based on the Genetic Algorithm (GA) using the RF method in its fitness function. The RF method is used because it can handle many input variables, can automatically handle missing values, and is not affected by noisy data.
- A GA-based FS in order to increase the performance of ML based models applied to the domain of intrusion detection systems. of Evolutionary Algorithm (EA) that is often used to solve several optimization tasks with a reduced computational.

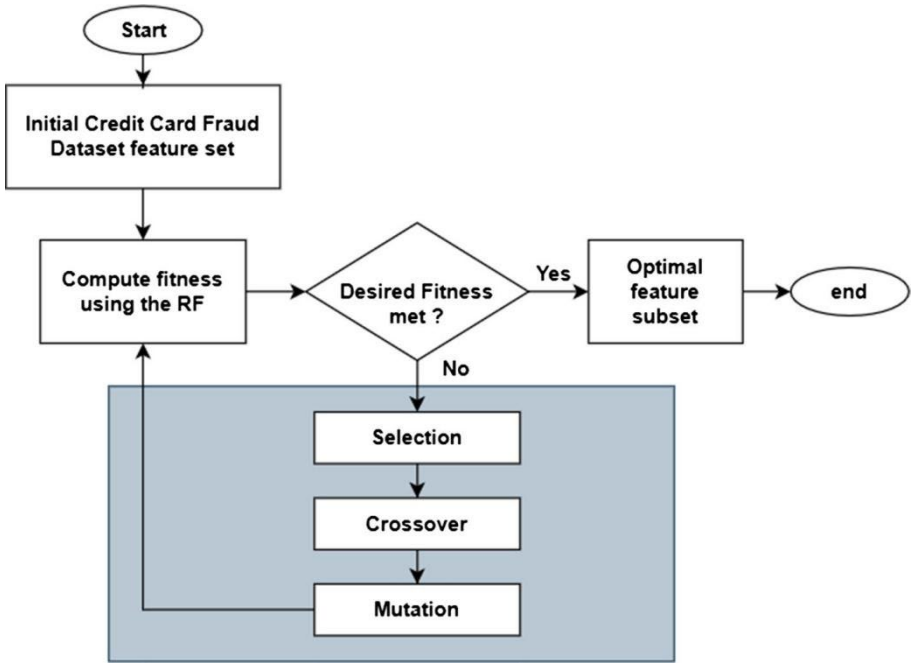

Fig 2.4 Architecture of CC fraud detection

2.5 Implementing Hybrid model in the Fraud Detection [5]

Proposed Idea:

- In this paper, we have studied the behavior pattern based on their previous transaction records. They classify all the attributes of transaction and construct the logical graph of behavior profile (LGBP).
- They define the state transition probability matrix to attain the features of transaction and construct behavior pattern for each user and propose a method to determine whether the transaction is done by genuine user or not.
- They don't require predictive model and their outlier detection mechanism helps to detect the card fraud using less memory and computation requirements.

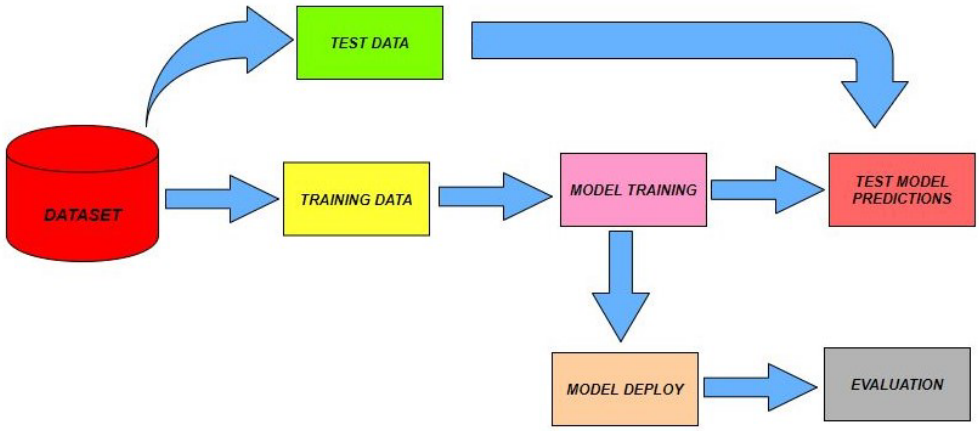

Fig2.5. Framework of the CC fraud detection model

Conclusion: We have done a comparative study of the results of K-Nearest Neighbors and Convoluted Neural Network and the hybrid model of both. Among the individual models, The KNN had the highest accuracy rate of 90.66%, followed by CNN with 88.12% accuracy. Upon hybridization, the resultant model had accuracy of 98%.The accuracy of the Convolutional Neural Network increased by 10% when made into a hybrid model with K-Nearest Neighbors, and would only improve if trained over larger balanced dataset.

2.6 Credit Card Fraud Detection using Deep Learning [6]

Proposed Idea:

- In this paper, we put forth a method of Fraud Detection which is completely based on Deep Learning. We first compare it with all the renowned methods such as Random Forest, Support Vector Machines, etc.
- Finally, we come across a conclusion that Neural Networks, even though harder to train, would be a perfect fit for the Model.

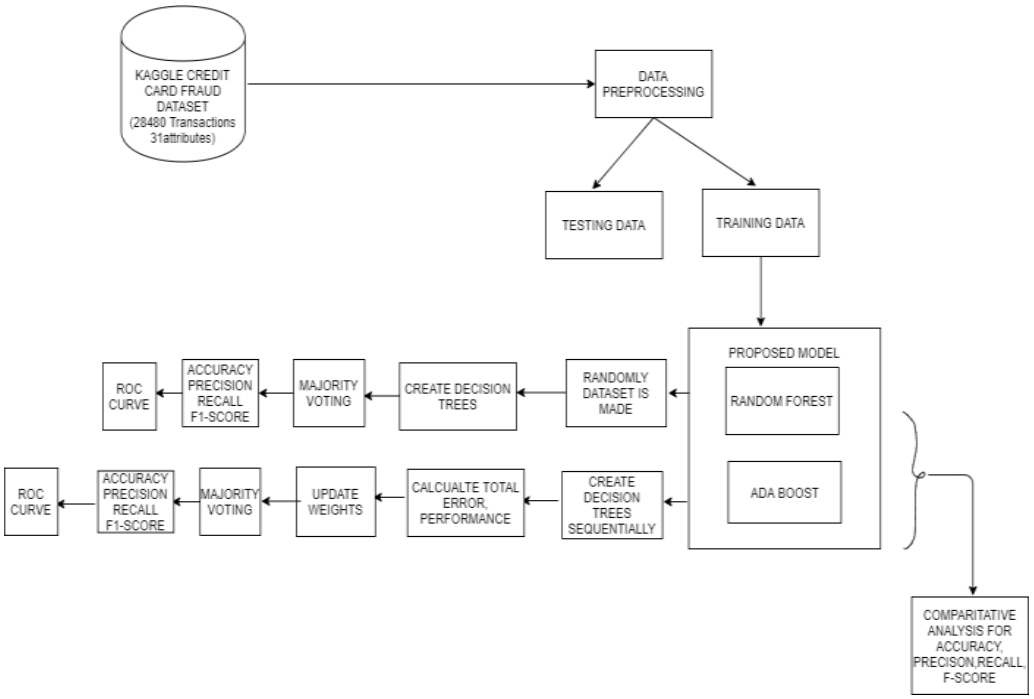

Fig 2.6. Architecture of the CC fraud detection framework model

Drawback: we can conclude that, Kera's based Deep Learning Neural Network proves to be a great alternative to other classifiers mentioned above. Also, no matter how accurate the trained model of the network might be, it will not show accurate results unless the skewness of the data is reduced.

2.7 Credit Card Fraud Detection using Machine Learning and Data Science [7]

Proposed idea:

- This article has listed out the most common methods of fraud along with their detection methods(outliers) and reviewed recent findings in this field.
- This paper has also explained in detail, how machine learning can be applied to get better results in fraud detection along with the algorithm, pseudocode, explanation its implementation and experimentation results.

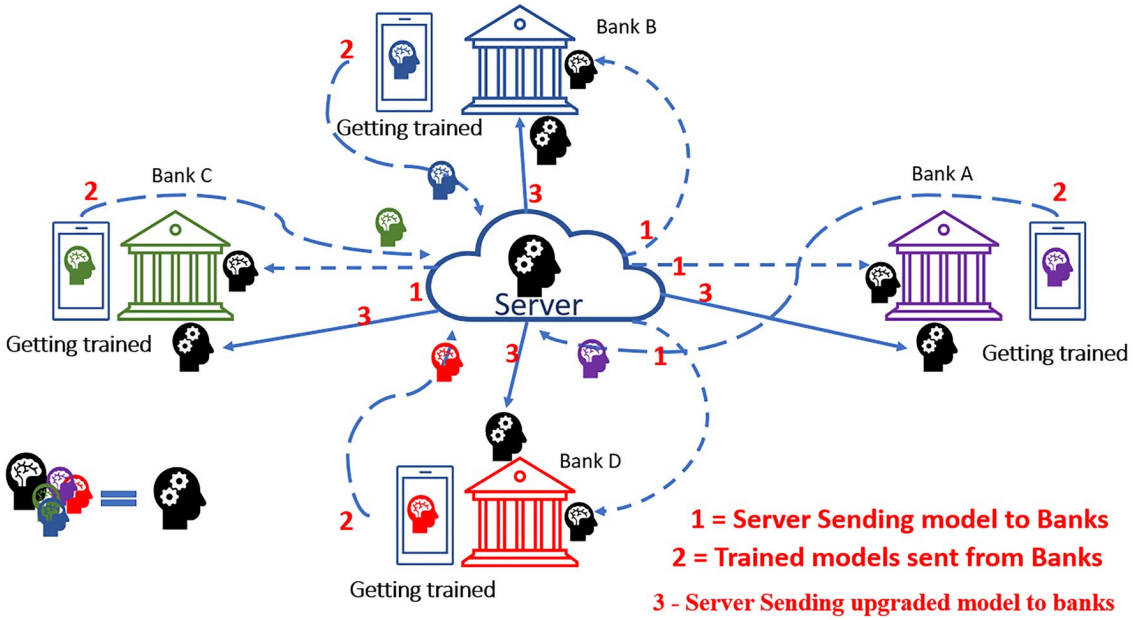

Fig 2.7. Workflow of training and testing our models

Drawbacks: This idea is difficult to implement in real life because it requires the cooperation from banks, which aren't willing to share information due to their market competition, and due to legal reasons and protection of data of their users.

2.8 Autonomous credit card fraud detection using machine learning approach [8]

Proposed idea:

- This paper proposes a Machine Learning models such as Naive bayes, SVM, ANN, and LSTM-RNN which have been utilized to detect fraud in the credit card system.
- The suggested system's performance is measured using sensitivity, precision, accuracy, and error rate.
- Traditional techniques are no longer effective in the age of big data. As a result, the team developed a model for detecting credit card fraud based on the Long Short-Term Memory technique using an actual data set of credit card fraud.
- This model was created to improve current detection tactics as well as detection accuracy in light of big data.
- It used deep learning techniques to quickly and effectively identify patterns, overcoming the difficulty of recognizing unexpected and sophisticated fraud practices.

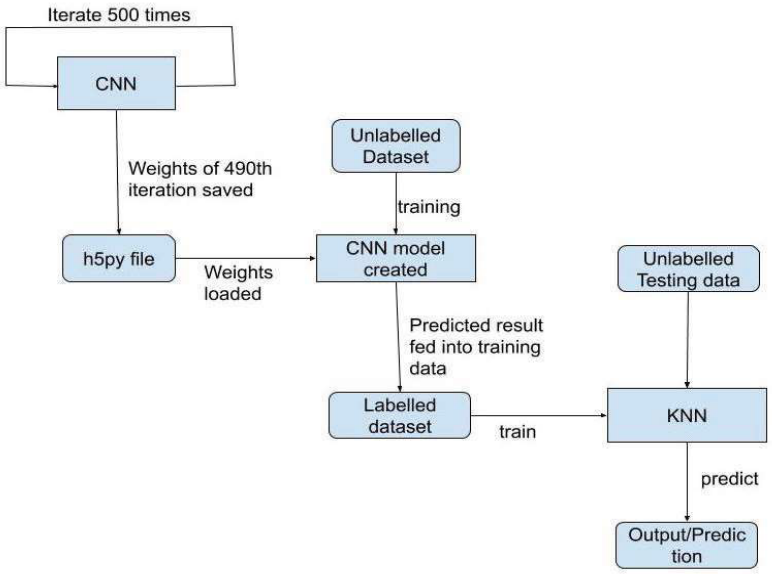

Fig 2.8. Overall process of proposed methodology

Drawbacks: The watershed algorithm used is highly sensitive to noise. If there is an image with noise, then it will influence the segmentation.

2.9 Imbalanced Classification Approaches for Credit Card Fraud Detection [9]

Proposed Idea:

- This paper report a rigorous experimentation and compare the performance of solutions that deal with the imbalance classification problem. Also, we identify their weaknesses to help researchers target the right issues for tackling the problem in the real world.
- Two techniques are employed in imbalanced classification approaches. The first technique is employed on data as a prepossessing step to balance classes, like oversampling, under sampling, etc. The second technique is used within the classification algorithm like Cost-Sensitive (CS) approaches or One-Class Classification (OCC).

  1.Random Oversampling (RO) – It is used to balance classes by simply replicating observations as needed until the balance between classes is reached.

  2.One Class Classification (OCC) – This approach uses only one class of the data (usually the minority class) and learns its characteristics.

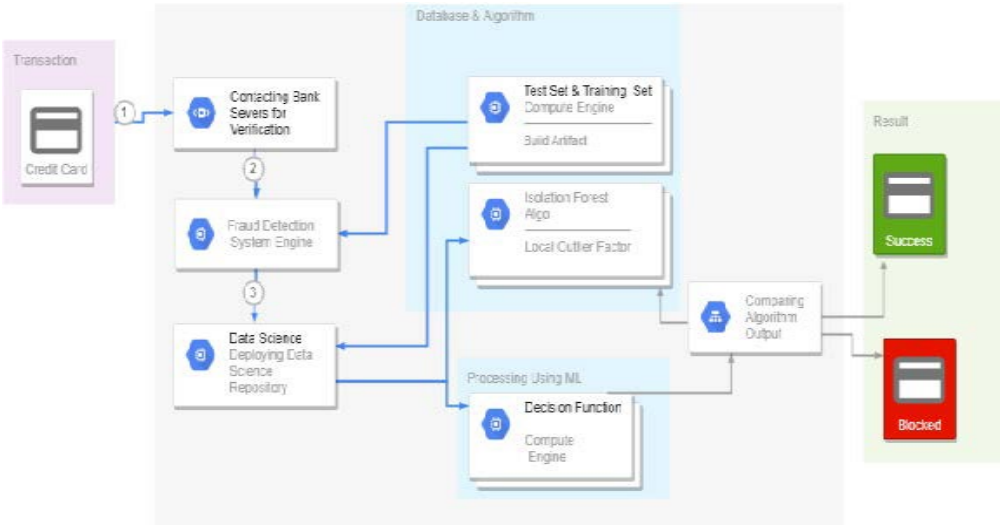

Fig 2.9. Framework of CC fraud detection

2.10 Credit Card Fraud Detection Using AdaBoost and Majority Voting [10]

Proposed idea:

This paper proposes single and hybrid machine learning algorithms for financial applications. Various financial applications and financial statements are reviewed.

1. Single Models - For credit card fraud detection, Random Forest (RF), Support Vector Machine, (SVM) and Logistic Regression (LOR) will be examined.

2. Hybrid Models - Hybrid models are combination of multiple individual models. A hybrid model consisting of the Multilayer Perceptron (MLP) neural network, SVM, LOR, and Harmony Search (HS) optimization.

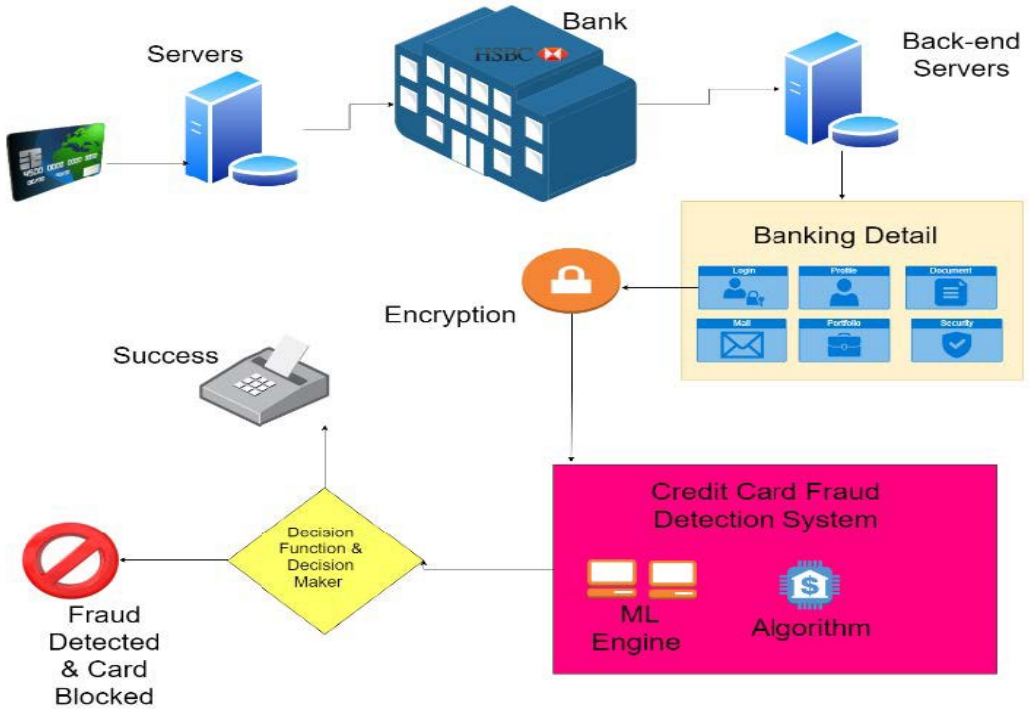

Fig 2.10. Block diagram of proposed system

Conclusion: A total of twelve algorithms are used in this experimental study. They are used in conjunction with the Ada Boost and Majority Voting methods.

Chapter 3

REQUIREMENT ENGINEERING

3.1 Software and Hardware Tools Used

3.1.1 Software Requirements

Python 3	Python is an interpreted, high-level, and general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant whitespace.
Pip	Pip is a package-management system written in Python used to install and manage software packages.
NumPy	NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.
Anaconda	Anaconda is a distribution of the Python and R programming languages for scientific computing, that aims to simplify package management and deployment.
Google Collab	Collaboratory is a product from Google Research. Collab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis.

Table 3.1. Software Requirements

3.1.2 Hardware Requirements

- Processor: Intel i5 10th gen or more.
- Ethernet connection (LAN) Oral wireless adapter (Wi-Fi).
- GPU - NVIDIA GeForce MX150- 4GBorabove.
- Memory (RAM): Minimum4 GB; Recommended16GB or above.

3.2 Conceptual Modeling

3.2.1 Use Case Diagram

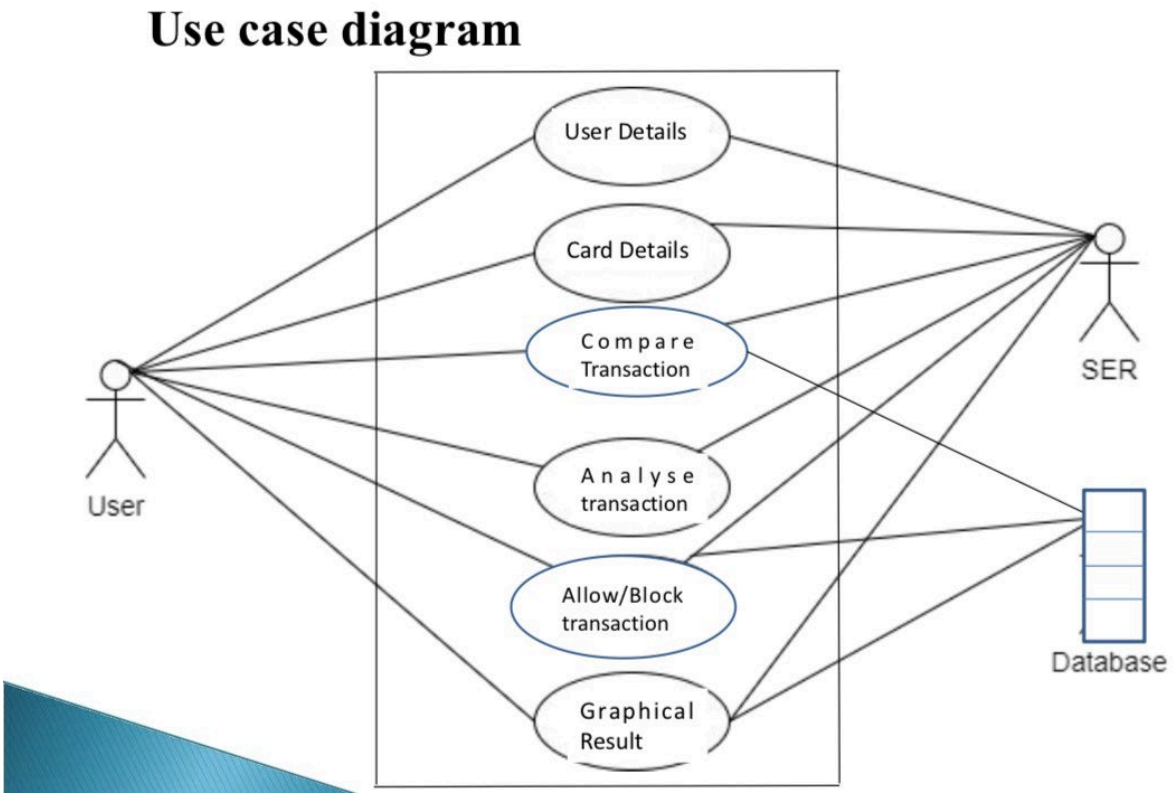

Fig 3.1 Use case diagram

3.2.2 Sequence Diagram

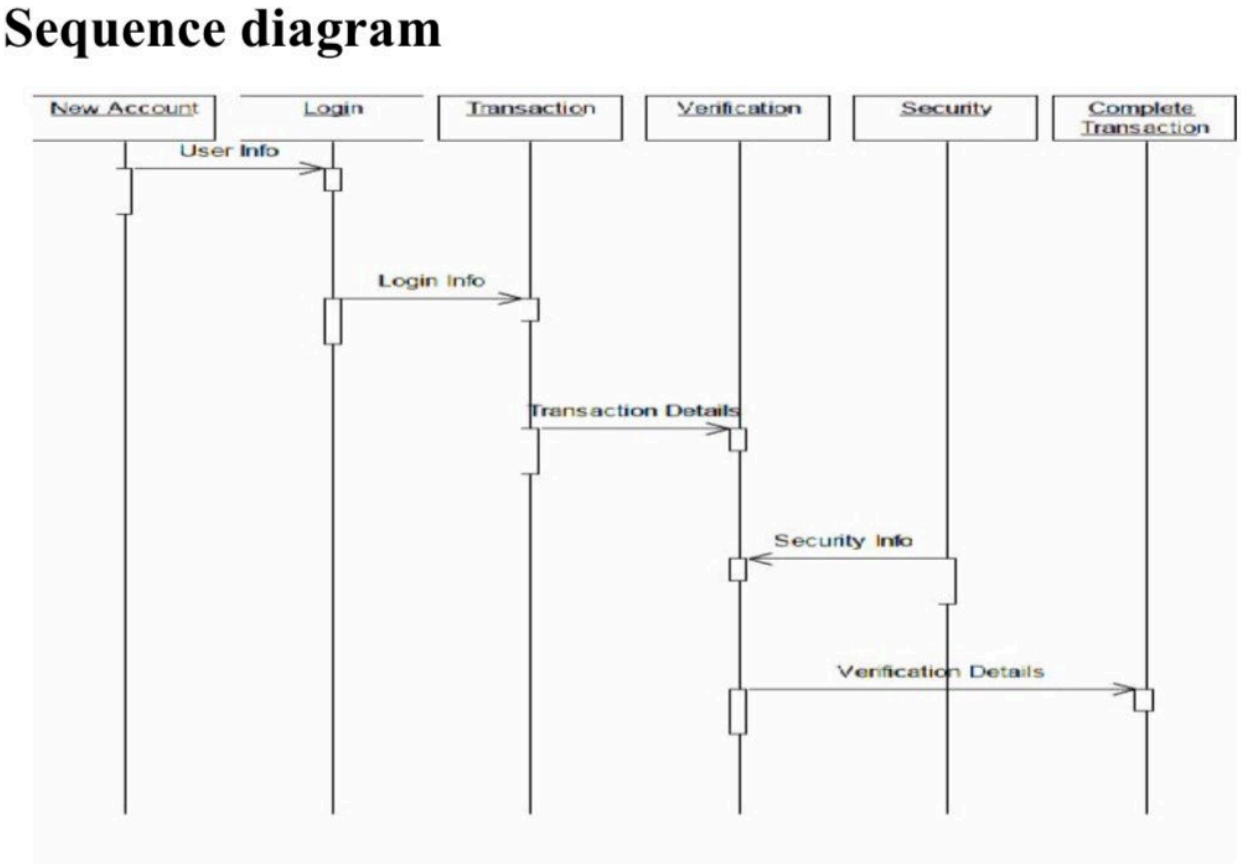

Fig 3.2 Sequence Diagram

3.2.3 Activity Diagram

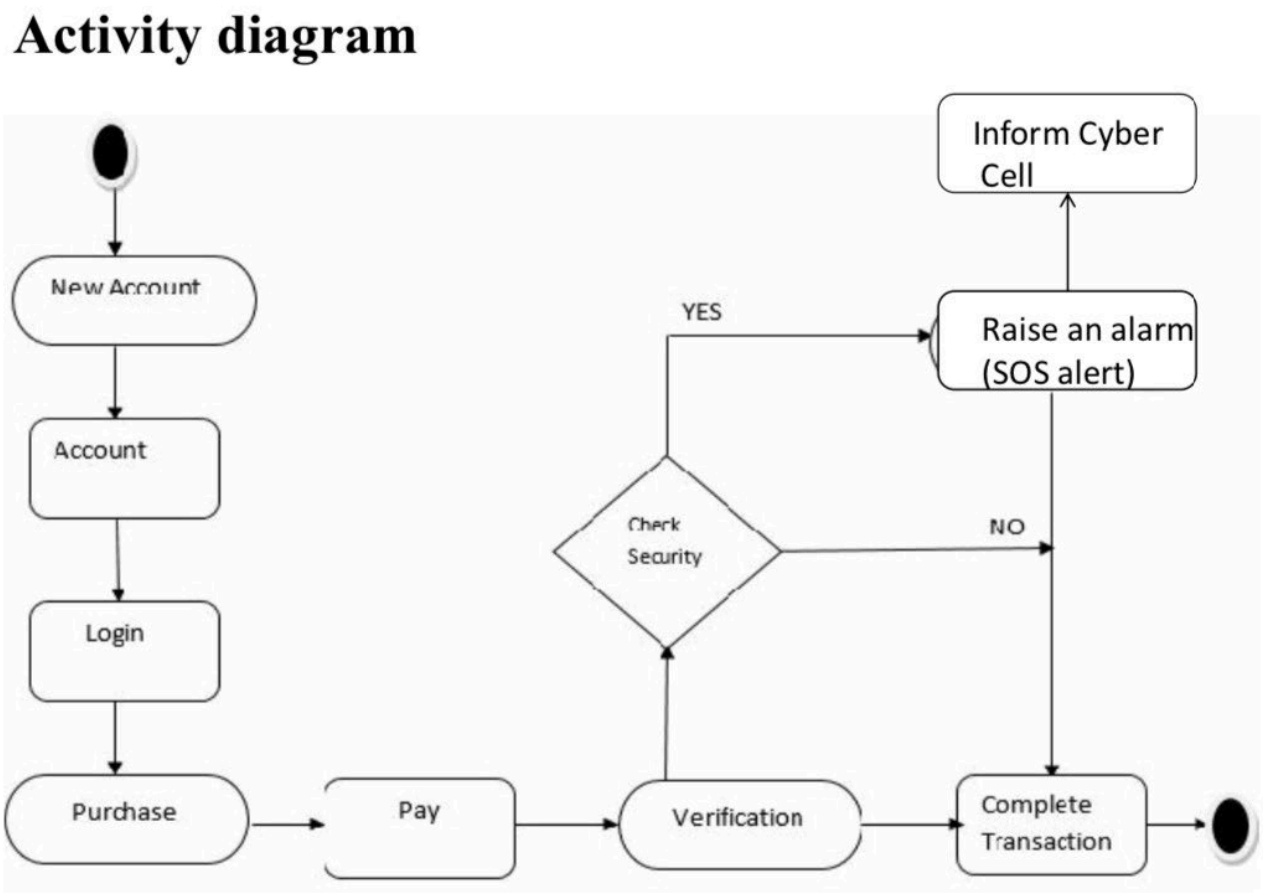

Fig 3.3 Activity Diagram

2.4 State Diagram

As seen in the figure 3.4, the state diagram displays the state in which our system will be at finite instances of time.

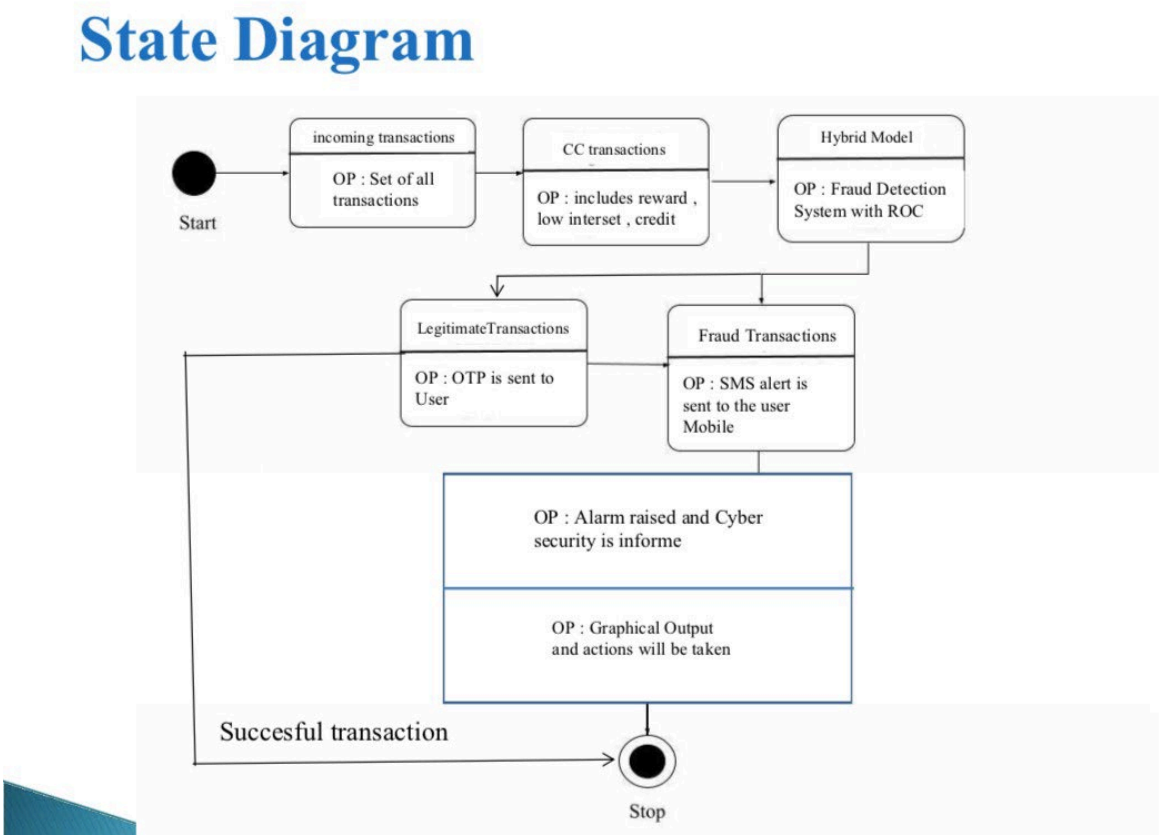

Fig 3.4. State Diagram

3.2.5 Class Diagram

Base Class:

- User

Sub Classes:

- Input1
- Input2
- Prediction
- Detection
- Display1
- Display2

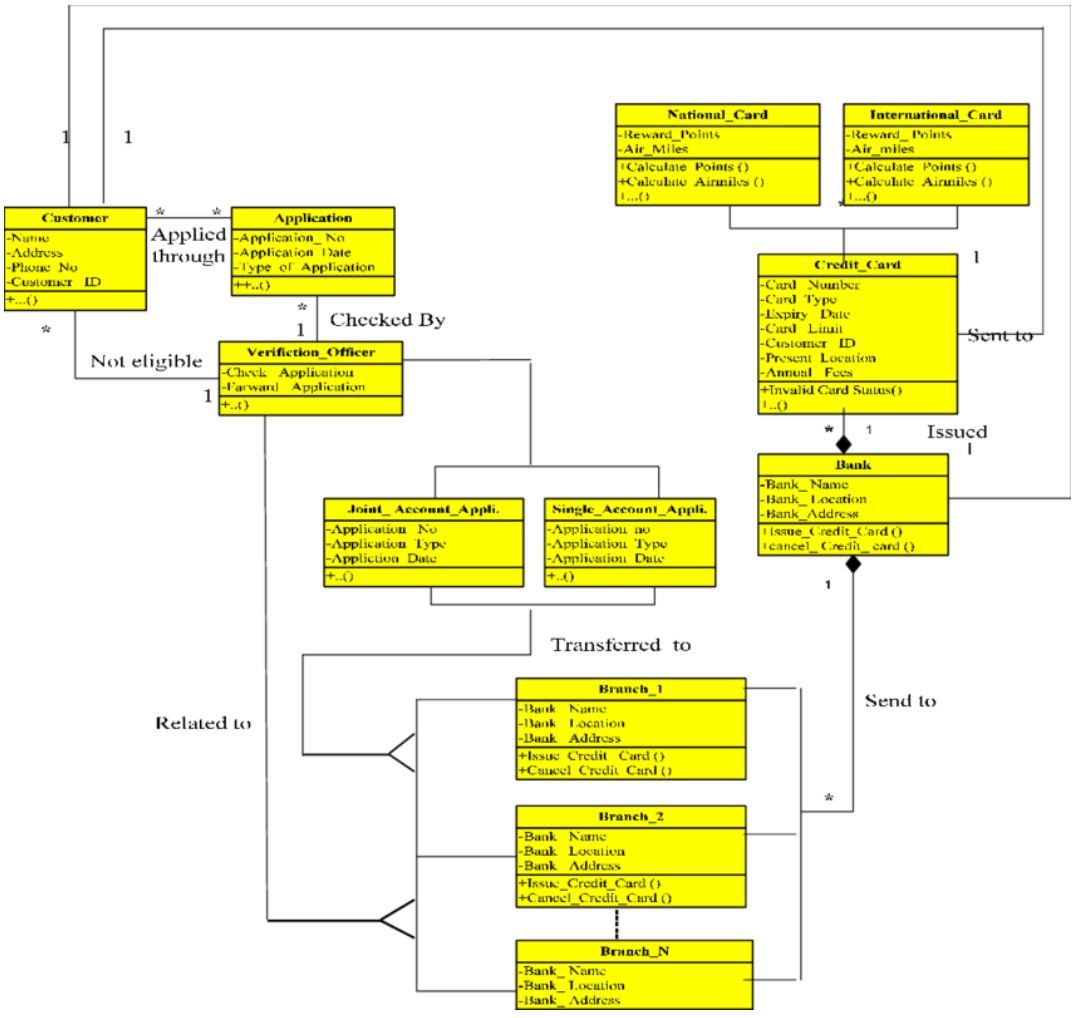

Fig 3.5. Class Diagram

3.3 Software Requirements Specification

3.3.1 Functional Requirements

- **Policy**: Should be able to implement the policy which is the core element of the RL as it alone can define the behavior of the agent. It should map the perceived states of the environment to the actions taken on those states.
- **Reward Signal:** The environment should be able to send an immediate signal to the learning agent at each state, and this signal is a reward signal. The agent's main objective is to maximize the total number of rewards for good actions.
- **Value Function:** The value functionality should be able to give information about how good the situation and action are and how much reward an agent can be expected. A value function should specify the good state and action for the future.

3.3.2 Non-Functional Requirements

- **Accuracy & Performance**: Learning algorithm accuracy and precision is important, and output is compared to the real true result.
- **Transparency**: It is often not clear how results are derived, causing issues in trust and transparency.
- **Reliability**: Further effort has to be put for reliability in the system, like looking at the reliability of individual ML predictions, focusing on reliability estimation.
- **Testability:** Systematic testing of the outcome of ML systems is necessary. Major focus should be given on applying ML systems to improve software testing strategies.

3.3.3 Domain Requirements

- The user must be able to input the required card details and obtain the output, to check the card is fraud or not.
- and must gather the data of all the fraud as well as non-fraudulent transactions throughout the month

Chapter 4

PROJECT PLANNING

As shown in figure 4.1 and 4.2, different phases involved in the project are mentioned alongwith their timeline.

- Problem analysis: We decided upon the problem to be solved for the project/ decideupon the topic for the project.
- Planning phase: In this phase, we laid out a plan to complete various activities relatedto the project.
- Project literature survey: In this phase, we have surveyed different types of technicalpaper related to the problem which we have picked for our project.
- Study of DL and ML concepts: We tried to learn the deep learning concepts requiredfor the project.
- Collection of datasets: We collected the ultrasound images datasets of Ovarian Images and numerical data from Kaggle.
- System Design and architecture phase: We design the architecture of the proposed system.
- Implementation phase: Here, we implement our proposed system using Python. Afterthe implementation is completed, the next phase is to Test the system.
- The last phase is preparation of the final report, in this phase we provide a final and detailed report of our project

4.1 Gantt Chart

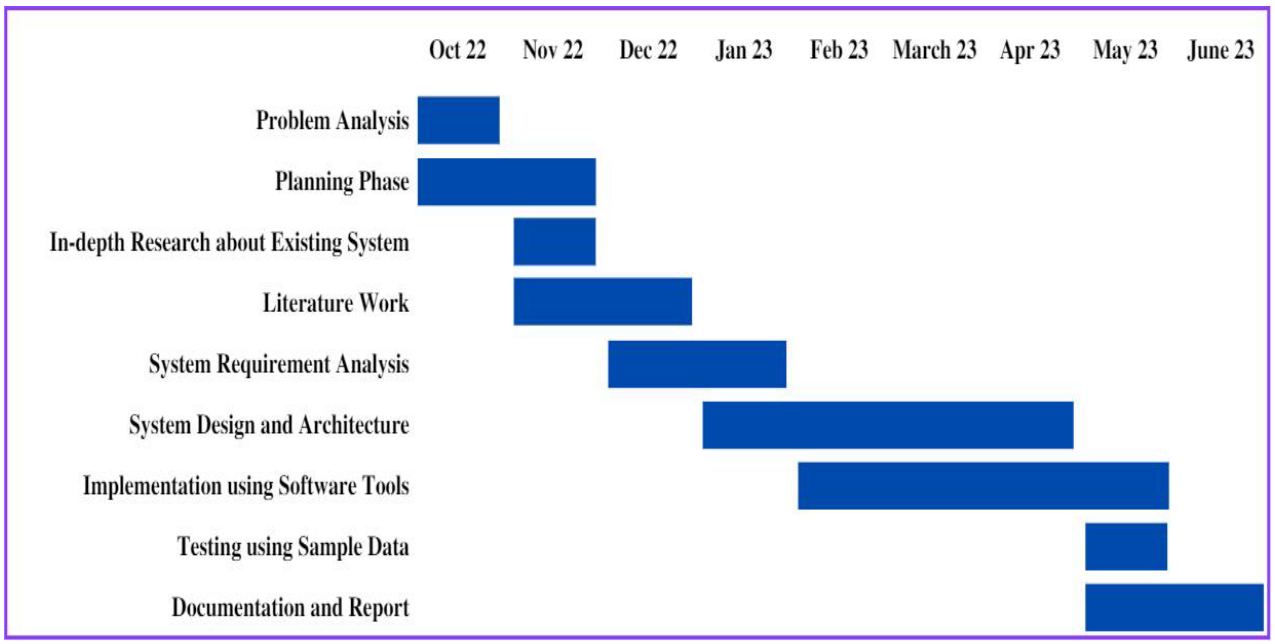

Fig 4.1 Gantt Chart

A Gantt chart is a type of bar chart that illustrates a project schedule. This chart lists the tasks to be performed on the vertical axis, and time intervals on the horizontal axis. The width of the horizontal bars in the graph shows the duration of each activity. In the fig 4.1, the Gantt chart of the current project is represented, where the planning stage started in the month of September 2022 and the is proposed to end by March 2023.

Chapter 5

APPLICATIONS

- Customer Service
- Banking
- Loan and Offerings
- Payment and Gateways
- Merchandise Payments

CONCLUSION

- In this project, Machine Learning techniques like Logistic Regression, Decision Tree and Random Forest were used to detect the fraud in credit card system.
- Sensitivity, Specificity, Accuracy and Error rate are used to evaluate the performance for the proposed system.

Future Enhancements:

- There is a very strong possibility of the system being adopted as a norm for the major banking and financial services applications as fraud detection and prevention is the major checkpoint in financial and banking sector.
- The above system is also likely to be embedded in other applications based, modified as per platform-specific/application specific environment.
- The banks, financial and retail institutes have faced huge losses owing to cause of a robust and accurate system to predict and prevent the fraudulent transactions going on in an institution.
- This in-turn affects the business capabilities and consumer trust of the company. Thus, the organizations have moved their focus onto implementing a system which can depict inconsistent transactions, providing banks a privilege to act upon it take necessary measures.

REFERENCES

[1] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and G. N. Surname, “Random forest for credit card fraud detection” , IEEE15thInternationalConference on Networking, Sensing and Control (ICNSC) , pp.123-345,2022 .

[2] Satvik Vats, Surya Kant Dubey, Naveen Kumar Pandey, “Tool for Effective Detection of Fraud in Credit Card System” , published in International Journal of Communication Network SecurityISSN:2231– 1882, Volume-2, Issue-1, 2022.

[3] Rinky D. Patel and Dheeraj Kumar Singh, “Credit Card Fraud Detection & Prevention of Fraud Using Genetic Algorithm”, published by International Journal of Soft Computing and Engineering(IJSCE)ISSN: 2231-2307, Volume-2, Issue-6, January 2019.

[4] Rinky D. Patel and Dheeraj Kumar Singh, “Credit Card Fraud Detection Prevention of Fraud Using Genetic Algorithm” , published by International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307,Volume-2, Issue-6, January 2019.

[5] Wen-Fang YU, Na Wang, “Research on Credit Card Fraud Detection Model Based on Distance Sum” , published by IEEE International Joint Conference on Artificial Intelligence, pp.243-256 , 2020.

[6] Andreas L. Prodromitids and Salvatore J. Stolfo; "Agent-Based Distributed Learning Applied to Fraud Detection"; Department of Computer Science-Columbia University, pp.145-167 , 2021.

[7] P. K. Chan, W. Fan, A. L. Prodromidis, and S. J. Stolfo, ‘‘Distributed data mining in credit card fraud detection,’’ IEEE Intel. Syst. Appl., pp. 67–74,2022.

[8] Soltani, N., Akbari, M.K., SargolzaeiJavan, M., “An user- based model for credit card fraud detection based on artificial immune system,” Artificial Intelligence and Signal Processing (AISP),16th CSI International Symposium on., IEEE, pp. 029- 033, 2021.

[9] S. Ghosh and D. L. Reilly, “Credit card fraud detection with a neural network”, Proceedings of the 27th Annual Conference on System Science, Volume 3: Information Systems: DSS/ Knowledge Based Systems, pp.621-630,2021.

[10] Masoumeh Zareapoor, seeja.K.R, M.Afshar.Alam, “Analysis of Credit Card Fraud Detection Techniques: based on Certain Design Criteria”, International Journal of Computer Applications , pp. 975 – 8887, Volume52–No.3, 2021.