

Comprehensive Analysis For Fraud Detection Of Credit Card Through Machine Learning

Submitted in partial fulfilment of the
requirements for the award of

Bachelor of Engineering degree in Computer Science and Engineering

By

Name : B.Prashanth
Reg No : 38110424



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SCHOOL OF COMPUTING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI 600 119

MARCH - 2022



SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of
"B.prashanth (38110424)" under my supervision from December 2021 to
March 2022.

DR.MERCY PAUL SELVAN M.E., PH.D.
Internal Guide

DR. L. LAKSHMANAN M.E., PH.D.
Head of the Department

Submitted for Viva-voce Examination held on_____

Internal Examiner

External Examiner

DECLARATION

We **B.prashanth** (38110424) hereby declare that the Project Report entitled “**comprehensive analysis for fraud detection of credit card through machine learning.**” done by us under the guidance of **DR. Mercy Paul Selvan M.E.,PHD** (internal) is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering / Technology degree in **Computer Science and Engineering.**

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to the **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E., Ph.D, Dean**, School of Computing **Dr. L. Lakshmanan M.E., Ph.D. ,** and **Dr.S.Vigneshwari M.E., Ph.D. Heads** of the Department of Computer Science and Engineering for providing me with the necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr.Mercy Paul Selvan M.E.,Ph.d.** for her valuable guidance, suggestions and constant encouragement that paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering who** were helpful in many ways for the completion of the project.

ABSTRACT

- 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.

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.

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 alerts 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.

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.

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

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 Vietnam's 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.

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.

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.

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.

PROPOSED SYSTEM

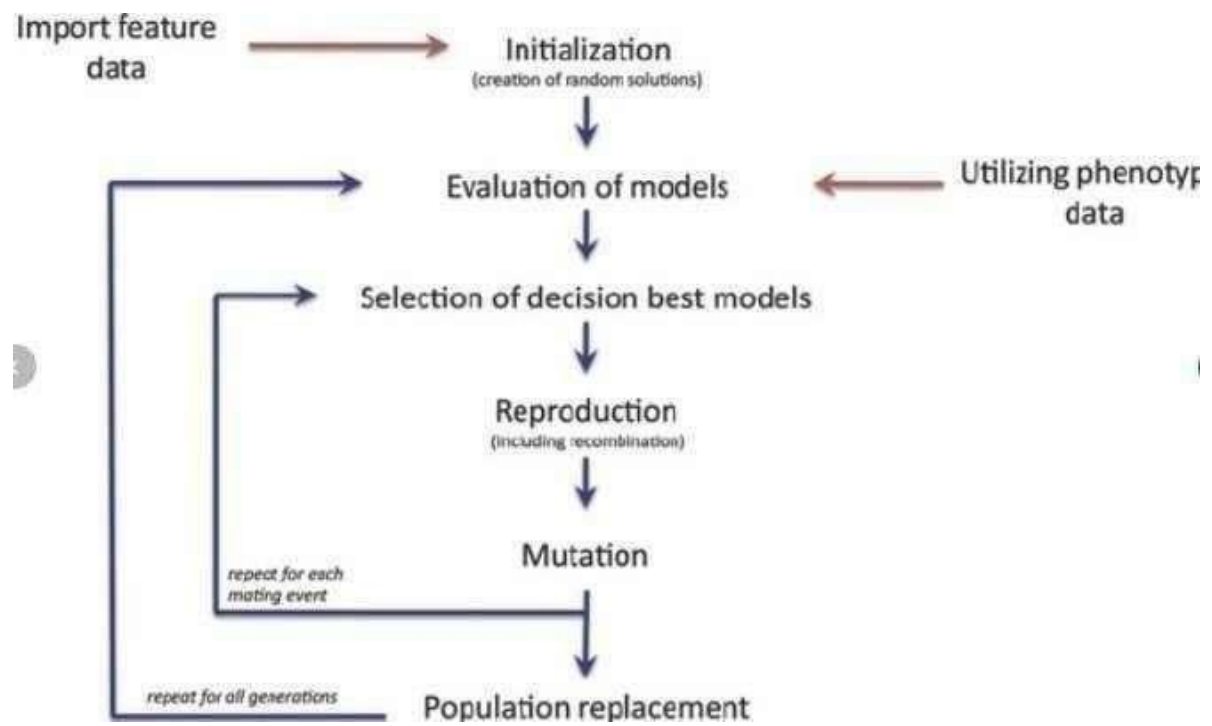
- The aim of the proposed system is to develop a application. which has capability to restrict and block the transaction performing by attacker from genuine user's credit card details.
- As we seen the existing system detects the fraud after fraud has been occurred . The proposed system tries to detect fraudulent transaction before transaction succeed.
- In proposed system we are using pattern recognition, 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 detected 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.

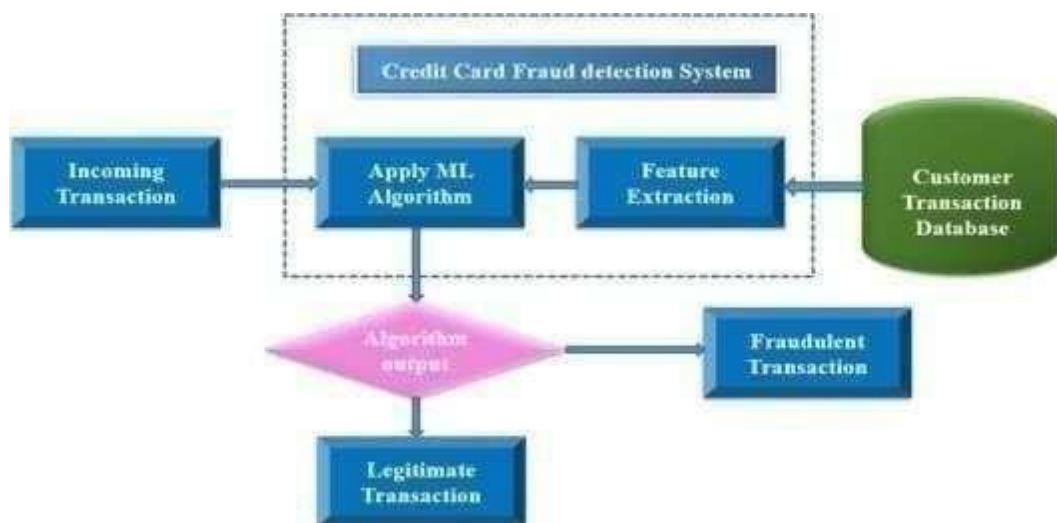
ALGORITHM USED

GENETIC ALGORITHM



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.

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.

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

MODULES

- ❖ User
GUI
- ❖ Critical Value
Identification
- ❖ Fraud Detection using Genetic
Algorithm

MODULES DESCRIPTION

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 be in the console screen of eclipse.

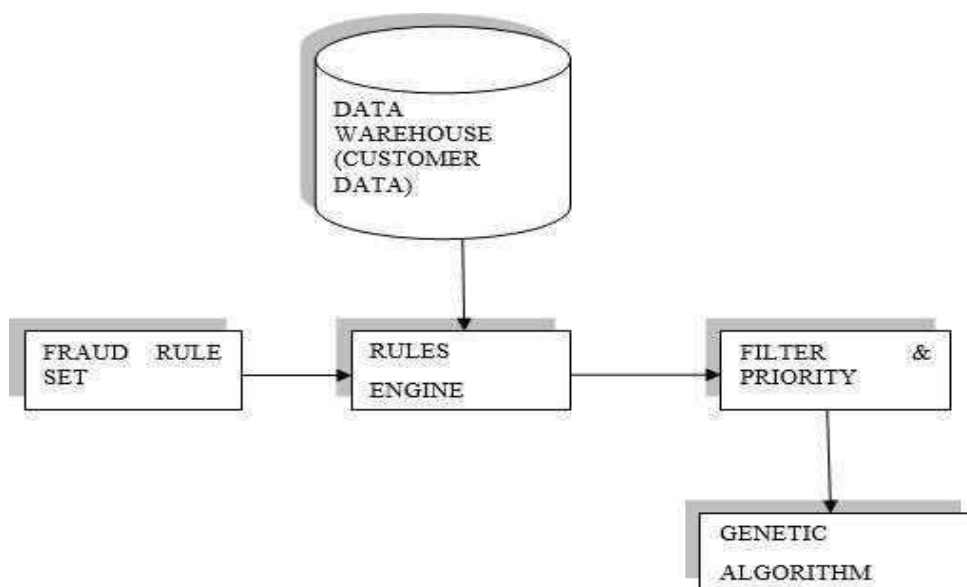
Critical Value Identification

- ▶ 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

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.



DATAFLOW 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

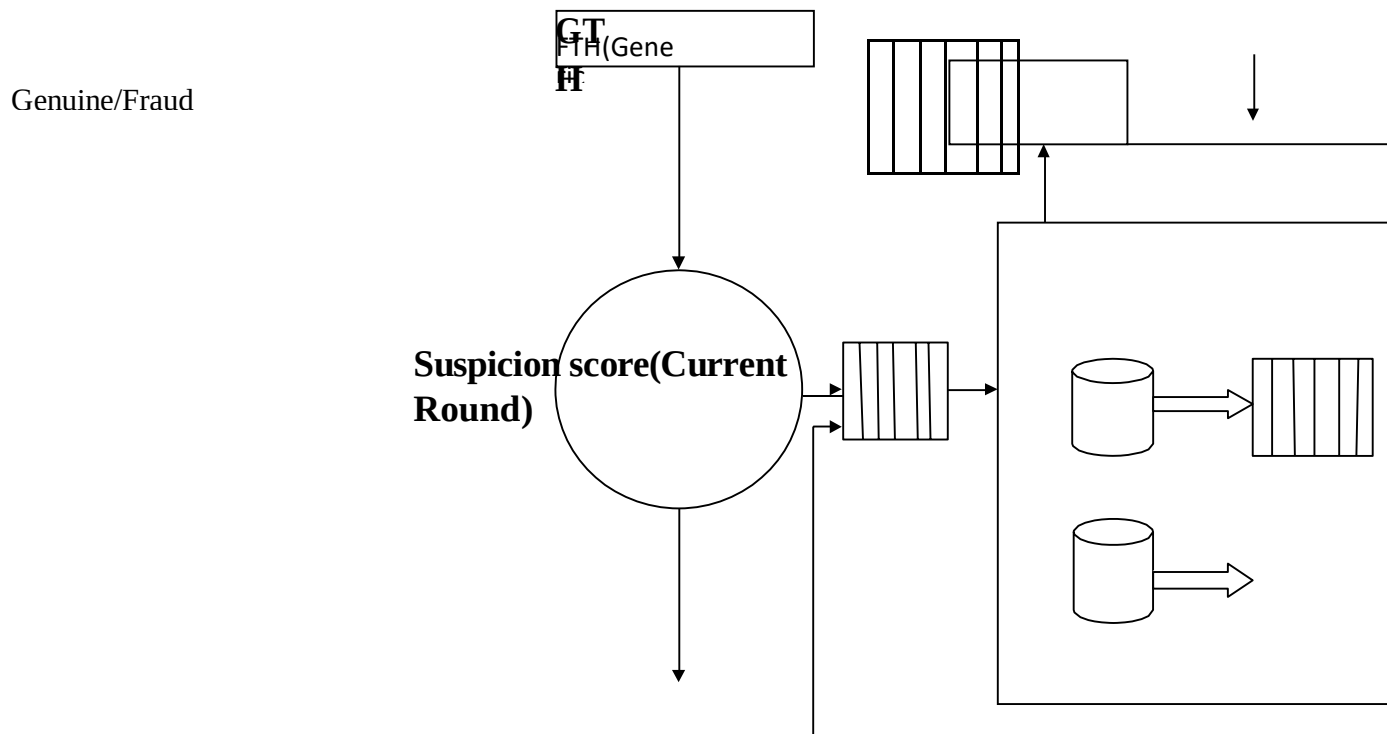
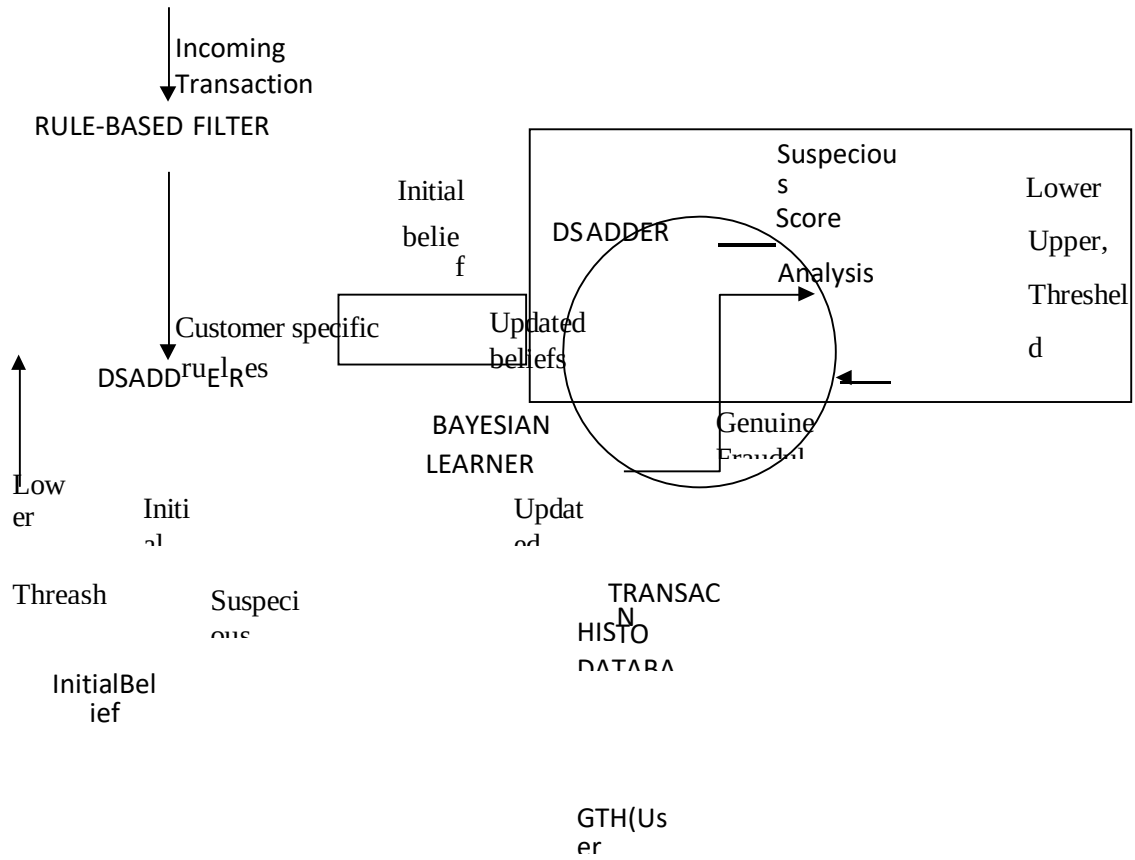
in.

Role of DFD:

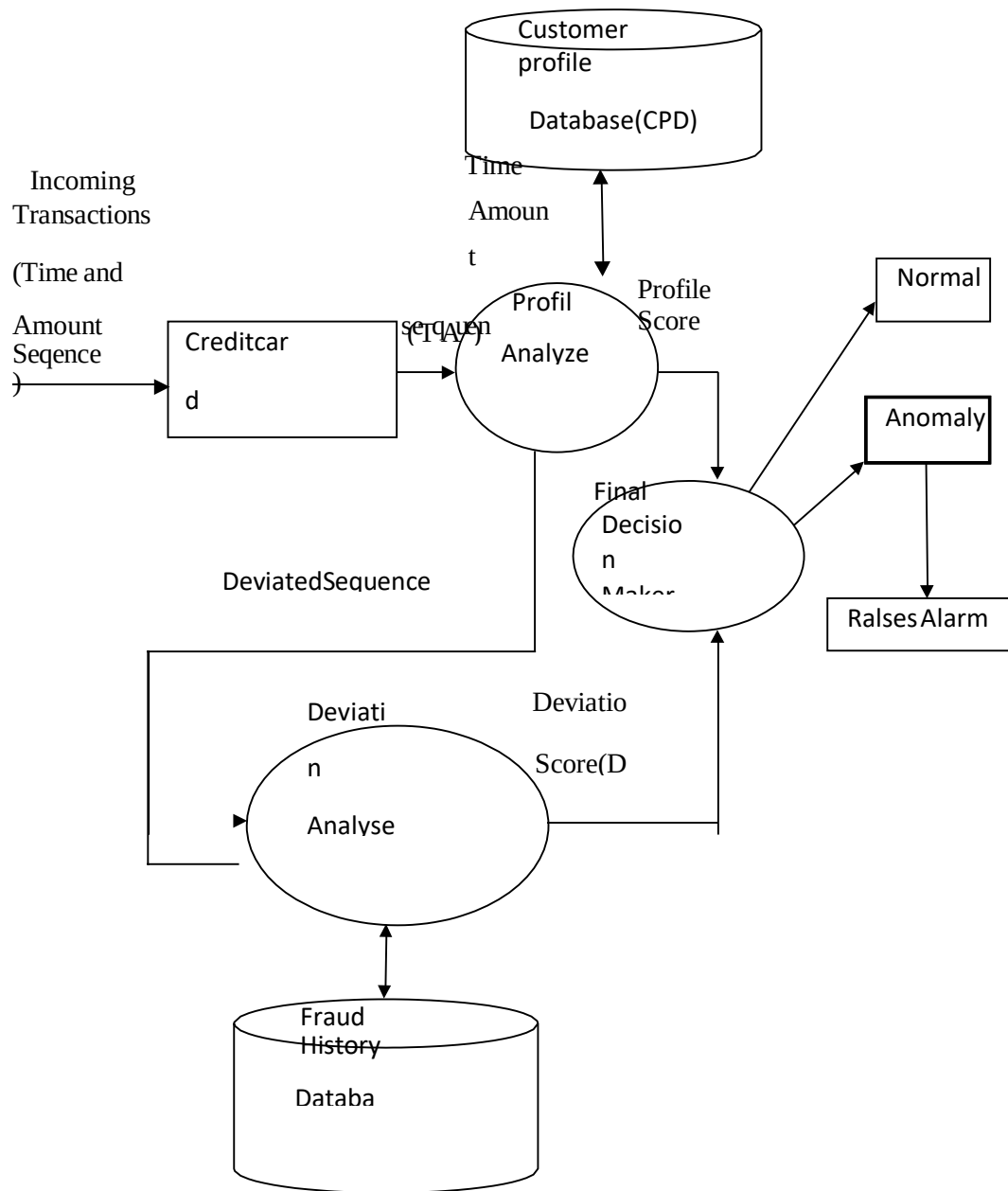
· It is a documentation support which is understood by both programmers and nonprogrammers. As DFD postulates only what processes are accomplished not how they are performed.

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

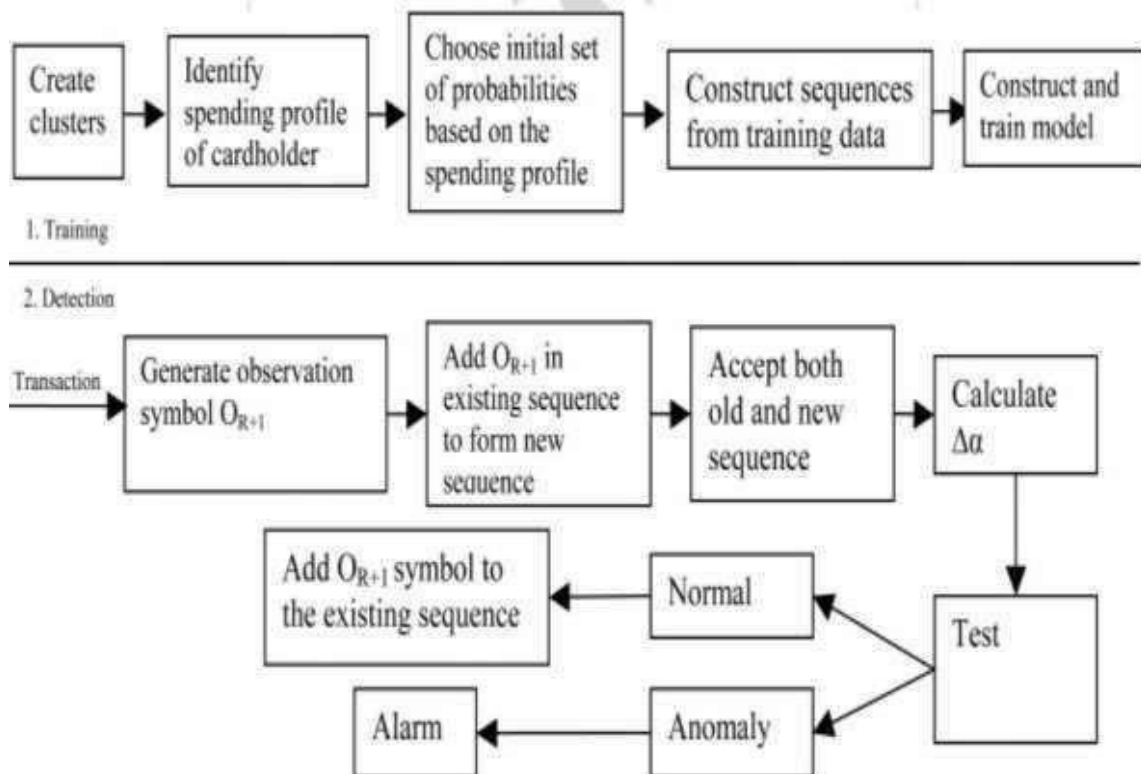
- It permits analyst to isolate areas of interest in the organization and study them by examining the data that enter the process and viewing how they ^{are} altered when they leave



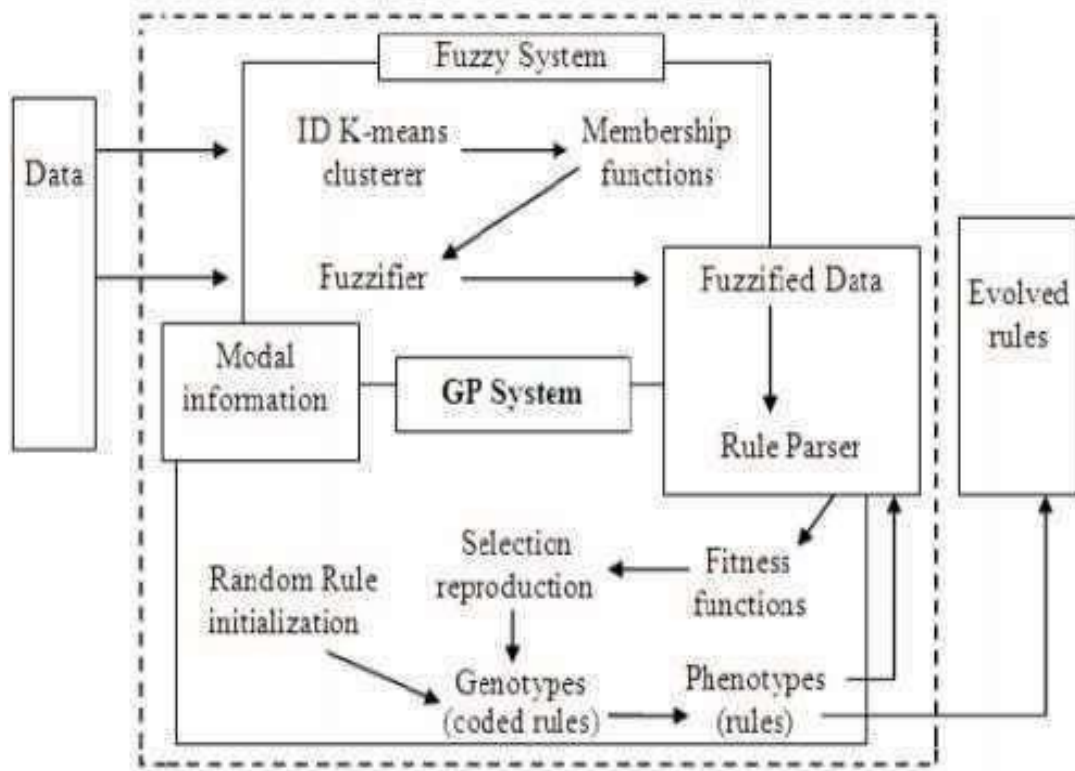
DFD-1



DFD-2



DFD-3



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 valueof 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.

RESULT: credit card is without a doubt an act of criminal dishonesty this article has out the most common methods of fraud along with their detection methods and reviewed recent findings in this field



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade 'A' By NAAC | 12B Status by UGC | Approved By AICTE
www.sathyabama.ac.in



School of Computing

Department of Information Technology

Certificate of Presentation

This is to certify that

Prashanth B

of

Sathyabama Institute of Science and Technology

has presented a paper entitled

Comprehensive Analysis for Fraud Detection of Credit Card through Machine Learning

in the AICTE sponsored

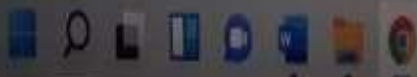
International Conference on Artificial Intelligence and Machine Learning (IAIM-2022).

held from 27th January 2022 - 29th January 2022.

Dr. R. SUBHASHINI
HOD [Information Technology]

Dr. T. SASIKALA
Dean [School of Computing]

Dr. T. SASIPRABA
Vice Chancellor



REFERENCES 1D. 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.

2V. 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.

3M. Vojtek and E. Kocenda, Credit Scoring Methods. Czech Journal of Economics and Finance, 56(3-7), 152-167, 2006.

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 System 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,Oject Oriented system Development, Tata McGraw-Hill Publication .
- 4 Jiwei Han etal., “Data Mining :Concepts and Techniques”, Morgan Kaufmaan Series ,2000