A Project Stage-II Report
On

" Credit Card Fraud Detection Using Machine Learning"

Submitted to the Savitribai Phule Pune University, Pune
In the Partial Fulfillment of the Requirements for the Award of Degree

of

BACHELOR OF ENGINEERING(Computer Science)

SUBMITTED BY

Pratichi	Exam No:**71726894C**
Rohan Dawkhar	Exam No:**71726782C**
Rahul Powar	Exam No:**71726888J**

Under the Guidance of

Prof. Amolkumar Jadhav
D. Y. Patil School of Engineering Academy

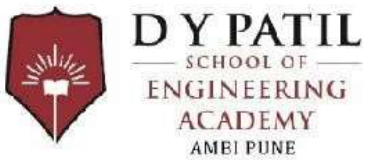
Department of Computer Science Engineering

D.Y. PATIL SCHOOL OF ENGINEERING ACADEMY

AMBI, PUNE

SAVITRIBAI PHULE PUNE UNIVERSITY

2019- 2020



CERTIFICATE

This is to certify that the Project Entitled

"CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING"

Submitted by

Pratichi	Exam No:**71726894C**
Rohan Dawkhar	Exam No:**71726782C**
Rahul Powar	Exam No**:71726888J**

is a bonafide student of this institute and the work has been carried out by him/her under the supervision of **Prof.Amolkumar Jadhav** 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** (Computer Engineering).

.

(Prof. Amolkumar Jadhav)
Guide
Dept. of Computer Engg.

(Prof.Vinod Bharat)
H.O.D
Dept. of Computer Engg.

(Dr. Kherde R V)
Principal
DYPSOEA

Place :

Date :

PROJECT APPROVAL SHEET

"Credit Card Fraud Detection Using Machine Learning"

Is successfully completed by

Pratichi	Exam no.71726894C
Rohan Dawkhar	Exam no.71726782C
Rahul Powar	Exam no.71726888J

At

Department of Computer Engineering

D Y Patil School of Engineering Academy,ambi

SAVITRIBAI PHULE PUNE UNIVERSITY

ACADEMIC YEAR 2019-20

Prof.Amolkumar Jadhav	Prof.Vinod Bharat
Internal Guide	H.O.D
Dept. of Computer Engg.	Dept. of Computer Engg.

Acknowledgment

Every engineering student looks toward the final year project as an opportunity by which he can implement the skill that he has eventually nurtured in the year by hard work dedication the milestone of completing the project would have been intractable without the help of few people who need to be acknowledge.

We owe this moment of satisfactions with a dear sense gratitude to our internal guide Prof. Amolkumar Jadhav who guided us at every stage, whose technical support and helpful attitude give us high moral support. We would also like to extend our sincere thanks to our H.O.D. Prof. Vinod Bharat for his guidance and constant encouragement.

We are highly obliged to the entire state of the Computer Science Engineering department and principal sir for their kind co-operation and help. We also take this opportunity to thank all our colleagues who waked our interest by giving useful suggestions and also possible help. At last but not least we are thankful to our friend colleagues and all the people directly or indirectly concerned with this project.

Thanking You,

Pratichi
Rohan Dawkhar
Rahul Powar

Abstract

Due to rapid growth in the field of cashless transactions or digital transactions, credit cards are widely used in almost every work and hence there are more chances of fraudulent transactions. These fraudulent transactions can be identified by analysing various behaviours of credit card customers from previous transaction history datasets. If any deviation is noticed in the behaviour from the available patterns, there is the possibility of fraudulent transaction. Machine learning techniques are widely used to detect the frauds. In this paper, we have used KNN technique to detect the frauds .The performance of this techniques is evaluated based on the accuracy, precision ,recall.

Contents

List of Figures

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

We will use some algorithmic techniques that are essential for detecting the credit card frauds in the banking field. Fraud detection in credit card is the 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 to credit card fraud detection such as genetic algorithm, artificial neural network frequent item set mining, machine learning algorithms, migrating birds optimization algorithm, comparative analysis of logistic regression, SVM, decision tree and random forest is carried out. We are going to use KNN algorithm and build a KNN based classifier machine to which training and testing data will be provided.

1.2 Motivation of the Project

In general, credit card fraud detection has been known as the process of identifying whether transactions are genuine or fraudulent. In the era of online transactions, this kind of frauds are increasing day by day even after having securities. As the data mining and machine learning techniques are vastly used to counter cyber criminal cases, scholars often embraced those approaches to study and detect credit card fraud activities. In this project, we are trying to develop a software/application which will use some algorithms having good accuracy to detect the frauds as till now there is no real time application has been seen on this. Our main aim is to make this software/application so convenient for the common people or all types of businesses so that they can easily access it and can find themselves safe and secured.

CHAPTER 2

LITERATURE SURVEY

2.1 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 had 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 the 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 to credit card fraud detection such as genetic algorithm, artificial neural network frequent item set mining, machine learning algorithms, migrating birds optimization algorithm, comparative analysis of logistic regression, SVM, decision tree and random forest is carried out. 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. Datasets in previous researches with real data in the literature is nowhere mentioned. Secondly, the improvement of methods is more difficult by the fact that the security concern imposes an 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 the legit transaction in the past may be a fraud in present or vice versa. This paper evaluates four advanced data mining approaches, Decision tree, support vector machines, Logistic regression and random forests and then a collative comparison is made to evaluate that which model performed best. 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

2.2 Literature Survey Papers

1. The Use of Predictive Analytics Technology to Detect Credit Card Fraud in Canada.

Author:- "Kosemani Temitayo Hafiz, Dr.Shaun Aghili, Dr. Pavol Zavarsky."

This research paper focuses on the creation of a scorecard from relevant evaluation criteria, features, and capabilities of predictive analytics vendor solutions currently being used to detect credit card fraud. The scorecard provides a side by side comparison of five credit card predictive analytics vendor solutions adopted in Canada. From the ensuing research findings, a list of credit card fraud PAT vendor solution challenges, risks, and limitations was outlined. All the sub topics should be numbered as shown above. Numbering should be made correctly.

2. BLAST-SSAHA Hybridization for Credit Card Fraud Detection.

Author:- "Amlan Kundu, Suvasini Panigrahi, Shamik Sural, Senior Member, IEEE, and Arun K. Majumdar"

This paper propose to use two-stage sequence alignment in which a profile Analyser (PA) first determines the similarity of an incoming sequence of transactions on a given credit card with the genuine cardholder's past spending sequences. The unusual transactions traced by the profile analyser are next passed on to a deviation analyser (DA) for possible alignment with past fraudulent behaviour. The final decision about the nature of a transaction is taken on the basis of the observations by these two analysers. In order to achieve online response time for both PA and DA, we suggest a new approach for combining two sequence alignment algorithms BLAST and SSAHA.

3. Research on Credit Card Fraud Detection Model Based on Distance Sum.

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

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. It 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. All the sub topics should be numbered as shown above. Numbering should be made correctly.

4. Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models.

Author:- "Navanshu Khare and Saad Yunus Sait"

This paper investigates and checks the performance of Decision tree, Random Forest, SVM and logistic regression on highly skewed credit card fraud data. Dataset of credit card transactions is sourced from European cardholders containing 284,786 transactions. These techniques are applied on the raw and pre-processed data.

5. Credit Card Fraud Detection Using Bayesian and Neural Networks

Author:-“Sam Maes , karl tuyls , Bram vanschoenwinkel and Bernard Manderick”

This paper discuss automated credit card fraud detection by means of machine learning. We apply two techniques suited for reasoning under uncertainity : artificial neural networks and Bayesian belief networks to the problem and show their significant results on real world financial data.

CHAPTER 3

PROBLEM STATEMENT & SCOPE

3.1 Problem statement

Finance fraud is a growing problem with far consequences in the financial industry and while many techniques have been discovered. Major problem is that online payment does not require physical card. Anyone who knows the details of the card can make fraud transactions. Card holder comes to know only after the fraud transaction is carried out.

3.1.1 Goals and objective

We will use some algorithmic techniques that are essential for detecting the credit card frauds in the banking field. Fraud detection in credit card is the 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 to credit card fraud detection such as genetic algorithm, artificial neural network frequent item set mining, machine learning algorithms, migrating birds optimization algorithm, comparative analysis of logistic regression, SVM, decision tree and random forest is carried out.

3.1.2 Purpose and Scope of Document

We describe what features are in the scope and what are not in the scope of the system to be developed. Our project is like a component which can be used on different ways in future. Component can be implemented in banking management system. This project can be used as inbuilt component in Banking Management System. It is standalone application in which we are trying to implement Classification algorithms such as random forest algorithm, Support Vector Machine etc.

3.1.3 Overview of responsibilities of Developer

1. Adaptive Project Framework -

   In this methodology, the project scope is a variable. Additionally, the time and the cost are constants for the project. Therefore, during the project

execution, the project scope is adjusted in order to get the maximum business value from the project.

2. Dynamic Systems Development Model (DSDM) AAA This is the successor of Rapid Application Development (RAD) methodology. This is also a subset of agile software development methodology and boasts about the training and documents support this methodology has. This method emphasizes more on the active user involvement during the project life cycle.

3. Extreme Programming (XP) -

   Lowering the cost of requirement changes is the main objective of ex-treme programming.XP emphasizes on ne scale feedback, continuous process, shared understand- ing and programmer welfare.In XP, there is no detailed requirements speci cation or software architecture built.

4. Feature Driven Development (FDD) - This methodology is more fo-cused on simple and well de ned processes, short iterative and feature driven delivery cycles.All the planning and execution in this project type take place based on the features.

5. Information Technology Infrastructure Library (ITIL) -

   This methodology is a collection of best practices in project management.ITIL covers a broad aspect of project management which starts from the organizational man- agement level.

6. Joint Application Development (JAD) -

Involving the client from the early stages with the project tasks is em-phasized by this methodology.The project team and the client hold JAD sessions collaboratively in order to get the contribution from the client.These JAD sessions take place during the entire project life cycle.

7. Lean Development (LD) -

Lean development focuses on developing change tolerance software.In this method, satisfying the customer comes as the highest priority.The team is motivated to provide the highest value for the money paid by the custumer.

9. Systems Development Life Cycle (SDLC) -

This is a conceptual model used in software development projects.In this method, there is a possibility of combining two or more project management methodologies for the best outcome.SDLC also heavily emphasizes on the use of documentation and has strict guidelines on it.

3.2 Usage Scenario

(a) Developer will create project then developer will accept real time data,build and deploy project

(b) End user will use that analyzed data for analysis purpose

3.2.1 User process

There are two users or actors which are as follows,

Developer- Who will create project,will accept real time data,Build project and also Deploy project.
End user-Who will use analyzed data for analysis purpose

CHAPTER 4

HARDWARE & SOFTAWARE REQUIREMENT

4.1 Proposed System

Machine learning is a collection of methods that can automatically identify patterns in data, and then use those patterns to predict future outcomes, or to perform other types of decision making below certain conditions. Machine learning introduces various algorithms, those enable machines to understand the current situations and on the basis of that machines can take appropriate decisions. Machine learning works independently and takes decision at its own. The main two types of machine learning are, supervised learning and unsupervised learning.

In this project, we have used a Kaggle provided dataset of simulated mobile based payment transactions. We analyze this data by categorizing it with respect to different types of transactions it contains. We also perform PCA - Principal Component Analysis - to visualize the variability of data in two dimensional spaces. The datasets contain transactions made by credit cards in September 2013 by European cardholders. These dataset present transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions. It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example dependent cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise."

4.2 Hardware and software requirement

Hardware :

High end GPU system for processing data

Minimum processor: i5

Minimum RAM : 8GB

Minimum Hard disk: 120 GB

Software :

Operating System: Windows

Jupyter Notebook

Programming Language: Python

4.3 Technologies Used

Python Programming Language :

Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. It was created by Guido van Rossum during 1985- 1990. Like Perl, Python source code is also available under the GNU General Public License (GPL). This tutorial gives enough understanding on Python programming language.

Python is a MUST for students and working professionals to become a great Software Engineer specially when they are working in Web Development Domain. I will list down some of the key advantages of learning Python:

- Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- Python is Interactive − You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
- Python is Object-Oriented − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
- Python is a Beginner's Language − Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

Characteristics of Python :

Following are important characteristics of Python Programming −

- It supports functional and structured programming methods as well as OOP.
- It can be used as a scripting language or can be compiled to byte-code for building large applications.
- It provides very high-level dynamic data types and supports dynamic type checking.
- It supports automatic garbage collection.
- It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

METHODOLOGIES/ALGORITHM DETAILS :

k-nearest neighbours algorithm :

In pattern recognition, the k-Nearest Neighbours algorithm (or k-NN for short) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression:

In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If then the object is simply assigned to the class of that single nearest neighbor. In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbours.

K-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The k-NN algorithm is among the simplest of all machine learning algorithms.

Algorithm:

The training examples are vectors in a multidimensional feature space, each with a class label.

The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples.

In the classification phase, k is a user-defined constant,

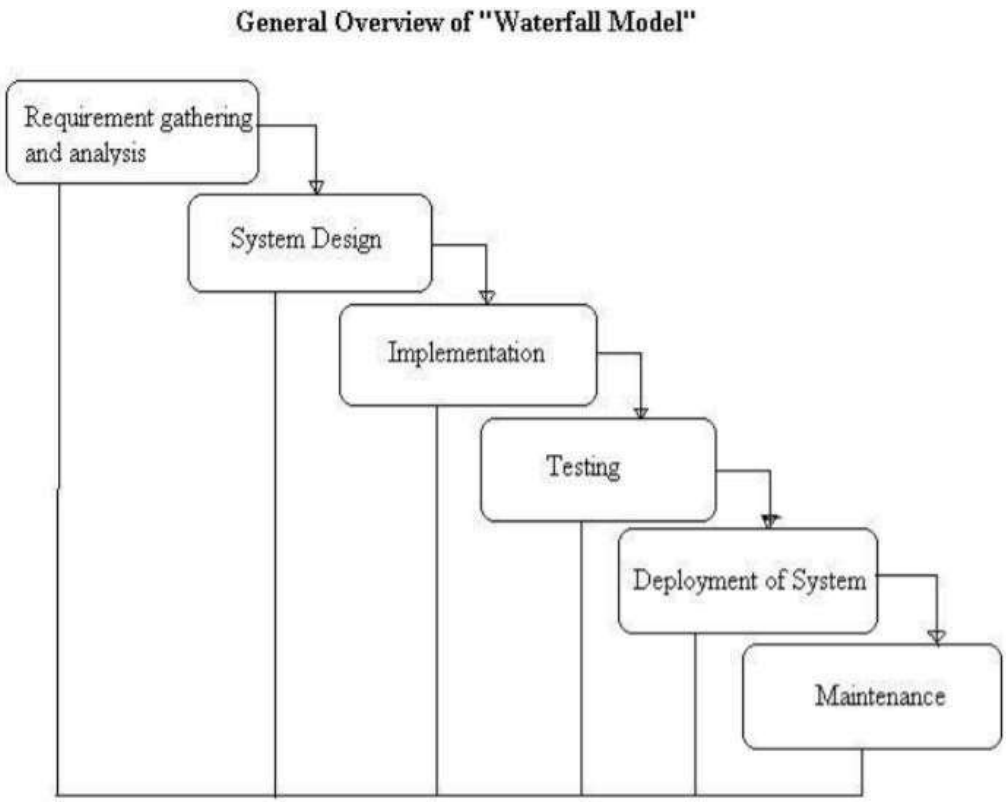
It is an unlabeled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point.

CHAPTER 5

PLANNING AND ESTIMATION

5.1 Software development Life Cycle

The entire project spanned for duration of 6 months. In order to effectively design and develop a cost effective model the Waterfall model was practice.



5.1.1 Requirement gathering and Analysis phase:

This phase started at the beginning of our project, we had formed groups and modularized the project. Important points of consideration were

1. Define and visualize all the objectives clearly.

2. Gather requirements and evaluate them Consider the technical requirements needed and then collect tech nical specifications of various peripheral components (Hardware) required.

3. Analyze the coding languages needed for the project

4. Define coding strategies.

5. Analyze future risks / problems.

6. Define strategies to avoid this risks else define altern ate solutions to this risks.

7. Check financial feasibility.

8. Define Gantt charts and assign time span for each phase. By studying the project extensively we developed a Gantt chart to track and schedule the project. Below is the Gantt chart of our project.

5.1.2 Timeline chart

Task Name	1 D	Start date	Finish date	duration	30/07/2019 To 19/08/2019	19/08/ To 26/08/19	27/08/2019 To 23/09/2019	24/08/2019 To 07/10/2019	08/10 To 15/10	08/10 To 15/10	08/10 To 15/10
Require-ment engineering ng	1	29/07/19	19/08/19	3 Weeks	⟺						
Problem Definition	2	12/08/19	26/08/19	1 Week		⟺					
Literature Survey	3	19/08/19	02/09/19	4 Weeks			⟺				
Analysis	4	02/09/19	02/09/19	2 Week				⟺			
Flow Chart	5	16/09/19	02/09/19	1 Week					⟺		
Block Diagram	6	30/09/19	07/10/19	2 weeks				⟺			
H/ W Specification	7	07/10/19	07/10/19	1 week						⟺	
S/W Specification	8	07/10/19	07/10/19	1 week							⟺

5.2 FEASIBILITY STUDY

Feasibility study is made to see if the project on completion will serve the purpose of the organization for the amount of work, effort and the time that spend on it. Feasibility study lets the developer foresee the future of the project and the usefulness. A feasibility study of a system proposal is according to its workability, which is the impact on the organization, ability to meet their user needs and effective use of resources. Thus when a new application is proposed it normally goes through a feasibility study before it is approved for development. The document provide the feasibility of the project that is being designed and lists various areas that were considered very carefully during the feasibility study of this project such as Technical, Economic and Operational feasibilities.

The following are its features:

5.2.1 TECHNICAL FEASIBILITY

The system must be evaluated from the technical point of view first. The assessment of this feasibility must be based on an outline design of the system requirement in the terms of input, output, programs and procedures. Having identified an outline system, the investigation must go on to suggest the type of equipment, required method developing the system, of running the system once it has been designed.

- Technical issues raised during the investigation are:
- Does the existing technology sufficient for the suggested one?
- Can the system expand if developed?

The project should be developed such that the necessary functions and performance are achieved within the constraints. The project is developed within latest technology. Through the technology may become obsolete after some period of time, due to the fact that never version of same software supports older versions, the system may still be used. So there are minimal constraints involved with this project. The system has been developed using Java the project is technically feasible for development.

5.2.2 ECONOMIC FEASIBILITY

The developing system must be justified by cost and benefit. Criteria to ensure that effort is concentrated on project, which will give best, return at the earliest. One of the factors, which

affect the development of a new system, is the cost it would require. The following are some of the important financial questions asked during preliminary investigation:

- The costs conduct a full system investigation.
- The cost of the hardware and software.
- The benefits in the form of reduced costs or fewer costly errors.

Since the system is developed as part of project work, there is no manual cost to spend for the proposed system. Also all the resources are already available, it give an indication of the system is economically possible for development.

5.2.3 BEHAVIORAL FEASIBILITY

This includes the following questions:

- Is there sufficient support for the users?
- Will the proposed system cause harm?

The project would be beneficial because it satisfies the objectives when developed and installed. All behavioral aspects are considered carefully and conclude that the project is behaviorally feasible.

5.2.4 RISK ANALYSIS PROCESS

Regardless of the prevention techniques employed, possible threats that could arise inside or outside the organization need to be assessed. Although the exact nature of potential disasters or their resulting consequences are difficult to determine, it is beneficial to perform a comprehensive risk assessment of all threats that can realistically occur to the organization. Regardless of the type of threat, the goals of business recovery planning are to ensure the safety of customers, employees and other personnel during and following a disaster. The relative probability of a disaster occurring should be determined. Items to consider in determining the probability of a specific disaster should include, but not be limited to: geographic location, topography of the area, proximity to major sources of power, bodies of water and airports, degree of accessibility to facilities within the organization, history of local utility companies in providing uninterrupted services, history of the area‘s susceptibility to natural threats, proximity to major highways which transport hazardous waste and combustible products. Potential exposures may be classified as natural, technical,

or human threats.

Examples include:

- Natural Threats: internal flooding, external flooding, internal fire, external fire, seismic activity, high winds, snow and ice storms, volcanic eruption, tornado, hurricane, epidemic, tidal wave, typhoon.

Technical Threats:

power failure/fluctuation, heating, ventilation or air conditioning failure, malfunction or failure of CPU, failure of system software, failure of application software, telecommunications failure, gas leaks, communications failure, nuclear fallout.

- Human Threats: robbery, bomb threats, embezzlement, extortion, burglary, vandalism, terrorism, civil disorder, chemical spill, sabotage, explosion, war, biological contamination, radiation contamination, hazardous waste, vehicle crash, airport proximity, work stoppage (Internal/External), computer crime.

All locations and facilities should be included in the risk analysis. Rather than attempting to determine exact probabilities of each disaster, a general relational rating system of high, medium and low can be used initially to identify the probability of the threat occurring. The risk analysis also should determine the impact of each type of potential threat on various functions or departments within the organization. A Risk Analysis Form, found here(PDF Format), can facilitate the process. The functions or departments will vary by type of organization. The planning process should identify and measure the likelihood of all potential risks and the impact on the organization if that threat occurred.

To do this, each department should be analyzed separately. Although the main computer system may be the single greatest risk, it is not the only important concern. Even in the most automated organizations, some departments may not be computerized or automated at all. In fully automated departments, important records remain outside the system, such as legal files, PC data, software stored on diskettes, or supporting documentation for data entry. The impact can be rated as: 0= No impact or interruption in operations, 1= Noticeable impact, interruption in operations for up to 8 hours, 2= Damage to equipment and/or facilities, interruption in operations for 8 - 48 hours, 3= Major damage to the equipment and/or facilities, interruption in operations for more than 48 hours. All main office and/or computer center

functions must be relocated. Certain assumptions may be necessary to uniformly apply ratings to each potential threat.

Following are typical assumptions that can be used during the risk assessment process:

1. Although impact ratings could range between 1 and 3 for any facility given a specific set of circumstances, ratings applied should reflect anticipated, likely or expected impact on each area.

2. Each potential threat should be assumed to be ―localized‖ to the facility being rated.

3. Although one potential threat could lead to another potential threat (e.g., a hurricane could spawn tornados), no domino effect should be assumed.

4. If the result of the threat would not warrant movement to an alternate site(s), the impact should be rated no higher than a ―2.‖

5. The risk assessment should be performed by facility. To measure the potential risks, a weighted point rating system can be used.

Functional requirement

In software engineering, a functional requirement defines a function of a software system or its component. A function is described as a set of inputs, the behavior, and outputs (see also software). Functional requirements may be calculations, technical details, data manipulation and processing and other specific functionality that define what a system is supposed to accomplish. Behavioral requirements describing all the cases where the system uses the functional requirements are captured in use cases.

Functional requirements are supported by non-functional requirements (also known as quality requirements), which impose constraints on the design or implementation (such as performance requirements, security, or reliability). Generally, functional requirements are expressed in the form "system must do ", while non-functional requirements are "system shall be ". The plan for implementing functional requirements is detailed in the system design. The plan for implementing nonfunctional requirements is detailed in the system architecture.

As defined in requirements engineering, functional requirements specify particular results of a system. This should be contrasted with non-functional requirements which specify overall characteristics such as cost and reliability. Functional requirements drive the application architecture of a system, while non-functional requirements drive the technical architecture of a system.

5.2 Non-functional requirement

In systems engineering and requirements engineering, a non-functional requirement is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviors. Thisshould be contrasted with functional requirements that define specific behavior or functions. The plan for implementing functional requirements is detailed in the system design. The plan for implementing nonfunctional requirements is detailed in the system architecture.

In general, functional requirements define what a system is supposed to do whereas non-functional requirements define how a system is supposed to be. Functional requirements are usually in the form of "system shall do ", while non-functional requirements are "system shall be ".

Non-functional requirements are often called qualities of a system. Other terms for non-functional requirements are "constraints", "quality attributes", "quality goals", "quality ofservice requirements" and "nonbehavioral requirements".

CHAPTER 6

TESTING

Testing

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decisionbranches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

- A unit is the smallest testable part of an application like functions, classes, procedures, interfaces. Unit testing is a method by which individual units of source code are tested to determine if they are fit for use.
- Unit tests are basically written and executed by software developers to make sure that code meets its design and requirements and behaves as expected.
- The goal of unit testing is to segregate each part of the program and test that the individual parts are working correctly.
- This means that for any function or procedure when a set of inputs are given then it should return the proper values. It should handle the failures gracefully during the course of execution when any invalid input is given.

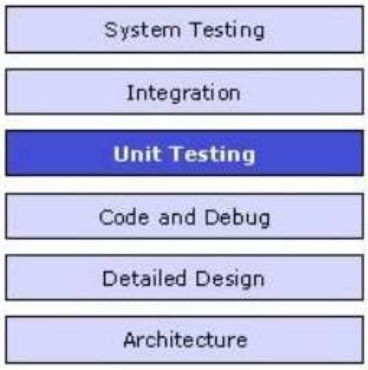


Fig 16.Unit Testing

Integration Testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

- Integration testing tests integration or interfaces between components, interactions to different parts of the system such as an operating system, file system and hardware or interfaces between systems.
- Also after integrating two different components together we do the integration testing. As displayed in the image below when two different modules Module A and Module B are integrated then the integration testing is done.

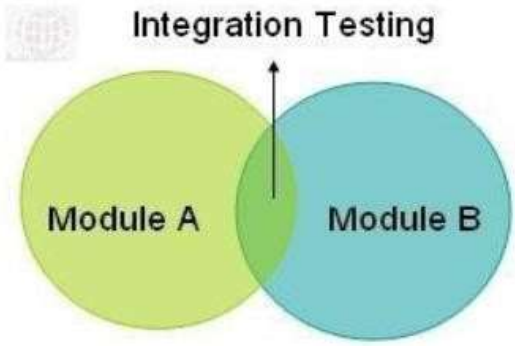


Fig 17.Integration Testing

Test Case1

Test Case	Data File
Objective	Read Data File
Expected Result	Read Data File Successfully

Test Case2

Test Case	Outliers Removal
Objective	Remove Outliers from all the volumes
Expected Result	Outliers removed successfully

Test Case 3

Test Case	Apply KNN
Objective	Find out the best value of k
Expected Result	k value found successfully

Test Case 4

Test Case	Performance Value
Objective	Report the value of performance
Expected Result	Recall score of machine generated successfully

CHAPTER 7

SYSTEM DESIGN

5.1 Architectural Design

For years, fraud has been a major issue in sectors like banking, medical, insurance, and many others. Due to the increase in online transactions through different payment options, such as credit/debit cards, PhonePe, Gpay, Paytm, etc., fraudulent activities have also increased. Moreover, fraudsters or criminals have become very skilled in finding escapes so that they can loot more. Since no system is perfect and there is always a loophole them, it has become a challenging task to make a secure system for authentication and preventing customers from fraud. So, Fraud detection algorithms are very useful for preventing frauds. Credit card fraud detection, which is a data mining problem, becomes challenging due to two major reasons - first, the profiles of normal and fraudulent behaviours change constantly and secondly, credit card fraud data sets are highly skewed. The performance of fraud detection in credit card transactions is greatly affected by the sampling approach on dataset, selection of variables and detection techniques used.

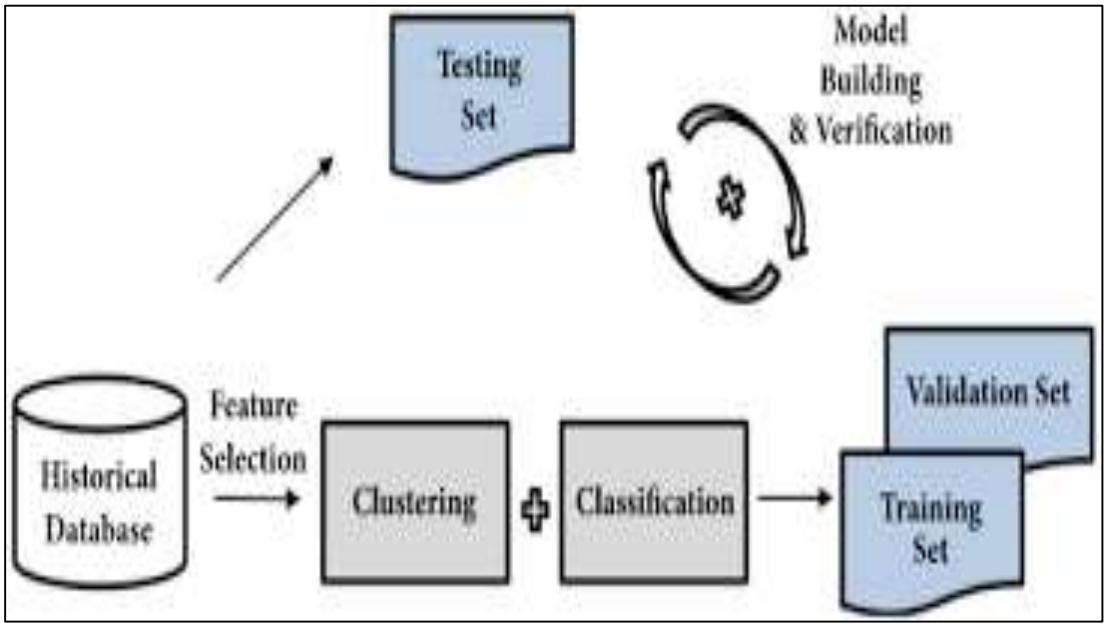


Figure 5.1: Architecture diagram

5.1.1 Data-flow Diagrams

In Data Flow Diagram we Show that ow 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 out-put of system .And the DFD 2 shows the detailed ow of the proposed system.

DFD-0

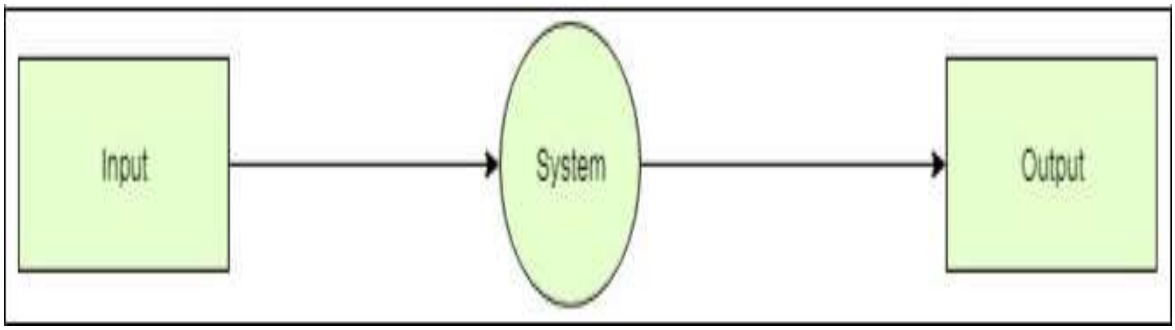


Figure 5.2: DFD-0 for fraud detection

DFD-1

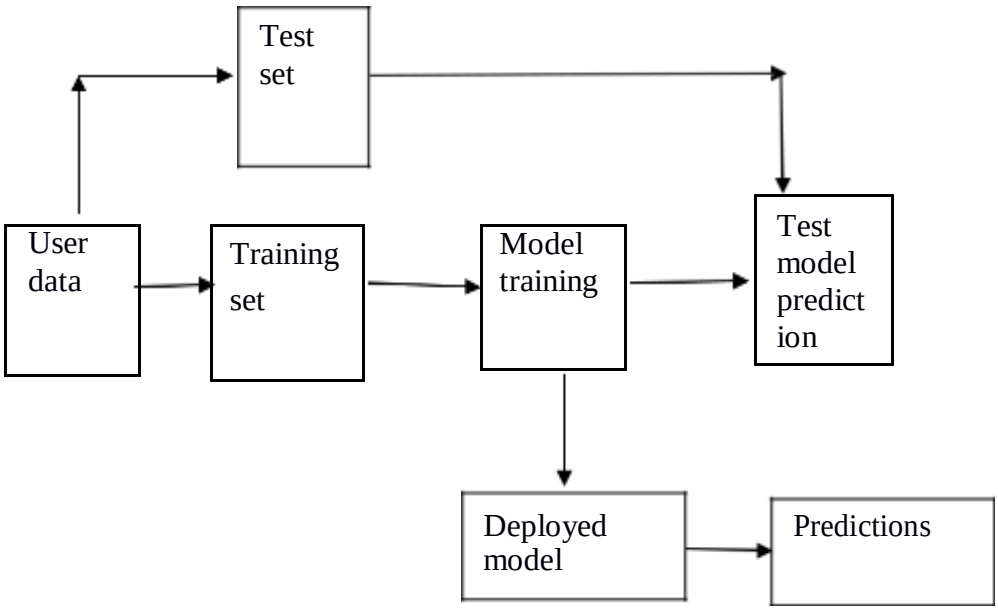


Figure 5.2: DFD-1 for fraud detection

DFD-2

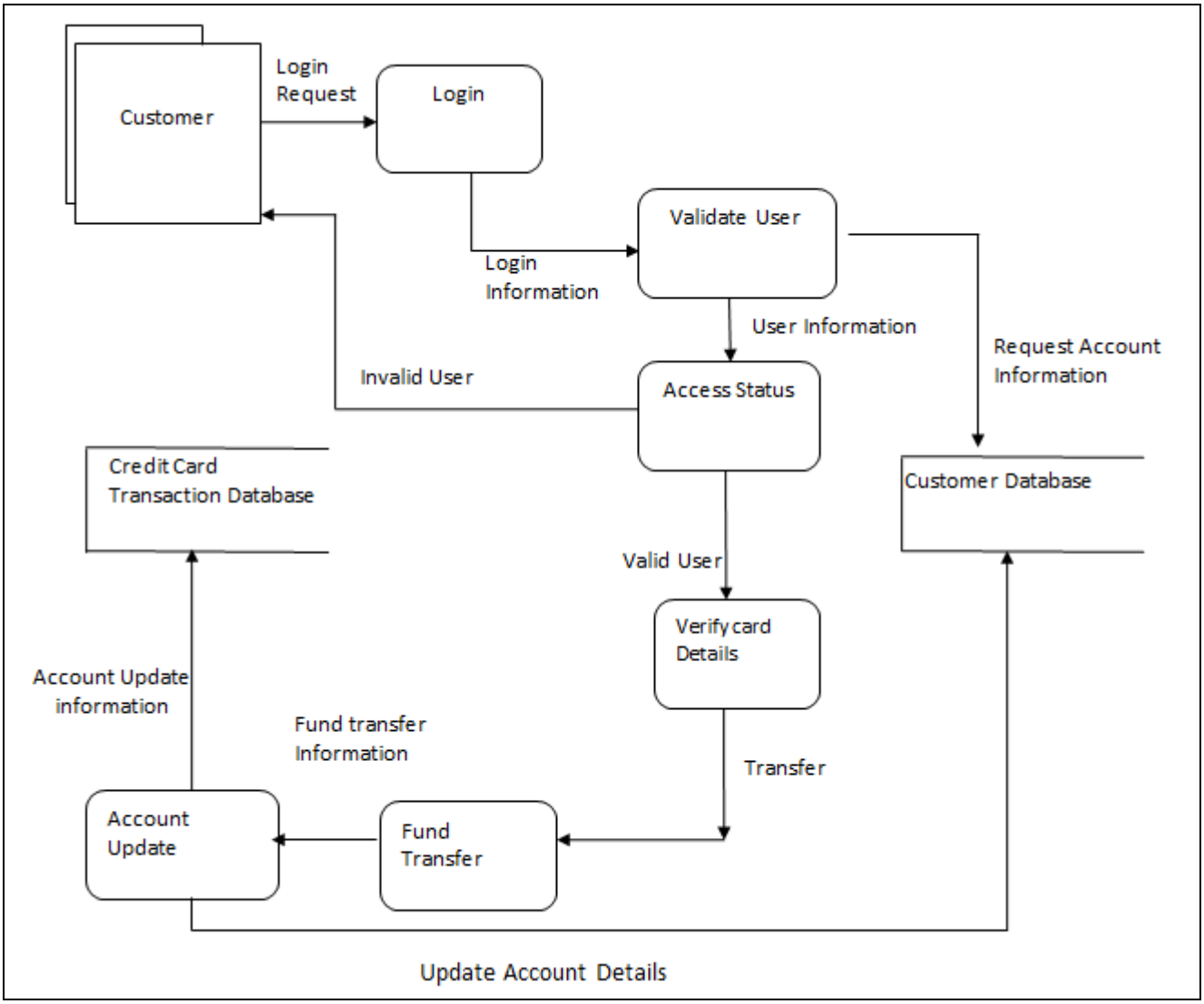


Figure 5.4: DFD-2 for fraud detection

5.1.2 State Chart diagram

A state diagram shows the behavior of classes in response to external stimuli. Speciffically a state diagram describes the behavior of a single object in response to a series of events in a system. Sometimes it's also known as a Harel state chart or a state machine diagram. This UML diagram models the dynamic ow of control from state to state of a particular object within a system.

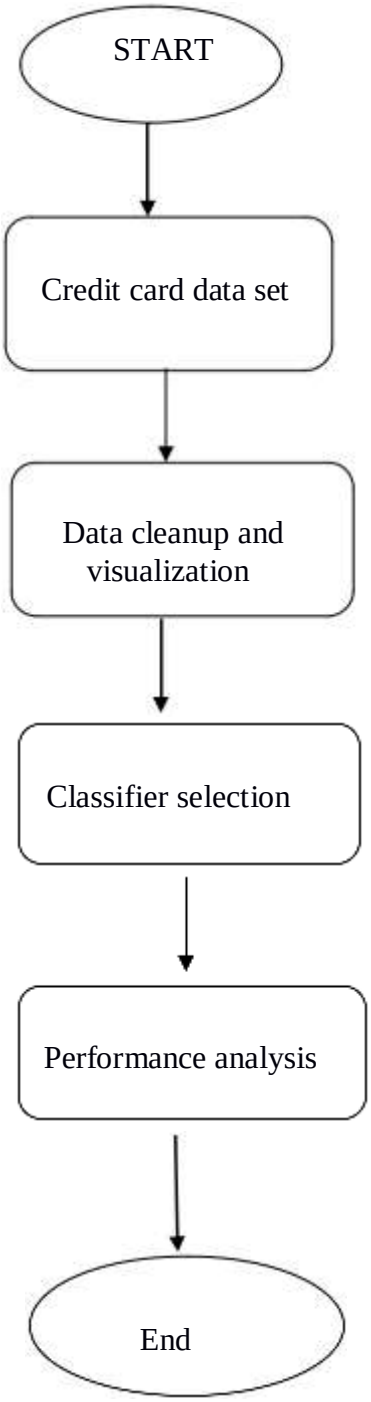


Figure 5.5: State Chart diagram for fraud detection

5.1.3 Use-case diagram

A use case diagram is a dynamic or behavior diagram in UML. Use case diagrams model the functionality of a system using actors and use cases. Use cases are a set of actions, services, and functions that the system needs to perform.

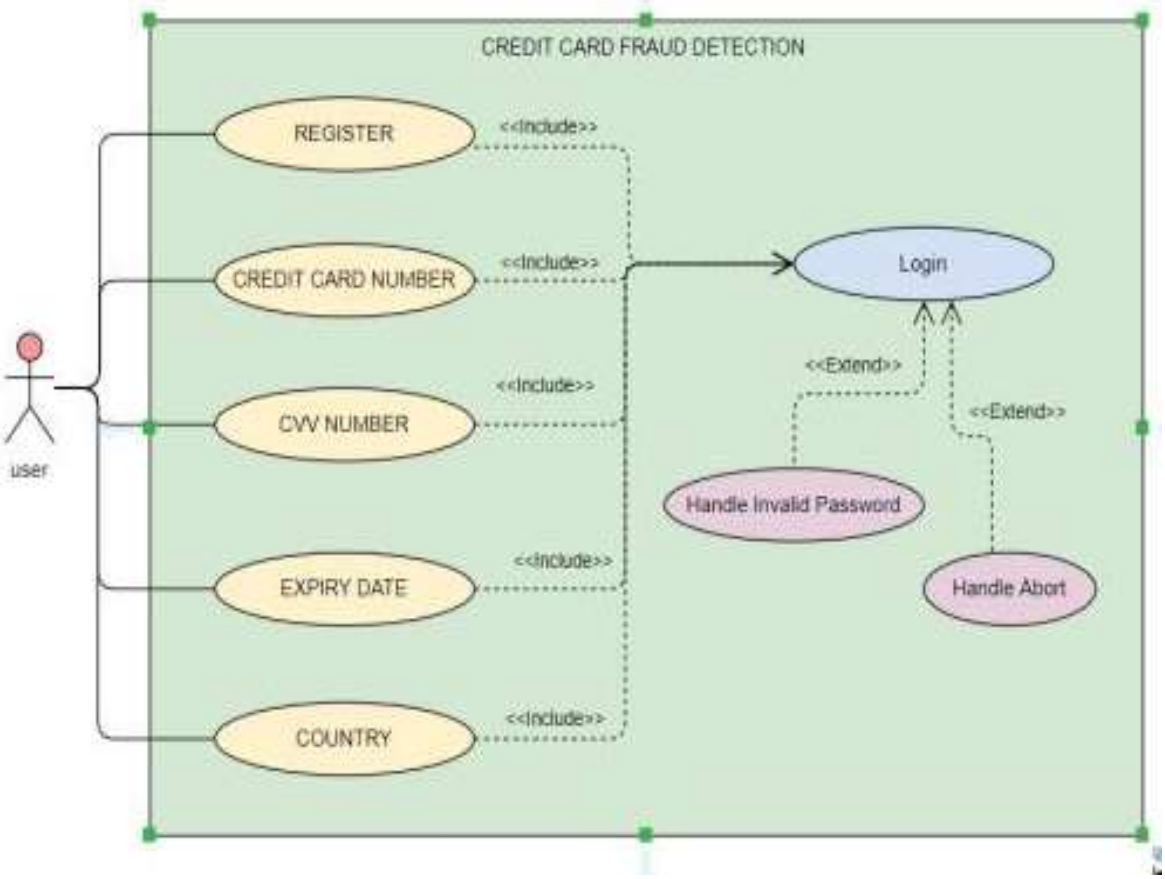


Figure 5.6: Use-case diagram for fraud detection

5.1.4 Activity Diagram

Activity diagrams are graphical representations of workrows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams are intended to model both computational and organizational processes (i.e. workrows). Activity diagrams show the overall flow of control

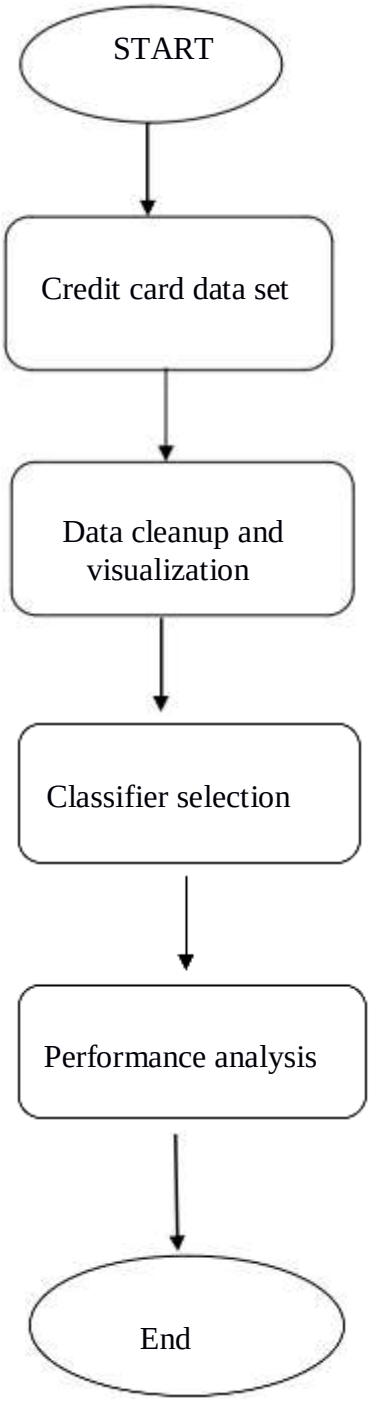


Figure 5.7: Activity Diagram for fraud detection

5.2 Non Functional Requirements:

Interface Requirements

Our System GUI Should be Easy for operating to data analysts/users.

Performance Requirements

Our system should work properly on large no of users access, and signals.

Software quality attributes such as availability

Data collection and data cleaning.

Removing outliers.

Find the recall score.

5.2.1 Design Constraints

We use Google collab for boxplots of our project.

5.2.2 Software Interface Description

Language: Python

Professional Environment: Google Collab

Database: Google Drive

System Type: 64-bit or 32-bit

Processor: Intel core i5, 2 GHz

Random Access Memory (RAM): 8 GB

Storage Capacity: 1 TB

IO device: mouse and keyboard

Device Name: Laptop or Computer

CHAPTER 8

RESULTS

Calculating value of performance

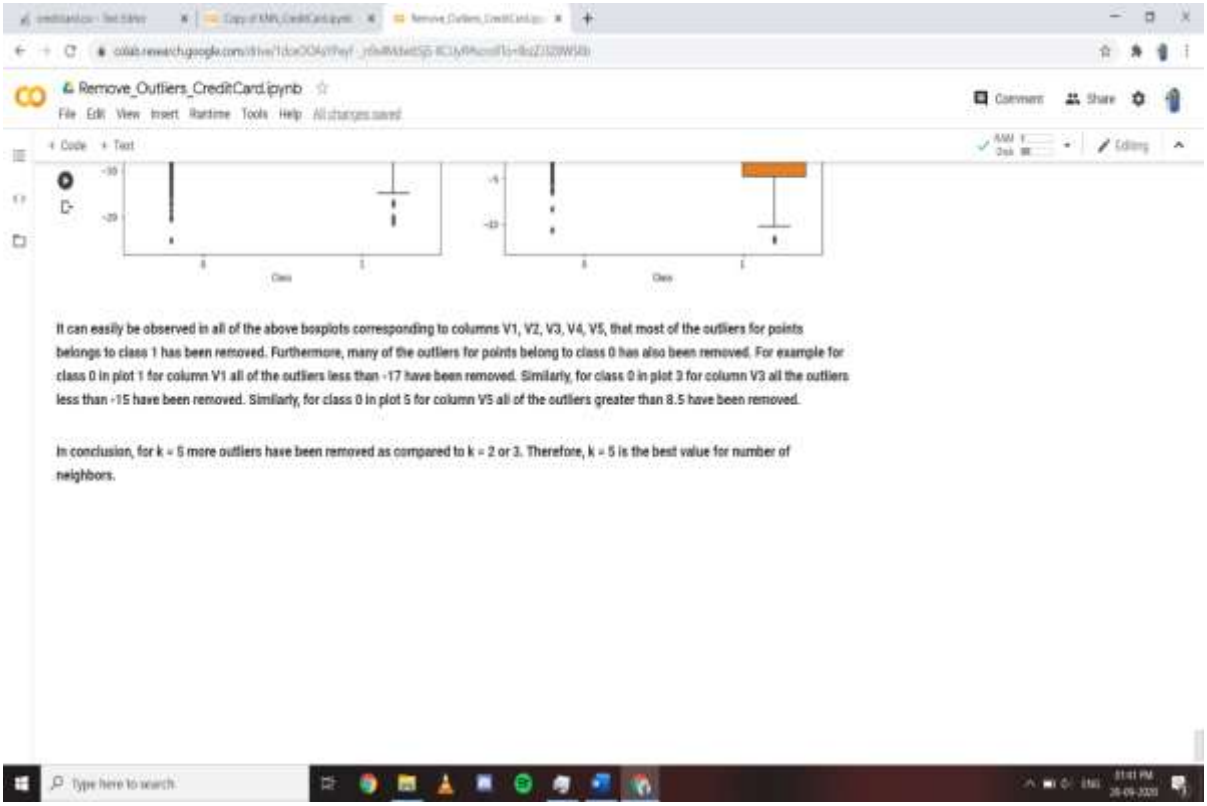


Fig.1 More outliers are removed for k=5 which is the best value of k

Fig.2 Reporting the value of machine's performance

CHAPTER 9

FUTURE MODIFICATIONS

&

CONCLUSION

9.1 Future modification

In the future, we plan to test the prototype in a real life environment, which would include HEIs, students and companies. In this way the presented concept could be further validated. Additionally, we plan to adapt this system to Artificial Intelligence so that each course would be assigned with a unique address and a pool of tokens. After completing the course obligations, students would get tokens from the course address and not directly from the institution. The course address would be a multi signature address between an institution and a professor**.**

9.2 Conclusion-

Fraud detection is a complex issue that requires a substantial amount of planning before throwing machine learning algorithms at it. Nonetheless, it is also an application of data science and machine learning for the good, which makes sure that the customer's money is safe and not easily tampered with.

Future work will include a comprehensive tuning of the Random Forest algorithm I talked about earlier. Having a data set with non-anonymized features would make this particularly interesting as outputting the feature importance would enable one to see what specific factors are most important for detecting fraudulent transactions.

The proposed solution is based on the Machine learning concepts. It transfers the higher education grading system from the current real-world physical records or traditional digital ones (e.g. databases) to an efficient, simplified, ubiquitous version, based on A.I technology. It is anticipated that such a system could potentially evolve into a unified, simplified and globally ubiquitous higher education credit and grading system.

CHAPTER 10

REFERENCES

1 WorldPay. (2015, Nov). Global payments report preview: your definitive guide to the world of online payments. Retrieved September 28, 2016, from http://offers.worldpayglobal.com/rs/850JOA856/images/GlobalPaymentsReportNov2015.pdf.

2 Federal Trade Commision. (2008). consumer sentinel network - data book for January - December 2008. Retrieved Oct 20, 2016. From https://www.ftc.gov/.

3 Bhatla, T.P., Prabhu, V., and Dua, A. (2003). understanding credit card frauds. Crads Business Review# 2003-1, Tata Consultancy Services.

4 The Nilson Report. (2015). Global fraud losses reach $16.31 Billion. Edition: July 2015, Issue 1068.

5 Y. Sahin and E. Duman, "Detecting credit card fraud by decision trees and support vector machines", Proceedings of the International MultiConference of Engineers and Computer Scientists 2011 Vol I, IMECS 2011, March 2011.

6 Elkan, C. (2001). Magical thinking in data mining: lessons from COIL challenge 2000. Proc. of SIGKDD01, 426-431.

7 Mohammed, J. Zaki., & Wagner, Meira Jr. (2014). Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press. ISBN 978-0-521-76633-3.

8 F. N. Ogwueleka. (2011). Data mining application in credit card fraud detection system. Journal of Engineering Science and Technology, Vol. 6, No. 3 (2011) 311 - 322.

9 V. Bhusari & S. Patil. (2011). Application of hidden markov model in credit card fraud detection. International Journal of Distributed and Parallel Systems (IJDPS) Vol.2, No.6.

10 S.J. Stolfo, D.W. Fan, W. Lee, A.L. Prodromidis, and P.K. Chan. (1998). Credit card fraud detection using meta-learning: issues and initial results, Proc. AAAI Workshop AI Methods in Fraud and Risk Management, pp. 83-90.

11 Sen, Sanjay Kumar., & Dash, Sujatha. (2013). Meta learning algorithms for credit card fraud detection. International Journal of Engineering Research and Development Volume 6, Issue 6, pp. 16-20.

12 Maes, Sam, Tuyls Karl, Vanschoenwinkel Bram & Manderick, Bernard. (2002). Credit card fraud detection using bayesian and neural networks. Proc. of 1st NAISO Congress on Neuro Fuzzy Technologies. Hawana.

13 A.C. Bahnsen, Aleksandar, Stojanovic., D. Aouada & Bjorn, Ottersten. (2013). Cost sensitive credit card fraud detection using bayes minimum risk.

CHAPTER 11

APPENDIX A

Title: Problem statement feasibility assessment using, satis ability analy-sis and NP Hard,NP-Complete or P type using modern algebra and rele-vant mathematical models. a. P-Problem: A problem is assigned to the P(polynomial time) class if there exits at least one algorithm to solve that problem, such that the number of steps of the algorithm is bounded by a polynomial in O(n),where n is the size input.

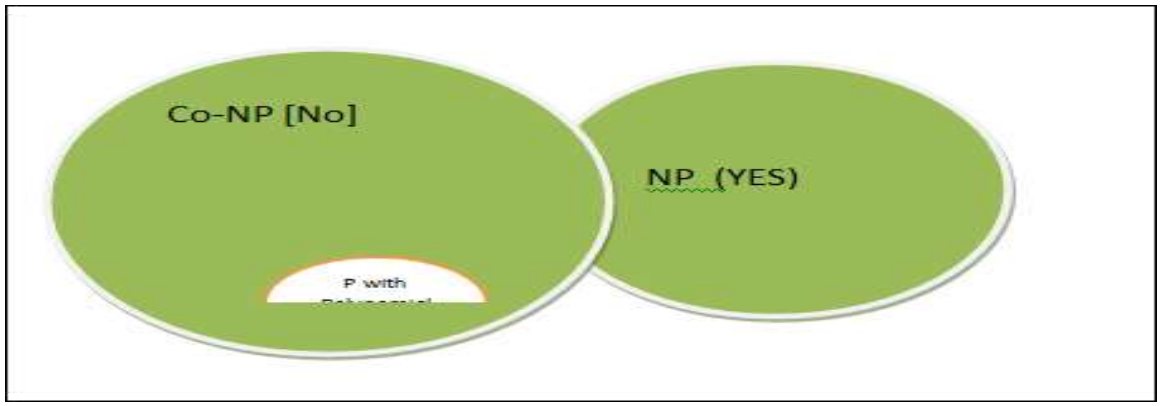


Figure 8.1: NP Complete 1

b. NP-Problem: A problem is assigned to NP (non-deterministic polynomial time) class if it is solvable in polynomial time by a non-deterministic Turing machine.

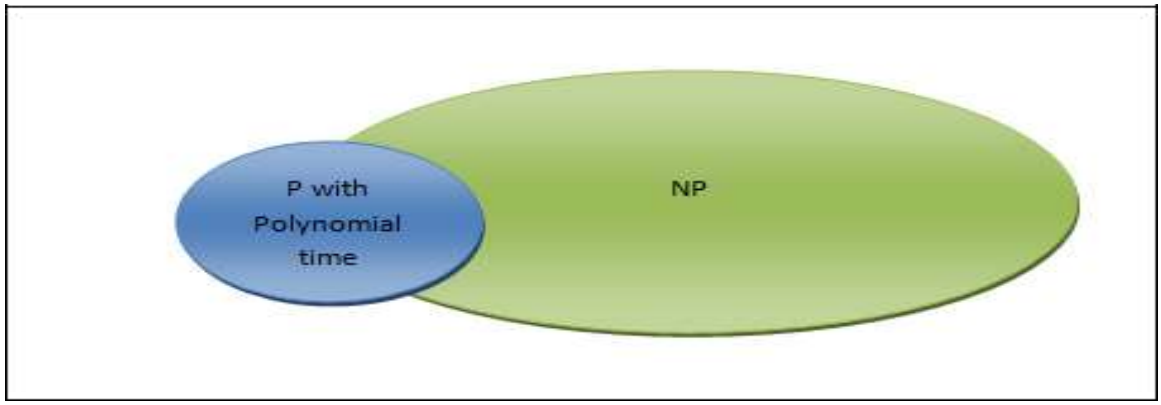
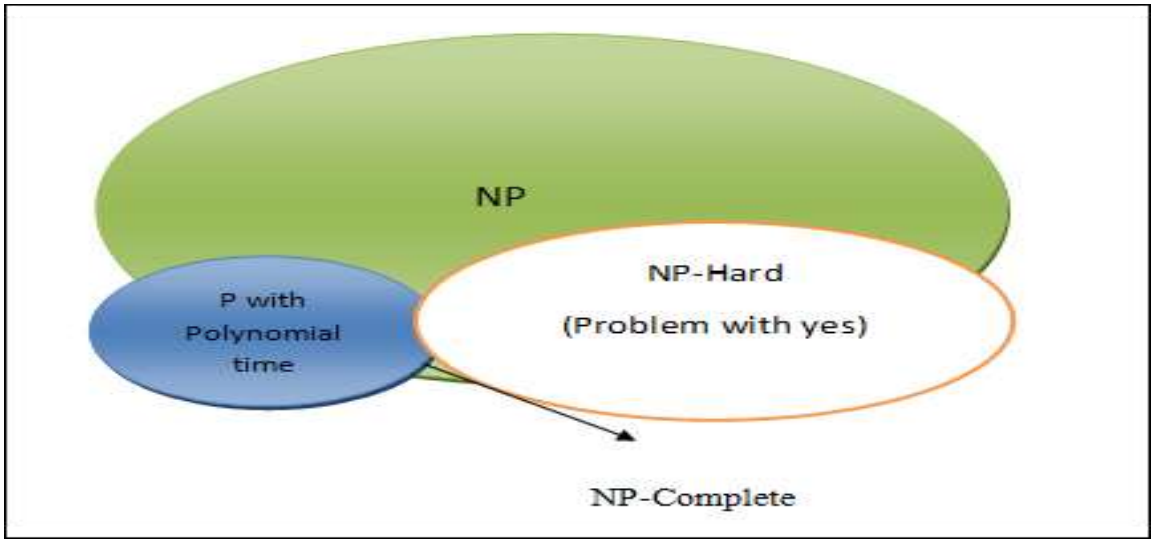


Figure 8.2: NP Complete 2

c. NP-Hard: Problem is said to be NP-Hard if an algorithm for solving it can be translated into one for solving any other NP-problem. It is much easier to show that a problem is NP than to show that it is NP-hard.



d. NP-Complete: A problem which is both NP NP-Hard is called an NP-Complete problem. In this system Binary conversion image segmen-tation is used to result will be NP-Hard.

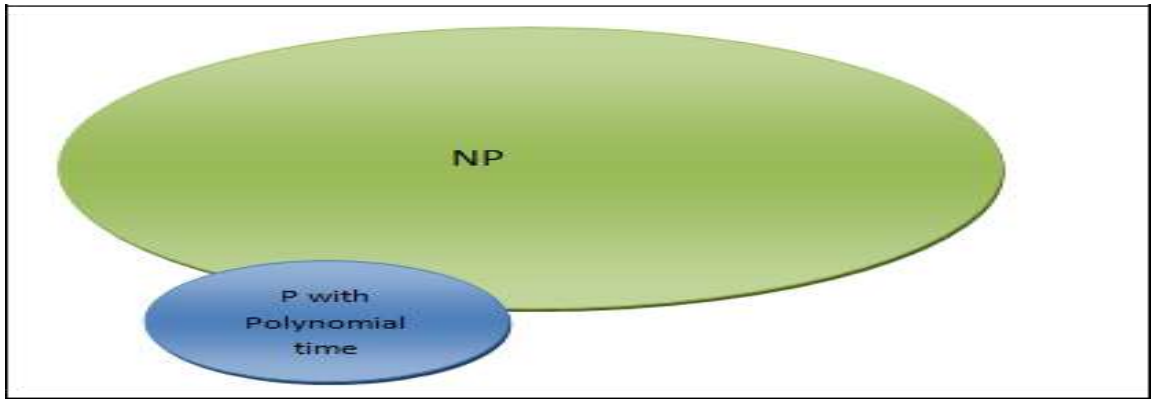


Figure 8.4: NP Complete 4

Conclusion: Hence we implemented Feasibility assessment using NP-Hard, NP-complete in project, stated that our protocols are much strong and can with stand to many of the challenging authentication attacks. Our main focus is to highlight the potential of our approach for real-world de-ployment: whether we can achieve a high level of usability with satisfactory and acceptable results.

CHAPTER 12

APPENDIX B

1 Xiaoyuan Liang, Xusheng Du, Guiling Wang, and Zhu Han, ^aA Deep Q Learning Network for Traffic lights^a Cycle Control in Vehicular Networks^a, IEEE Transactions On Vehicular Technology 0018-9545 (c) 2018.

2 Rusheng Zhang , Akihiro Ishikawa , Wenli Wang , Benjamin Striner y,

and Ozan Tonguz, aIntelligent^ TrafficSignal Control: Using Reinforcement Learning with Partial Detection^a, arXiv:1807.01628v2 4 Feb 2019.

3 Omid Avate pour, Froogh Sadry, aTra^ c Management System Using IoT

Technology - A Comparative Review^a, 978-1-5386-5398-2/18/31.00 A 2018 IEEE 1041 [4] Mohamad Belal Natafgi, Mohamad Osman, Asser Sleiman Haidar Lama Hamandi, ^aSmart Traffic light System Using Machine Learning^a , IMCET 2018.

4 Chi-Man Vong, Pak-Kin Wong, Zi-Qian Ma, Ka-In Wong, ^aApplication of RFID Technology and the Maximum Spanning Tree Algorithm for Solving Vehicle Emissions in Cities on Internet of Things^a, 2014 IEEE World Forum on Internet of Things (WF-IoT).

CHAPTER 13

APPENDIX C

10.1 Plagiarism Report

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	130	Date	December 16,2019
Characters	806	Exclude Url	

0%	100%	0	9
Plagiarism	Unique	Plagiarized Sentences	Unique Sentences

Content Checked For Plagiarism

In general, credit card fraud detection is the process of identifying whether the transactions are genuine or fraudulent. In the era of online transactions, this kind of frauds are increasing day by day even after having securities.As the data mining and machine learning techniques are used to counter cyber-criminal activities, technicians often accept those approaches to study and detect credit card frauds.In this project ,we are trying to develop a software/application which will use some algorithms having good accuracy to detect the frauds as till now there is no real time application has been seen on this. Our main aim is to make this software/application so convenient for the common people or all types of businesses so that they can easily access it and can find themselves safe and secured.

PLAGIARISM SCAN REPORT

Words	105	Date	December 16,2019
Characters	635	Exclude Url	

0%	100%	0	7
Plagiarism	Unique	Plagiarized Sentences	Unique Sentences

Content Checked For Plagiarism

We describe what features are in the scope and what are not in the scope of the system to be developed. Our project is like a component which can be used on different ways in future.Component can be implemented in banking management system. It can be also be used in web based application and as Android App for more generalized purpose. In Scope Project can be used in every transaction monitoring sectors. This project can be used as inbuilt component in Banking Management System. It is standalone application in which we are trying to implement Classification algorithms such as random forest algorithm, Support Vector Machine,etc.