

COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS FOR FRAUD DETECTION IN BLOCKCHAIN

GUIDE:.. AMANDEEP SINGH K, M.E., (PhD).

CHIRUMAMILLA BALARAMA KRISHNA [38110110]

UMMA GOPALA KRISHNA REDDY [38110608]

IAbstract

The use of credit cards is prevalent in modern day society. But it is obvious that the number of credit card fraud cases is constantly increasing in spite of the chip cards worldwide integration and existing protection systems. This is why the problem of fraud detection is very important now. The credit card fraud detection features uses user behavior and location scanning to check for unusual patterns. These patterns include user characteristics such as user spending patterns as well as usual user geographic locations to verify his identity. If any unusual pattern is detected, the system requires revivification.

In this project, a technique for 'Credit Card Fraud Detection' is developed. As fraudsters are increasing day by day. And fallacious transactions are done by the credit card and there are various types of fraud. So to solve this problem combination of technique is used like Genetic Algorithm, Behavior Based Technique and SET (Secure Electronic Transaction), Machine learning, Data Mining. By this transaction is tested individually and whatever suits the best is further proceeded. And the foremost goal is to detect fraud by filtering the above techniques to get better result. In this project the general description of the developed fraud detection system and comparisons between models based

are (pattern recognition). In the last section of this paper the results of evaluative testing and corresponding conclusions are considered. When a Invalid user(fraud) uses a bank transaction , While transaction first bank authoriser check weather the user is valid user or a invalid user.if the user is invalid then the bank authoriser block the transaction.

CHAPTER 1

INTRODUCTION

In recent years, the prevailing data mining concerns people with credit card fraud detection model based on data mining. Since our problem is approached as a classification problem, classical data mining algorithms are not directly applicable. So an alternative approach is made by using general purpose meta heuristic approaches like genetic algorithms.

This project is to propose a credit card fraud detection system using genetic algorithm. Genetic algorithms are evolutionary algorithms which aim at obtaining better solutions as time progresses. When a card is copied or stolen or lost and captured by fraudsters it is usually used until its available limit is depleted. Thus, rather than the number of correctly classified transactions, a solution which minimizes the total available limit on cards subject to fraud is more prominent. It aims in minimizing the false alerts using genetic algorithm where a set of interval valued parameters are optimized.

1.1 PROBLEM DEFINITION

To develop a credit card fraud detection system using genetic algorithm. During the credit card transaction, the fraud is detected and the number of false alert is being minimized by using genetic algorithm. Instead of maximizing the numbers of correctly classified transactions we defined an objective function

where the misclassification costs are variable and thus, correct classification of some transactions are more important than correctly classifying the others.

The algorithm begins with multi-population of randomly generated chromosomes. These chromosomes undergo the operations of selection, crossover and mutation. Crossover combines the information from two parent chromosomes to produce new individuals, exploiting the best of the current generation, while mutation or randomly changing some of the parameters allows exploration into other regions of the solution space. Natural selection via a problem specific cost function assures that only the best fit chromosomes remain in the population to mate and produce the next generation. Upon iteration, the genetic algorithm converges to a global solution.

1.2 PROPOSED ALGORITHMS

Genetic algorithm

Genetic algorithms are evolutionary algorithms which aim at obtaining better solutions as time progresses. Since their first introduction by Holland, they have been successfully applied to many problem domains from astronomy to sports, from optimization to computer science, etc. They have also been used in data mining mainly for variable selection and are mostly coupled with other data mining algorithms. In this study, we try to solve our classification problem by using only a genetic algorithm solution.

Pseudo code of genetic algorithm

Initialize the population

Evaluate initial population

Repeat

Perform competitive selection

Apply genetic operators to generate new solutions

Evaluate solutions in the population

Until some convergence criteria is satisfied.

Selection process

Selection is used for choosing the best individuals, that is, for selecting those chromosomes with higher fitness values. The selection operation takes the current population and produces a 'mating pool' which contains the individuals which are going to reproduce. There are several selection methods, like biased selection, random selection, roulette wheel selection, tournament selection. In this work the following selection mechanisms are used.

Tournament Selection

Tournament selection has been used in this as it selects optimal individuals from diverse groups. It selects t individuals from the current population uniformly at random, forms a tournament and the best individual of a group wins the tournament and is put into the mating pool for recombination. This process is repeated the number of times necessary to achieve the desired size of intermediate population. The tournament size controls the selection strength. The larger the tournament size, the stronger is the selection process.

Elitist Selection

In order to make sure that the best individuals of the solution are passed to further generations, and should not be lost in random selection, this selection operator is used. So we used a few best chromosomes from each generation, based on the higher fitness value and are passed to the next generation of population.

Reproduction

To generate a second generation population of solutions from those selected through genetic operators: crossover (also called recombination), and/or mutation.

For each new solution to be produced, a pair of "parent" solutions is selected for breeding from the pool selected previously. By producing a "child" solution using the above methods of crossover and mutation, a new solution is created which typically shares many of the characteristics of its "parents". New parents are selected for each new child, and the process continues until a new population of solutions of appropriate size is generated. Although reproduction methods that are based on the use of two parents are more "biology inspired", some research suggests more than two "parents" are better to be used to reproduce a good quality chromosome.

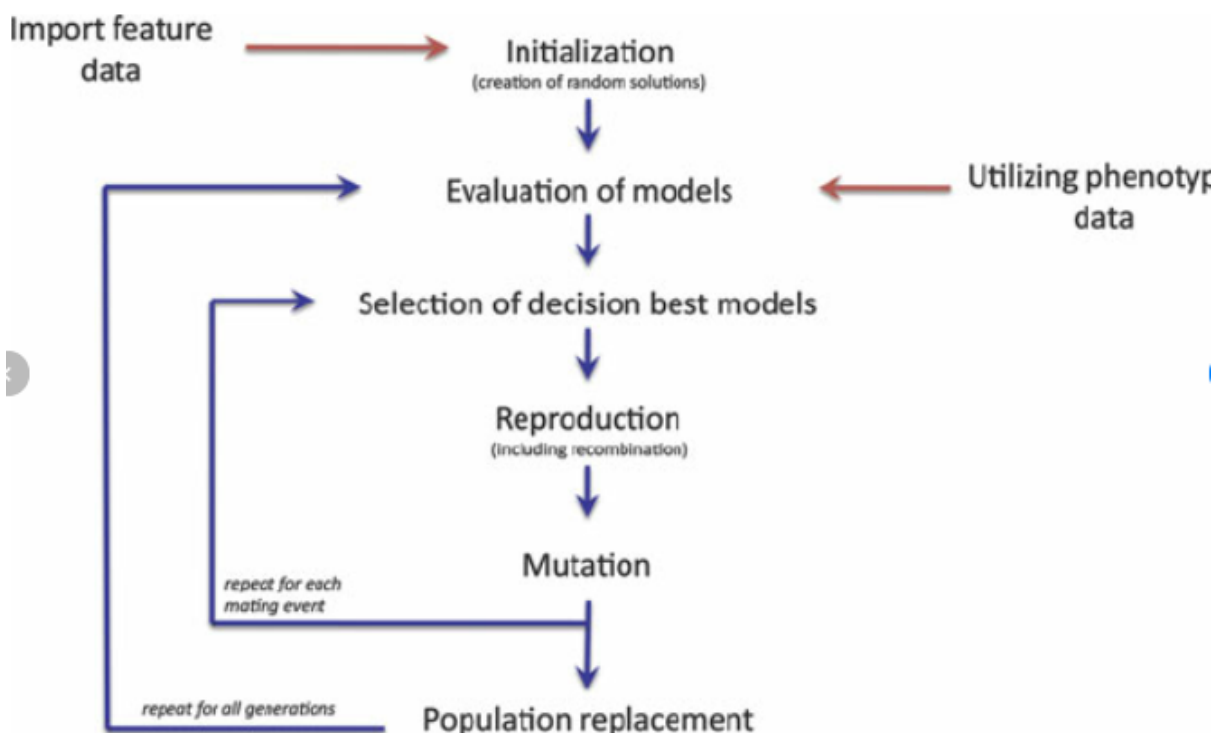
These processes ultimately result in the next generation population of chromosomes that is different from the initial generation. Generally the average fitness will have increased by this procedure for the population, since only the best organisms from the first generation are selected for breeding, along with a small proportion of less fit solutions, for reasons already mentioned above.

Although Crossover and Mutation are known as the main genetic operators, it is possible to use other operators such as regrouping, colonization-extinction, or migration in genetic algorithms.

Termination

This generational process is repeated until a termination condition has been reached. Common terminating conditions are:

- A solution is found that satisfies minimum criteria
- Fixed number of generations reached
- Allocated budget (computation time/money) reached
- The highest ranking solution's fitness is reaching or has reached a plateau such that successive iterations no longer produce better results
- Manual inspection
- Combinations of the above



CHAPTER 2

LITERATURE SURVEY

Fraud detection has been usually seen as a data mining problem where the objective is to correctly classify the transactions as legitimate or fraudulent. For classification problems many performance measures are defined most of which are related with correct number of cases classified correctly.

A more appropriate measure is needed due to the inherent structure of credit card transactions. When a card is copied or stolen or lost and captured by fraudsters it is usually used until its available limit is depleted. Thus, rather than the number of correctly classified transactions, a solution which minimizes the total available limit on cards subject to fraud is more prominent.

Since the fraud detection problem has mostly been defined as a classification problem, in addition to some statistical approaches many data mining algorithms have been proposed to solve it. Among these, decision trees and artificial neural networks are the most popular ones. The study of Bolton and Hand provides a good summary of literature on fraud detection problems.

However, when the problem is approached as a classification problem with variable misclassification costs as discussed above, the classical data mining algorithms are not directly applicable; either some modifications should be made on them or new algorithms developed specifically for this purpose are needed. An alternative approach could be trying to make use of general purpose meta heuristic approaches like genetic algorithms.

Title: Statistical classification methods in consumer credit scoring

Author: D. J. Hand and W. E. Henley

Credit scoring is the term used to describe formal statistical methods used for classifying applicants for credit into 'good' and 'bad' risk classes. Such methods have become increasingly important with the dramatic growth in consumer credit in recent years. A wide range of statistical methods has been applied, though the literature available to the public is limited for reasons of commercial confidentiality. Particular problems arising in the credit scoring context are examined and the statistical methods which have been applied are reviewed.

Title: A comparison of neural networks and linear scoring models in the credit union environment

Author: V. S. Desai, J. N. Crook and G. A. Overstreet

The purpose of the present paper is to explore the ability of neural networks such as multilayer perceptrons and modular neural networks, and traditional techniques such as linear discriminant analysis and logistic regression, in building credit scoring models in the credit union environment. Also, since

funding and small sample size often preclude the use of customized credit scoring models at small credit unions, we investigate the performance of generic models and compare them with customized models. Our results indicate that customized neural networks offer a very promising avenue if the measure of performance is percentage of bad loans correctly classified. However, if the measure of performance is percentage of good and bad loans correctly classified, logistic regression models are comparable to the neural networks approach. The performance of generic models was not as good as the customized models, particularly when it came to correctly classifying bad loans. Although we found significant differences in the results for the three credit unions, our modular neural network could not accommodate these differences, indicating that more innovative architectures might be necessary for building effective generic models.

Title: Credit Scoring Methods. Czech Journal of Economics and Finance

Author: M.Vojtek and E. Kocenda

The paper reviews the best-developed and most frequently applied methods of credit scoring employed by commercial banks when evaluating loan applications. The authors concentrate on retail loans – applied research in this segment is limited, though there has been a sharp increase in the volume of loans to retail clients in recent years. Logit analysis is identified as the most frequent credit-scoring method used by banks. However, other nonparametric methods are widespread in terms of pattern recognition. The methods reviewed have potential for application in post-transition countries.

Title: A survey of credit and behavioral scoring: forecasting: financial risk of lending to customers

Author: L. C. Thomas

Credit scoring and behavioural scoring are the techniques that help organisations decide whether or not to grant credit to consumers who apply to them. This article surveys the techniques used — both statistical and operational research based — to support these decisions. It also discusses the need to incorporate economic conditions into the scoring systems and the way the systems could change from estimating the probability of a consumer defaulting to estimating the profit a consumer will bring to the lending organisation — two of the major developments being attempted in the area. It points out how successful has been this under-researched area of forecasting financial risk. 2000 Elsevier Science B.V. All rights reserved.

Title: Neural nets versus conventional techniques in credit scoring in Egyptian banking

Author: H. Abdou, J. Pointon, and A. El-Masry

The number of Non-Performing Loans has increased in recent years, paralleling the current financial crisis, thus increasing the importance of credit scoring models. This study proposes a three stage hybrid Adaptive Neuro Fuzzy Inference System credit scoring model, which is based on statistical techniques and Neuro Fuzzy. The proposed model's performance was compared with conventional and commonly utilized models. The credit scoring models are tested using a 10-fold cross-validation process with the credit card data of an international bank operating in Turkey. Results demonstrate that the proposed model consistently performs better than the Linear Discriminant Analysis, Logistic Regression Analysis, and Artificial Neural Network (ANN) approaches, in terms of average correct classification rate and estimated misclassification cost. As with ANN, the proposed model has learning ability; unlike ANN, the model does not stay in a black box. In the proposed model, the interpretation of

Independent variables may provide valuable information for bankers and consumers, especially in the explanation of why credit applications are rejected.

Title: A credit scoring model for Vietnams retail banking market

Author: T.H.T. Dinh, and S. Kleimeier

As banking markets in developing countries are maturing, banks face competition not only from other domestic banks but also from sophisticated foreign banks. Given the substantial growth of consumer credit and increased regulatory attention to risk management, the development of a well-functioning credit assessment framework is essential. As part of such a framework, we propose a credit scoring model for Vietnamese retail loans. First, we show how to identify those borrower characteristics that should be part of a credit scoring model. Second, we illustrate how such a model can be calibrated to achieve the strategic objectives of the bank. Finally, we assess the use of credit scoring models in the context of transactional versus relationship lending.

CHAPTER 3

SYSTEM REQUIREMENTS

HARDWARE REQUIRMENTS

System : Pentium IV 2.4GHZ

HardDisk : 40 Gb

Ram : 512 Mb

SOFTWARE REQUIRMENTS

Operating system : Window 7(32 bit)

Coding Language : java

Tool : Netbeans

3.3 SOFTWARE ENIVORNMENT

Java is one of the world's most important and widely used computer languages, and it has held this distinction for many years. Unlike some other computer languages whose influence has weared with passage of time, while Java's has grown.

3.3.1 APPLICATION OF JAVA

Java is widely used in every corner of world and of human life. Java is not only used in software but is also widely used in designing hardware controlling software components. There are more than 930 million JRE downloads each year and 3 billion mobile phones run java.

Following are some other usage of Java:

1. Developing Desktop Applications
2. Web Applications like Linkedin.com, Snapdeal.com etc
3. Mobile Operating System like Android
4. Embedded Systems
5. Robotics and games etc.

3.3.2 FEATURES OF JAVA

The prime reason behind creation of Java was to bring portability and security feature into a computer language. Beside these two major features, there were many other features that played an important role in moulding the final form of this outstanding language. Those features are;

Simple

Java is easy to learn and its syntax is quite simple, clean and easy to understand. The confusing and ambiguous concepts of C++ are either left out in Java or they have been re-implemented in a cleaner way.

eg: Pointers and Operator Overloading are not there in java but were an important part of C++.

Object Oriented

In java everything is Object which has some data and behaviour. Java can be easily extended as it is based on Object Model.

Robust

Java makes an effort to eliminate error prone codes by emphasizing mainly on compile time error checking and runtime checking. But the main areas which Java improved were Memory Management and mishandled Exceptions by introducing automatic Garbage Collector and Exception Handling.

Platform Independent

Unlike other programming languages such as C, C++ etc. which are compiled into platform specific machines. Java is guaranteed to be write- once, run-anywhere language.

On compilation Java program is compiled into byte code. This byte code is platform independent and can be run on any machine, plus this byte code format also provide security. Any machine with Java Runtime Environment can run Java Programs.

Secure

When it comes to security, Java is always the first choice. With java secure features it enable us to develop virus free, temper free system. Java program always runs in Java runtime environment with almost null interaction with system OS, hence it is more secure.

Multi-Threading

Java multithreading feature makes it possible to write program that can do many tasks simultaneously. Benefit of multithreading is that it utilizes same memory and other resources to execute multiple threads at the same time, like While typing, grammatical errors are checked along.

Architectural Neutral

Compiler generates byte codes, which have nothing to do with a particular computer architecture, hence a Java program is easy to interpret on any machine.

Portable

Java Byte code can be carried to any platform. No implementation dependent features. Everything related to storage is predefined, example: size of primitive data types

High Performance

Java is an interpreted language, so it will never be as fast as a compiled language like C or C++. But, Java enables high performance with the use of just-in-time compiler.

3.4 COLLECTION FRAMEWORK

Collection framework was not part of original Java release. Collections was added to J2SE 1.2. Prior to Java 2, Java provided adhoc classes such as Dictionary, Vector, Stack and Properties to store and manipulate groups of objects. Collection framework provides many important classes and interfaces to collect and organize group of alike objects.

3.5 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

The feasibility study investigates the problem and the information needs of the stakeholders. It seeks to determine the resources required to provide an information systems solution, the cost and benefits of such a solution, and the feasibility of such a solution.

The goal of the feasibility study is to consider alternative information systems solutions, evaluate their feasibility, and propose the alternative most suitable to the organization. The feasibility of a proposed solution is evaluated in terms of its components.

3.5.1 ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour

into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

3.5.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

3.5.3 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity.

CHAPTER 4

SYSTEM ANALYSIS

4.1 PURPOSE

The purpose of this document is to define the requirements of credit card fraud detection. In detail, this document will provide a general description of our project, including user requirements, product perspective, and overview of requirements, general constraints. In addition, it will also provide the specific requirements and functionality needed for this project - such as interface, functional requirements and performance requirements.

4.2 SCOPE

The scope of this SRS document persists for the entire life cycle of the project. This document defines the final state of the software requirements agreed upon by the customers and designers. Finally at the end of the project execution all the functionalities may be traceable from the SRS to the product. The document describes the functionality, performance, constraints, interface and reliability for the entire life cycle of the project.

4.3 EXISTING SYSTEMS

- In case of credit card fraud detection the existing system is detect the fraud after fraud has been happen. Existing system maintain the large amount of data when customer comes to know about inconsistency in transaction he/she made complaint and then fraud detection system start it working. It first tries to detect that fraud has actually occur after that it

transactions that was used to fraud detection mechanism developed by master and visa cards

- A machine learning paradigm classification, with Credit Card Fraud Detection being the base.
- Intrusion detections to track fraud location and so on. In case of existing system there is no confirmation of recovery of fraud and Customer satisfaction.
- Secure electronic system used to analyze the behaviour of legitimate users.
- Data Mining mechanisms to classify and preprocess the users data.
- Genetic algorithms.

DISADVANTAGES OF EXISTING SYSTEM

- Each payment system has its limits regarding the maximum amount in the account, the number of transactions per day and the amount of output.
- If Internet connection fails, you can not get to your online account.
- If you follow the security rules the threat is minimal. The worse situation when the system of processing company has been broken, because it leads to the leak of personal data on cards and its owners.
- The information about all the transactions, including the amount, time and recipient are stored in the database of the payment system. And it means the intelligence agency has an access to this information. Sometimes this is the path for fraudulent activities.

4.4 PROPOSED SYSTEM

- The aim of the proposed system is to develop a application. which has capability to restrict and block the transaction performing by attacker from genuine user's credit card details.
- As we seen the existing system detects the fraud after fraud has been occurred . The proposed system tries to detect fraudulent transaction before transaction succeed.
- In proposed system we are using pattern recongnization, which works on transaction behavior of user. By Using pattern recongnization, after certain transactions we find one threshold value by using this threshold value we can compare current transaction with threshold value if transaction is quite different from user behavior then check whether it is genuine or fraud OTP (full form) and security questions are used.
- To get security from hackers we are providing encryption at registration time for password this encryption is done by Secure Hash Algorithm (SHA) algorithm,other encryption algorithms.

ADVANTAGES OF PROPOSED SYSTEM

- Online alerts are available if suspicious activity is detect on your card. The issuer will notify if any unusual charges come through on the card.
- Credit scores are used in numerous ways and when a person uses a credit card responsibly and makes timely payments without over limits or late payments, their credit score rises.
- By making payments on time and keeping the balance low on the card, a person will save money on the amount of interest charged by the card issuer.

- Eco friendly,no personal presence is required,it treats with growing technology,time saving.

CHAPTER 5

SYSTEM DESIGN

5.1 INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

What data should be given as input?

How the data should be arranged or coded?

The dialog to guide the operating personnel in providing input.

Methods for preparing input validations and steps to follow when error occur.

5.2 OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output

design it is determined how the information is to be displayed for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

The output form of an information system should accomplish one or more of the following objectives.

Convey information about past activities, current status or projections of the

Future.

Signal important events, opportunities, problems, or warnings.

Trigger an action.

Confirm an action

5.3 DATA FLOW DIAGRAM

Data Flow Diagram (DFD) is a two-dimensional diagram that describes how data is processed and transmitted in a system. The graphical depiction recognizes each source of data and how it interacts with other data sources to reach a mutual output. In order to draft a data flow diagram one must

- Identify external inputs and outputs
- Determine how the inputs and outputs relate to each other
- Explain with graphics how these connections relate and what they result

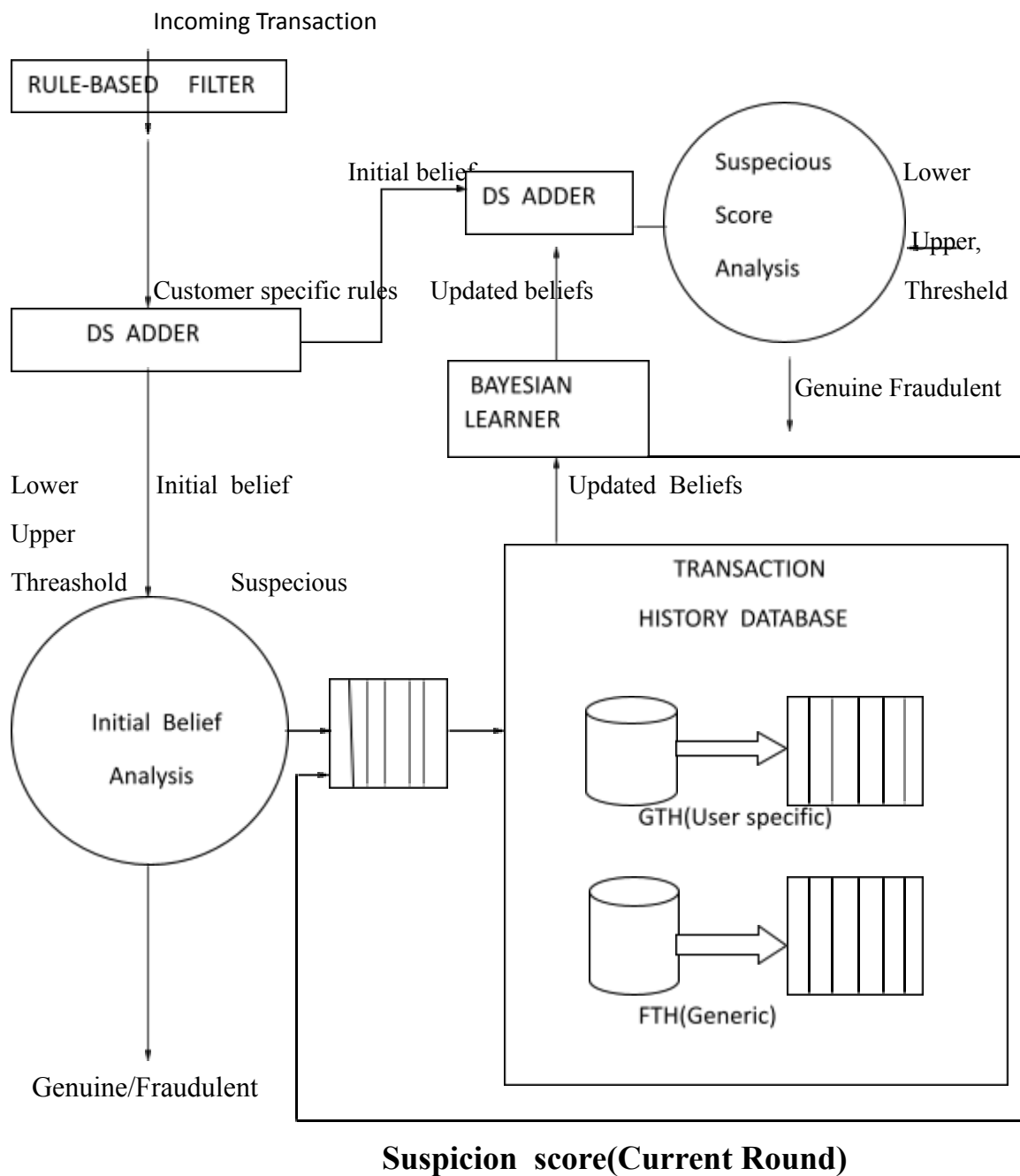
in.

Role of DFD:

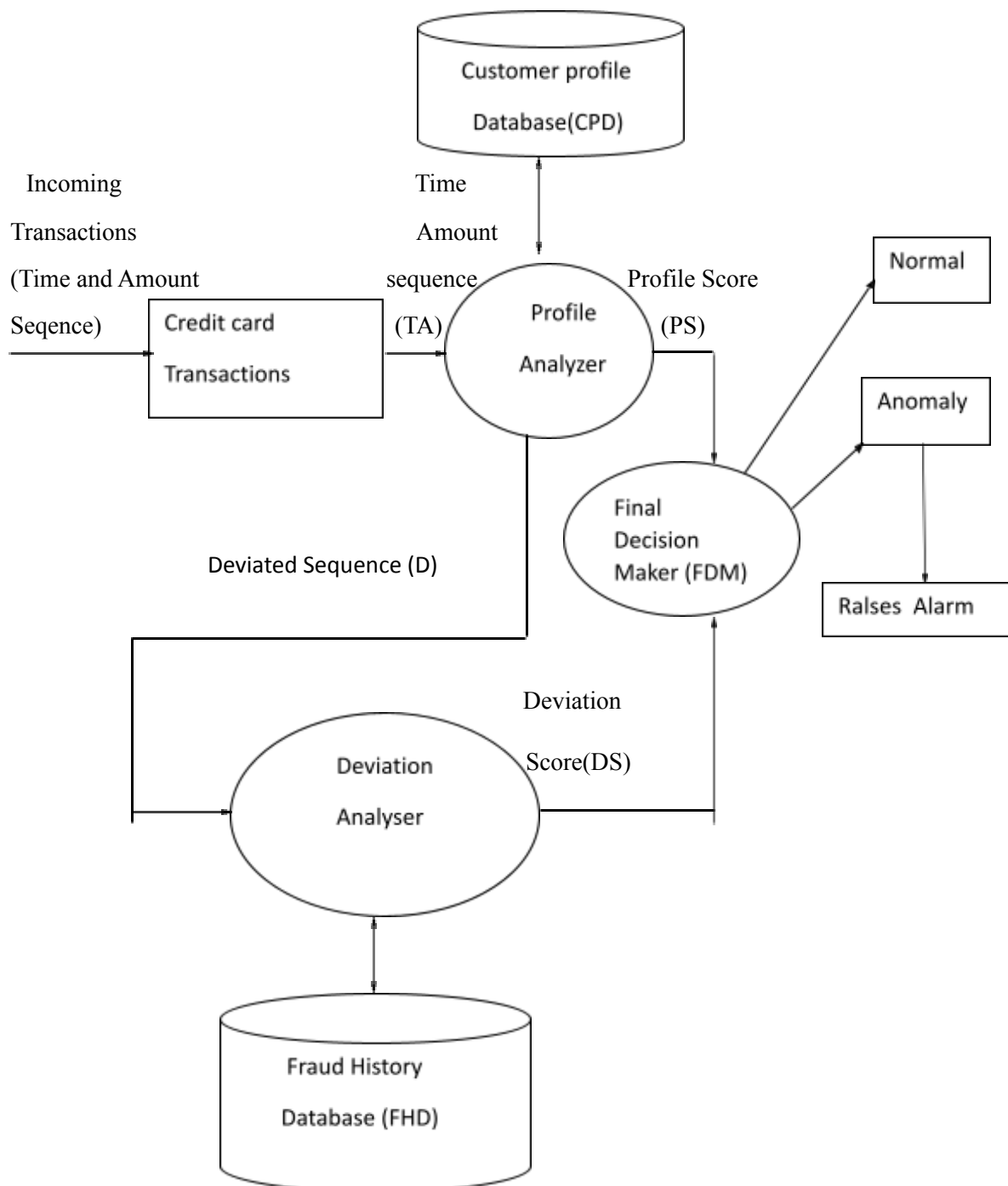
- It is a documentation support which is understood by both programmers and nonprogrammers. As DFD postulates only what processes are accomplished not how they are performed.

- A physical DFD postulates where the data flows and who processes the data.

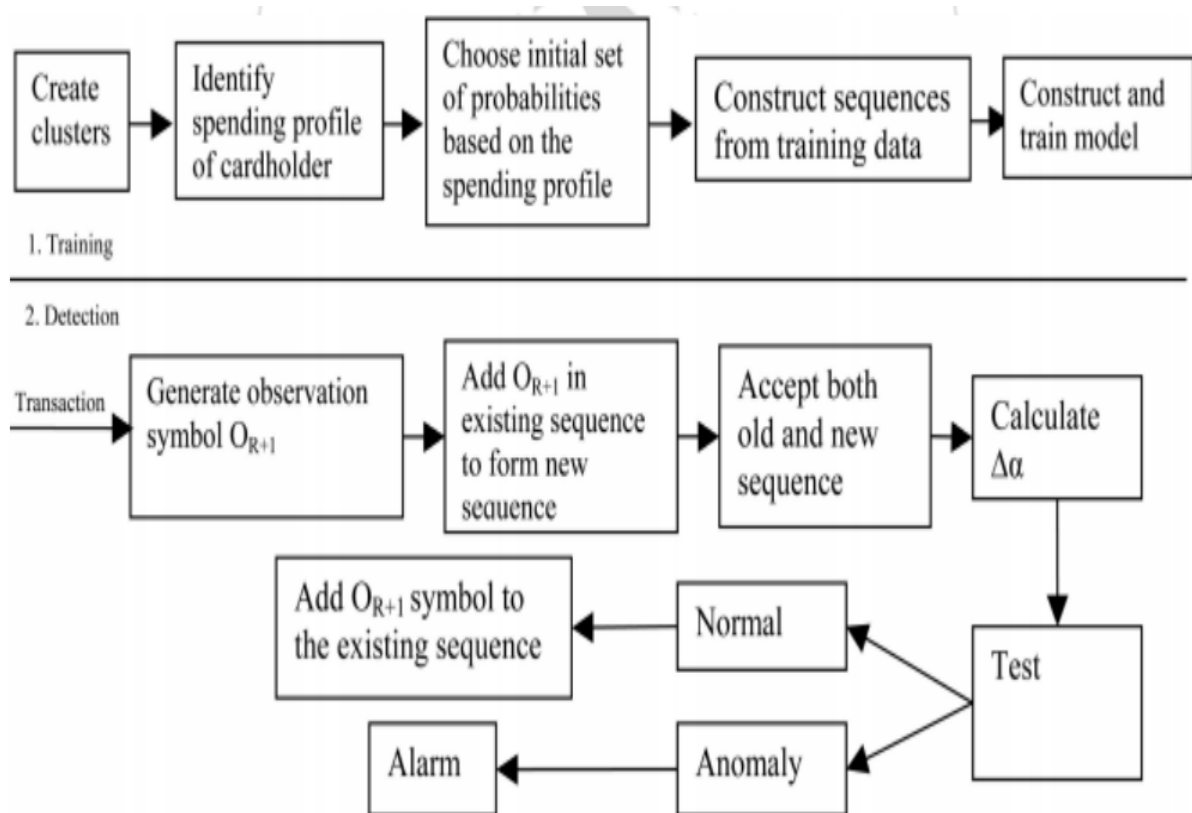
- It permits analyst to isolate areas of interest in the organization and study them by examining the data that enter the process and viewing how they are altered when they leave.



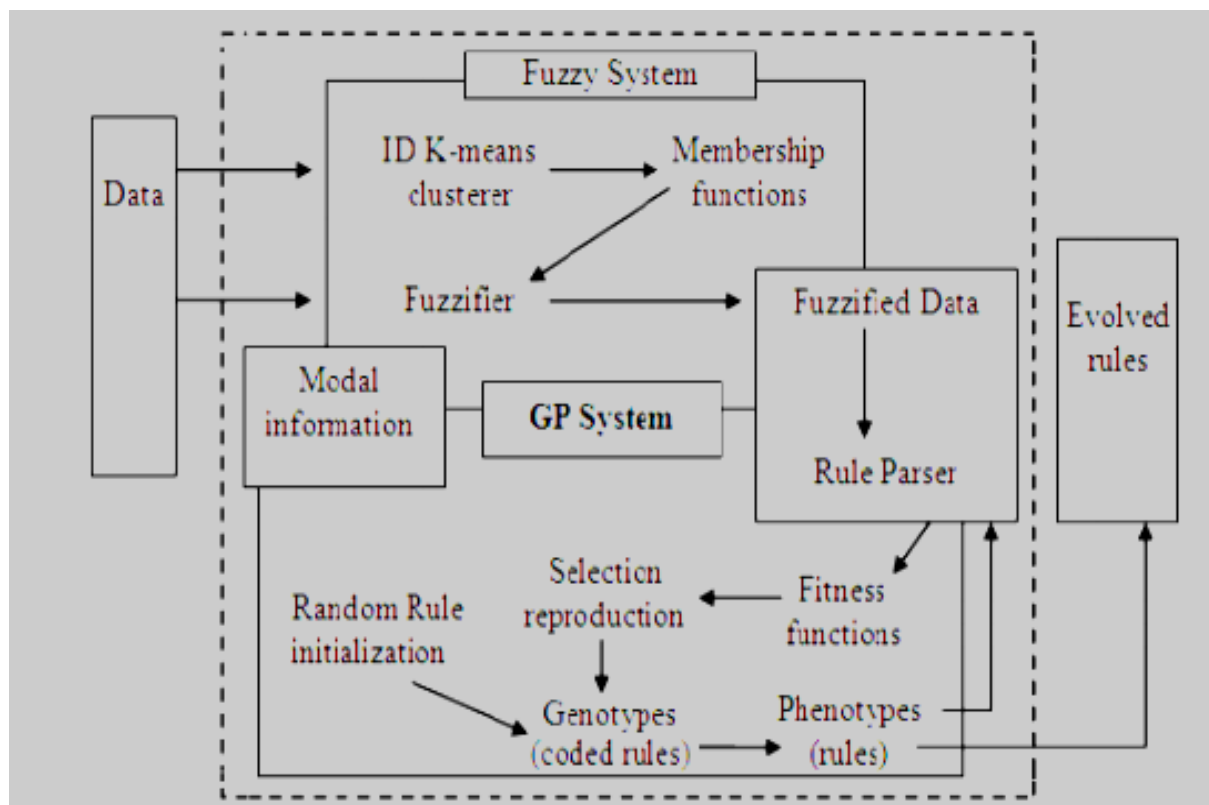
DFD-1



DFD-2



DFD-3



CHAPTER 6

MODULES

6.1 MODULES

- ❖ User GUI

❖ Critical Value Identification

❖ Fraud Detection using Genetic Algorithm

MODULES DESCRIPTION

6.1.1 User GUI

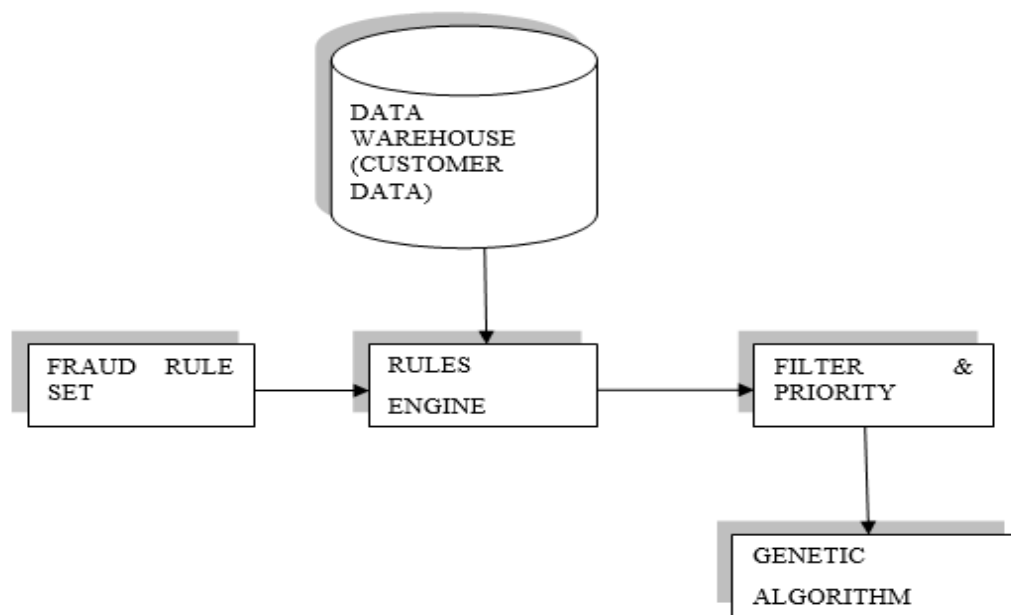
In this module, User Interface module is developed using Applet Viewer. This module is developed to user to identify the credit card fraud using genetic algorithm technique. So the user interface must be capable of providing the user to upload the dataset and make manipulations and finally must show the user whether fraud has been detected or not. Only final output will be in applet screen. All the generation details (crossover and mutation) will b in the console screen of eclipse.

6.1.2 Critical Value Identification

- ▶ Based on CC usage Frequency
- ▶ Based on CC usage Location
- ▶ Based on CC Over Draft
- ▶ Based on CC Book Balance
- ▶ Based on CC Average Daily Spending

6.1.3 Fraud Detection using Genetic Algorithm

In this module the system must detect whether any fraud has been occurred in the transaction or not. It must also display the user about the result.



CHAPTER 7

SYSTEM IMPLEMENTATION

7.1 SYSTEM ARCHITECTURE

Describing the overall features of the software is concerned with defining the requirements and establishing the high level of the system. During architectural design, the various web pages and their interconnections are identified and designed. The major software components are identified and decomposed into processing modules and conceptual data structures and the interconnections among the modules are identified. The following modules are identified in the proposed system.



The above architecture describes the work structure of the system.

- The customer data in the data warehouse is subjected to the rules engine which consists of the fraud rule set.

- The filter and priority module sets the priority for the data and then sends it to the genetic algorithm which performs its functions and generates the output.

7.2 Problem Statement:

The Credit Card Fraud Detection Problem includes modeling past credit card transactions with the knowledge of the ones that turned out to be fraud. This model is then used to identify whether a new transaction is fraudulent or not. Our aim here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications.

CHAPTER 8

SYSTEM TESTING

8.1 Test plan

Software testing is the process of evaluation a software item to detect differences between given input and expected output. Also to assess the feature of a software item. Testing assesses the quality of the product. Software testing

is a process that should be done during the development process. In other words software testing is a verification and validation process.

8.2 Verification

Verification is the process to make sure the product satisfies the conditions imposed at the start of the development phase. In other words, to make sure the product behaves the way we want it to.

8.3 Validation

Validation is the process to make sure the product satisfies the specified requirements at the end of the development phase. In other words, to make sure the product is built as per customer requirements.

8.4 Basics of software testing

There are two basics of software testing: black box testing and white box testing.

8.5 Black box Testing

Black box testing is a testing technique that ignores the internal mechanism of the system and focuses on the output generated against any input and execution of the system. It is also called functional testing.

8.6 White box Testing

White box testing is a testing technique that takes into account the internal mechanism of a system. It is also called structural testing and glass box

testing. Black box testing is often used for validation and white box testing is often used for verification.

8.7 Types of testing

There are many types of testing like

- Unit Testing
- Integration Testing
- Functional Testing
- System Testing
- Stress Testing
- Performance Testing
- Usability Testing
- Acceptance Testing
- Regression Testing
- Beta Testing

8.7.1 Unit Testing

Unit testing is the testing of an individual unit or group of related units. It falls under the class of white box testing. It is often done by the programmer to test that the unit he/she has implemented is producing expected output against given input.

8.7.2 Integration Testing

Integration testing is testing in which a group of components are combined to produce output. Also, the interaction between software and hardware is tested in integration testing if software and hardware components have any relation. It may fall under both white box testing and black box testing.

8.7.3 Functional Testing

Functional testing is the testing to ensure that the specified functionality required in the system requirements works. It falls under the class of black box testing.

8.7.4 System Testing

System testing is the testing to ensure that by putting the software in different environments (e.g., Operating Systems) it still works. System testing is done with full system implementation and environment. It falls under the class of black box testing.

8.7.5 Stress Testing

Stress testing is the testing to evaluate how system behaves under unfavorable conditions. Testing is conducted at beyond limits of the specifications. It falls under the class of black box testing.

8.7.6 Performance Testing

Performance testing is the testing to assess the speed and effectiveness of the system and to make sure it is generating results within a specified time as in performance requirements. It falls under the class of black box testing.

8.7.7 Usability Testing

Usability testing is performed to the perspective of the client, to evaluate how the GUI is user-friendly? How easily can the client learn? After learning

how to use, how proficiently can the client perform? How pleasing is it to use its design? This falls under the class of black box testing.

8.7.8 Acceptance Testing

Acceptance testing is often done by the customer to ensure that the delivered product meets the requirements and works as the customer expected. It falls under the class of black box testing.

8.7.9 Regression Testing

Regression testing is the testing after modification of a system, component, or a group of related units to ensure that the modification is working correctly and is not damaging or imposing other modules to produce unexpected results. It falls under the class of black box testing

REQUIREMENT ANALYSIS

Requirement analysis, also called requirement engineering, is the process of determining user expectations for a new modified product. It encompasses the tasks that determine the need for analyzing, documenting, validating and managing software or system requirements. The requirements should be documentable, actionable, measurable, testable and traceable related to identified business needs or opportunities and define to a level of detail, sufficient for system design.

FUNCTIONAL REQUIREMENTS

It is a technical specification requirement for the software products. It is the first step in the requirement analysis process which lists the requirements of particular software systems including functional, performance and security

requirements. The function of the system depends mainly on the quality hardware used to run the software with given functionality.

Usability

It specifies how easy the system must be use. It is easy to ask queries in any format which is short or long, porter stemming algorithm stimulates the desired response for user.

Robustness

It refers to a program that performs well not only under ordinary conditions but also under unusual conditions. It is the ability of the user to cope with errors for irrelevant queries during execution.

Security

The state of providing protected access to resource is security. The system provides good security and unauthorized users cannot access the system there by providing high security.

Reliability

It is the probability of how often the software fails. The measurement is often expressed in MTBF (Mean Time Between Failures). The requirement is needed in order to ensure that the processes work correctly and completely without being aborted. It can handle any load and survive and survive and even capable of working around any failure.

Compatibility

It is supported by version above all web browsers. Using any web servers like localhost makes the system real-time experience.

Flexibility

The flexibility of the project is provided in such a way that it has the ability to run on different environments being executed by different users.

Safety

Safety is a measure taken to prevent trouble. Every query is processed in a secured manner without letting others to know one's personal information.

NON- FUNCTIONAL REQUIREMENTS

Portability

It is the usability of the same software in different environments. The project can be run in any operating system.

Performance

These requirements determine the resources required, time interval, throughput and everything that deals with the performance of the system.

Accuracy

The result of the requesting query is very accurate and high speed of retrieving information. The degree of security provided by the system is high and effective.

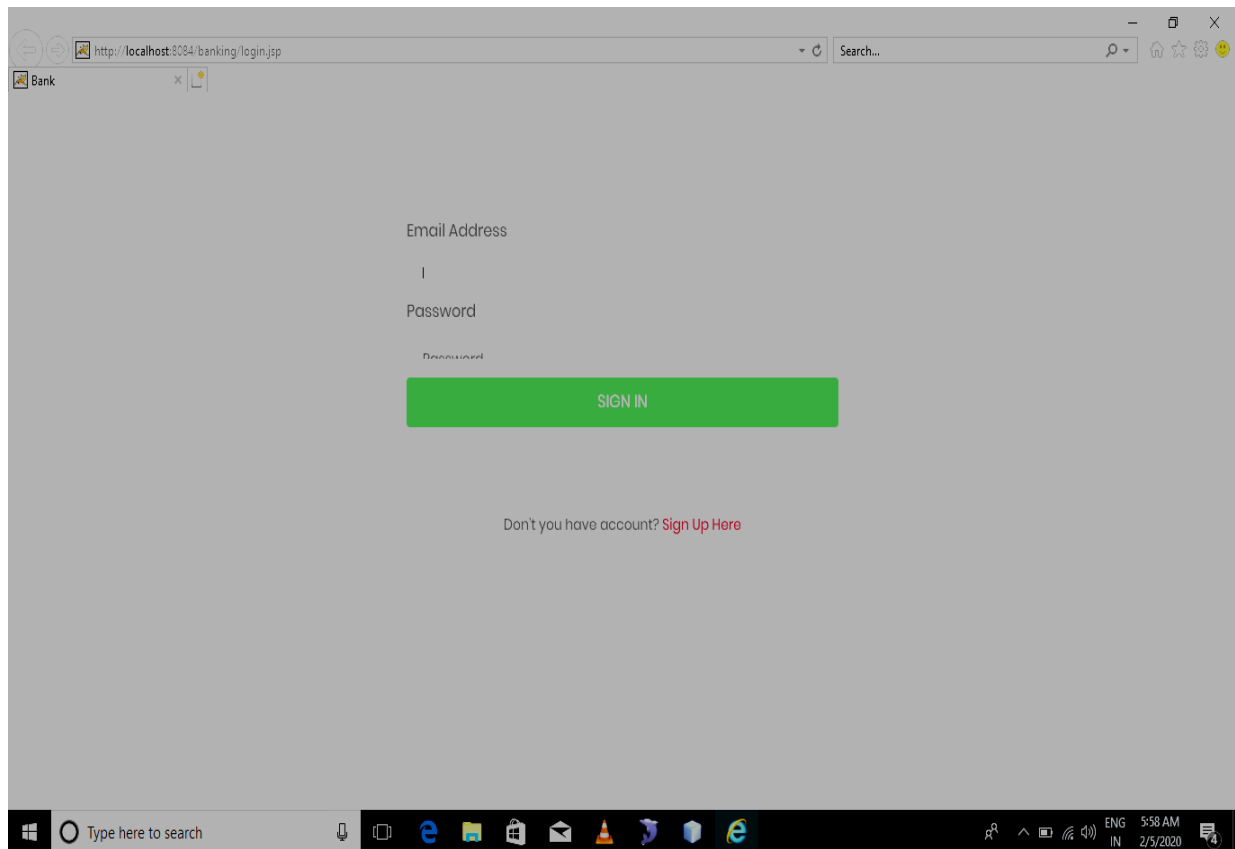
Maintainability

Project is simple as further updates can be easily done without affecting its stability. Maintainability basically defines that how easy it is to maintain the system. It means that how easy it is to maintain the system, analyse, change and test the application. Maintainability of this project is simple as further updates can be easily done without affecting its stability.

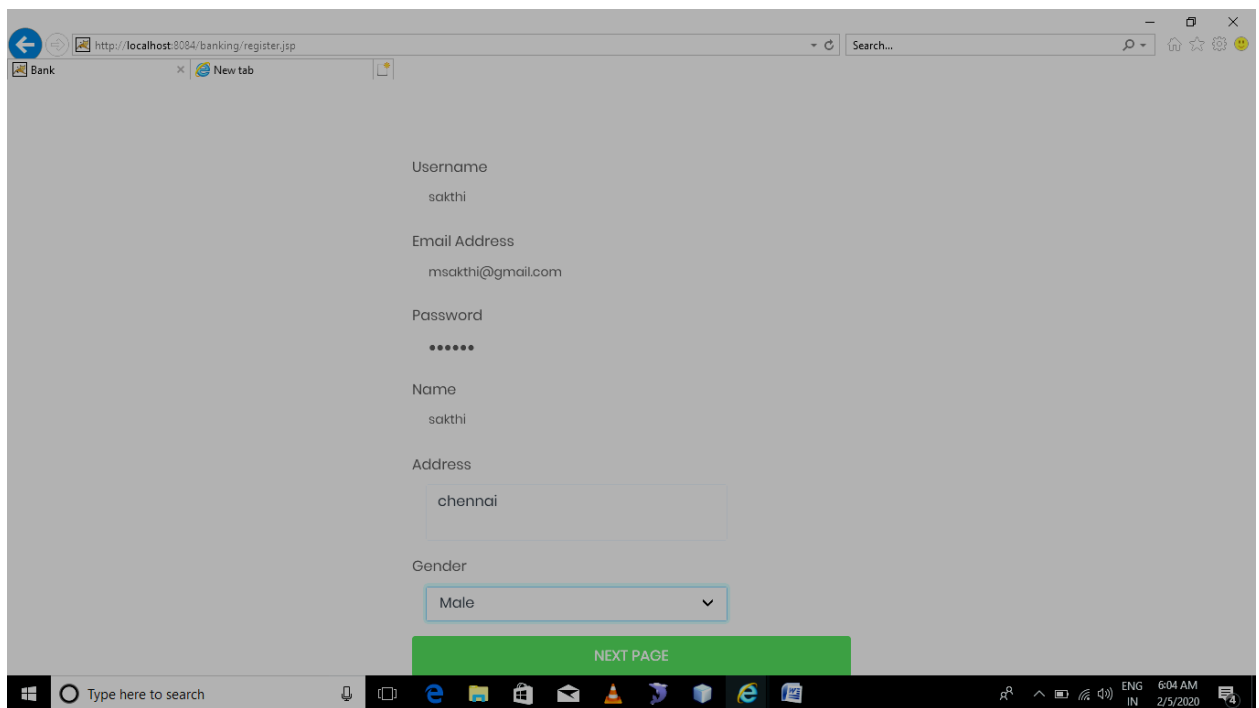
RESULTS

SCREENSHOTS

HOME PAGE



ADMIN SIGNUP



A screenshot of a web browser window displaying an admin signup form. The browser's address bar shows the URL `http://localhost:8084/banking/register.jsp`. The form is titled "ADMIN SIGNUP" and contains several input fields. The "Username" field is filled with "sakthi". The "Email Address" field is filled with "msakthi@gmail.com". The "Password" field is masked with seven dots. The "Name" field is filled with "sakthi". The "Address" field is filled with "chennai". The "Gender" field is a dropdown menu with "Male" selected. Below the form is a green button labeled "NEXT PAGE". The Windows taskbar is visible at the bottom, showing the search bar and various application icons. The system clock in the bottom right corner indicates the time is 6:04 AM on 2/5/2020.

ADMIN SIGNUP

Username
sakthi

Email Address
msakthi@gmail.com

Password
.....

Name
sakthi

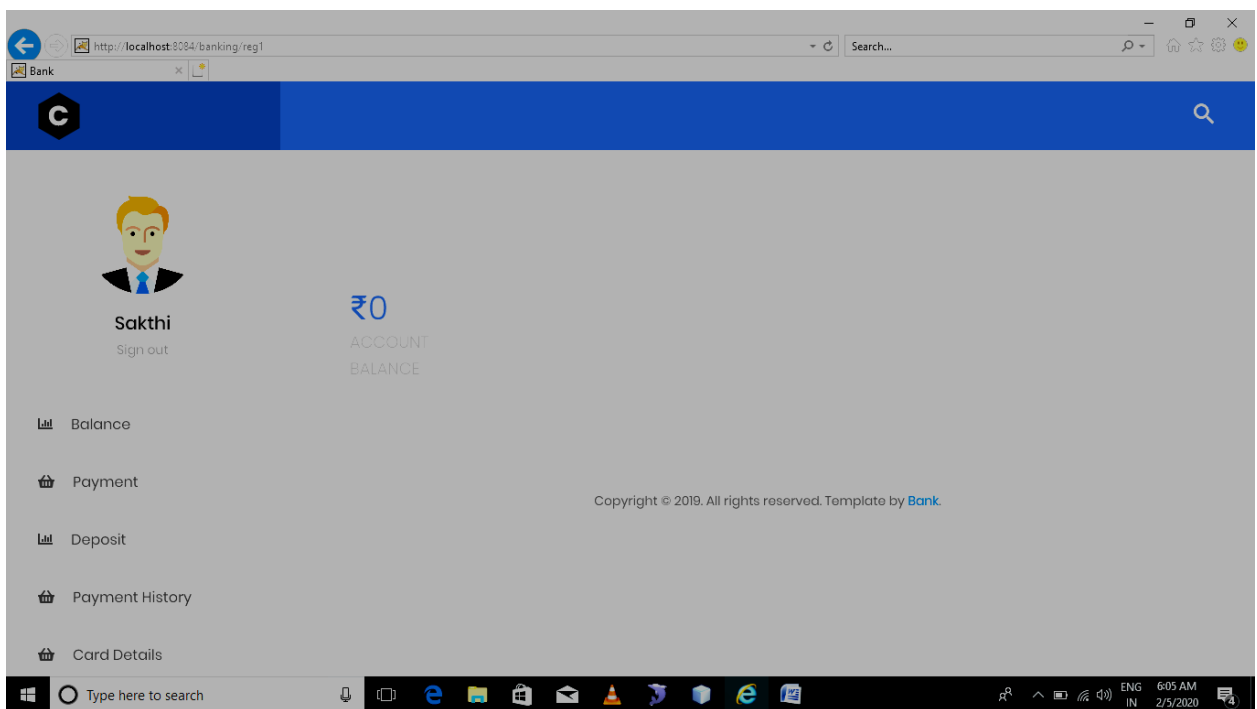
Address
chennai

Gender
Male

NEXT PAGE

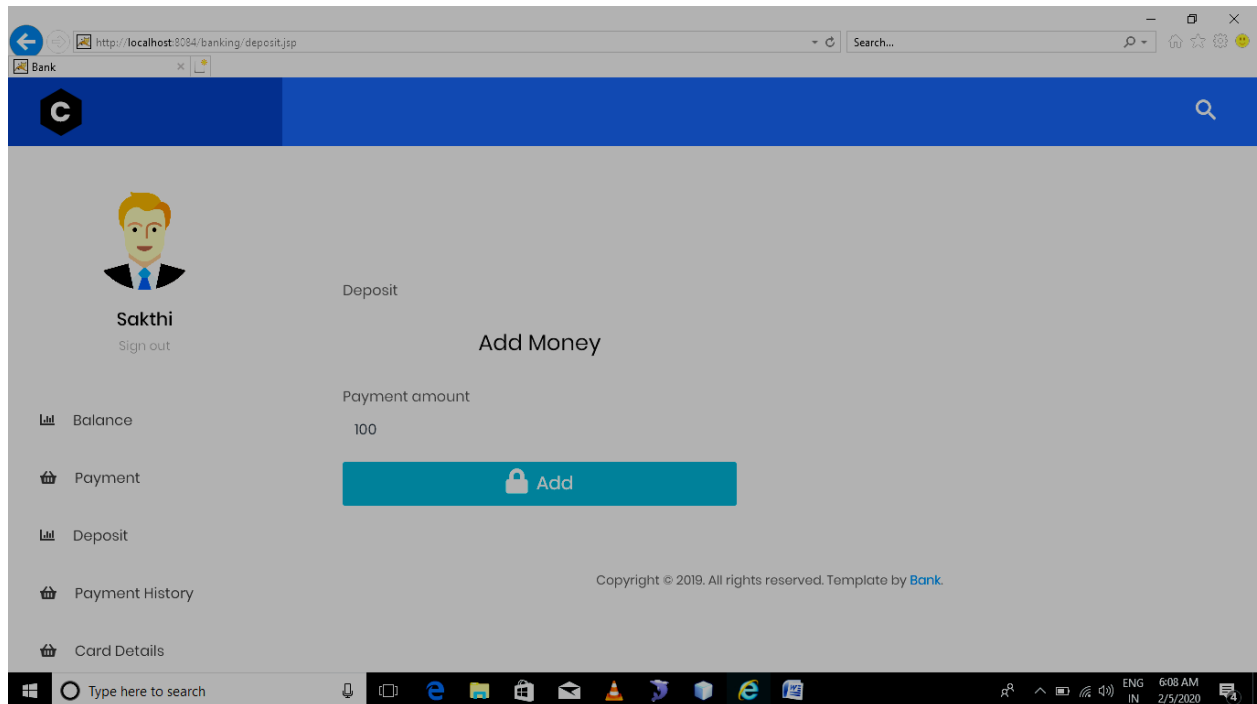
ADMIN BALANCE

In the account user are balance checking to the webpage and used one or more to account.



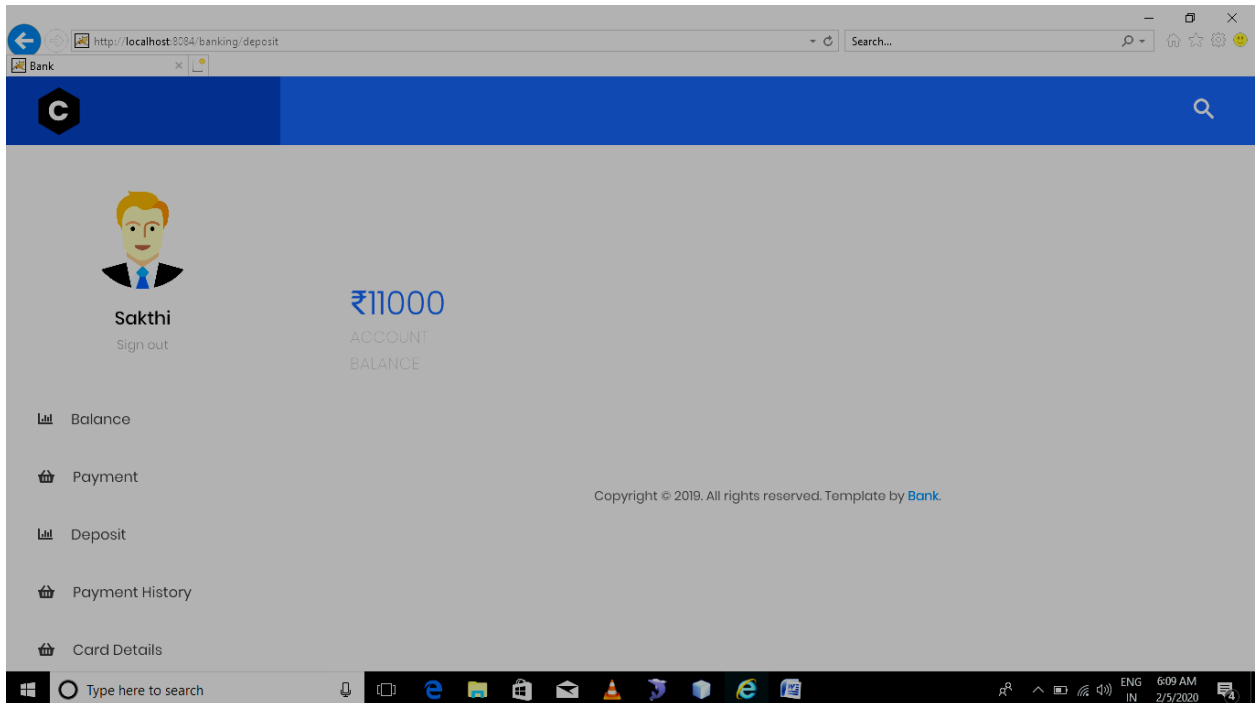
DEPOSIT PAGE

User Deposit money in our external account. Because in you could shopping there money contain in our account. If there zero account balance in our account cannot purchase any shopping, etc.....



AVAILABLE BALANCE

In there user will be deposit some , amount for using shopping and Balance will be increased



CHAPTER 9

CONCLUSION

In this paper, we have discussed that How Hidden Markov Model will be useful to detect fraudulent online transaction through credit card. The proposed Fraud Detection System is also scalable for handling vast volumes of transactions data processing. The HMM based credit card fraud detection

system is not having complex process to perform fraud check like the existing system. Proposed Fraud detection system gives genuine and fast result than existing system. The Hidden Markov Model makes the processing of detection very easy and tries to remove the complexity.

FUTURE ENHANCEMENT

We have discussed an application of HMM in credit card fraud detection. The different steps in credit card transaction processing are represented as the underlying stochastic process of an HMM. We have study the ranges of transaction amount as the observation symbols, whereas the types of item have been considered to be states of the HMM. We have study the suggested a method for finding the spending profile of cardholders, as well as application of this knowledge in deciding the value of observation symbols and initial estimate of the model parameters. It has also been explained how the HMM can detect whether an incoming transaction is fraudulent or not. The system is also scalable for handling large volumes of transactions.

REFERENCES

[1] D. J. Hand and W. E. Henley, Statistical classification methods in consumer credit scoring: A review. Journal of the Royal Statistical Society: Series A (Statistics in Society),160(3), 523–541,1997.

- [2] V. S. Desai, J. N. Crook and G. A. Overstreet, A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, 95(1), 24–37, 1996.
- [3] M. Vojtek and E. Kocenda, Credit Scoring Methods. *Czech Journal of Economics and Finance*, 56(3-7), 152-167, 2006.
- [4] L. C. Thomas, A survey of credit and behavioral scoring: forecasting: financial risk of lending to customers. *International Journal of Forecasting*, 16, 149-172, 2000.
- [5] H. Abdou, J. Pointon, and A. El-Masry, Neural nets versus conventional techniques in credit scoring in Egyptian banking. *Expert Systems with Applications*, 35, 1275-1292, 2008.
- [6] T.H.T. Dinh, and S. Kleimeier, A credit scoring model for Vietnams retail banking market. *International Review of Financial Analysis*, 16, 471- 495, 2007.

Websites:

- [1] http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/tcw2/report.html
- [2] http://www.kxcad.net/cae_MATLAB/toolbox/gads/f6691.html

[3]<http://java.sun.com/developer/onlineTraining/Programming/BasicJava1/front.html>

[4]<http://www.easywayserver.com/blog/user-login-in-jsp/>

[5]<http://www.faqs.org/patents/app/20100094765>

Textbooks:

[1]Pressman, Roger S. Software engineering: a practitioner's approach / Roger S. Pressman.—5th ed.p. cm (McGraw-Hill series in computer science).

[2]E.Balagurasamy, Programming with java, Tata McGraw-Hill Publication.

[3]Ali Bahrami, Object Oriented system Development, Tata McGraw-Hill Publication .

[4]Jiwei Han et.al., “Data Mining :Concepts and Techniques”, MorganKaufmaan Series,2000