MACHINE INTELLIGENCE TO IMPROVE IDENTITY CRIME DETECTION

THESIS

Submitted

in partial fulfillment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY
IN THE DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

By

V MAREESWARI
(Regn. No. SP 12 CSD 046)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

St. PETER'S INSTITUTE OF HIGHER EDUCATION AND RESEARCH

St. PETER'S UNIVERSITY

CHENNAI 600 054

FEBRUARY 2017

MACHINE INTELLIGENCE TO IMPROVE IDENTITY CRIME DETECTION

THESIS

Submitted

in partial fulfillment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY
IN THE DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

By

V MAREESWARI
(Regn. No. SP 12 CSD 046)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

St. PETER'S INSTITUTE OF HIGHER EDUCATION AND RESEARCH

St. PETER'S UNIVERSITY

CHENNAI 600 054

FEBRUARY 2017

CERTIFICATE

I hereby certify that the thesis entitled, **"MACHINE INTELLIGENCE TO IMPROVE IDENTITY CRIME DETECTION"** revised and resubmitted to the St. Peter's University, for the award of Degree of Doctor of Philosophy is the record of research work done by the candidate **V.Mareeswari** under my guidance and that the thesis has not formed previously the basis for the award of any degree, diploma, associateship, fellowship or other similar titles.

Place: **DR. G.GUNASEKARAN**

Date: **SUPERVISOR**

DECLARATION

Certified that the thesis entitled **"MACHINE INTELLIGENCE TO IMPROVE IDENTITY CRIME DETECTION"** is the bonafide record of independent work done by me under the supervision of **Dr.G.Gunasekaran**. Certified further that the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred earlier.

DR. G.GUNASEKARAN **V.MAREESWARI**

SUPERVISOR

Place:

Date:

ACKNOWLEDGEMENTS

The completion of this research work would not have been possible without the encouragement, help and support of many individuals. It is my privilege to thank the people who have supported and guided me throughout this research work.

I am highly indebted to my supervisor **Dr.G.Gunasekaran**, Professor and Principal, Meenakshi College of Engineering, Chennai, for giving me an opportunity to work under his guidance. I would like to express my sincere gratitude to him for his valuable guidance, insightful suggestions and constructive inputs.

I am grateful to **Dr.V.Cyril Raj**, Head, Department of Computer Science and Engineering, Dr.M.G.R. Educational and Research Institute University, Chennai and **Dr.T.Ravi**, Professor and Head, Department of Computer Science and Engineering, KCG College of Technology, Chennai for guiding my research work as Doctoral Committee Members for my research.

I am grateful to Management, **Dr.S.Ravichandran,** Vice Chancellor, **Dr.M.A.Dorai Rangaswamy,** Registrar, **Dr.D.S.Ramachandra Murthy,** Director (R&D), **Dr.S.Gunasekaran,** Dean (R&D) and **Dr.L.Mahesh Kumar,** Director (Academic) of St. Peter's University, Chennai, for giving me an opportunity to carry out my research in the university.

I also thank **Dr.S.Pushpa**, Professor and Head, Computer Science and Engineering and the staff members of the Department of Computer Science and Engineering, St. Peter's University fortheir help during this research.

Finally, I would like to thank all those who were directly or indirectly helpful in carrying out this research.

V. MAREESWARI

ABSTRACT

Credit card application fraud detection has been an extensive research area in the past two decades and still continues to be the area of interest in various sectors of data mining. Various detection algorithms and techniques have been proposed for fraud detection producing results with varying efficiency and accuracy. The algorithms range from a simple to varying levels of implementation and structural complexity, each employing their own detection techniques. Further, what these techniques have in common is that they all aim at analyzing a single revision of software. The motivation behind this research is the fraud detection through various hybrid techniques. Moreover, the research focuses on detecting fraud through refactoring techniques and consistent modification of fraud.

In the first stage of the work endeavours to focus the hybrid Fast Throughput Multi-Pattern Matching algorithm (FTMPM), which is used to match the large amount of attributes. The main target that focused on this FTMPM is to safeguard the credit application in the initial stage of the credit life cycle. The implementation of Multi pattern matching algorithm in order to compare the attributes makes the identification process reliable with less time complexity. Here the two main challenges are time constraints and accuracy have been achieved with balanced data load. This FTMPM has not achieved the optimization level, because it worked along with the CD and SD approaches. To improve the performance in terms of the optimization, the second stage of the work introduces the new approach called as an Improved Sheep Flock Heredity Algorithm.

The second stage of the work proposes an Improved Sheep Flock Heredity Algorithm (ISFHA), which improves the efficiency of the credit card application fraud detection method by verifying and validating the optimized parameters, such as single and multiple attributes. The attributes of every application [offline/online] are verified by using a newly developed procedure MLMA-[Multi-Level-Multi-Agent] and it is verified all the attribute values are best one or not. For optimizing the attributes the ISFH-[Improved Sheep Flock Heredity] algorithm is used and those attributes are validated according to the time and response with optimal value.

The final stage of the work proposes the Hybrid Elephant Swarm Optimization Algorithm (HESOA), which increases the speed of fraud detection. The main aim of the work is to detect and prevent the credit card application fraud. In this method, based on the attributes of credit card application validation is performed, the prevention of fraud transaction and for analyzing the system performance, the uses of hybrid elephant swarm optimization are proposed based on the heuristic search algorithm for the credit card fraud detection. The search algorithm is used to find the similarity among the neighboring attributes and elephant swarm optimization is used for finding the optimal path and best fitness. Proposed system gives the best accuracy results and shows the data handling capacity for large databases. Results show the accuracy of 99.32% in terms of detection which is comparatively improved compared to other existing methods.

Keywords: BL, CD, Data Mining,EBS, FTMPM, HESOA, ISFHA,JMI, RBS, SD, WL

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AGA	-	Artificial genetic algorithm
AIS	-	Artificial Immune Systems
ANN	-	Artificial Neural Network
BIS	-	Biological Immune System
BL	-	Bayesian learner
BN	-	Bayesian Nets
CAS	-	Computer Algebra System
CEM	-	Cost Effective Method
CFLANN	-	Chebyshev functional link artificial neural network
CKT	-	Credit K -tuple Table
CLI	-	Command-Line Interfaces
CLPSO	-	Comprehensive Learning Particle Swarm Optimizer
DA	-	Deviation Analyzer
DD-SRPSO	-	Directionally Driven Self-Regulating Particle Swarm Optimization
DM	-	Decision Making
DMSPSO	-	Dynamic Multi-Swarm Particle Swarm Optimizer
DST	-	Dempster–ShaferTheory
DT	-	Decision Tree
EBS	-	Event Based Scheduling
EDI	-	Electronic Data Interchange
EFD	-	Early Fraud Detection
FDM	-	Final Decision Maker
FDS	-	Fraud Detection System
FDSCC	-	Fraud Detection System for Credit Card
FFD	-	Financial Fraud Detection
FIPS	-	Fully Informed Particle Swarm
FKT	-	Fraud K -tuple Table

FRST	-	Fast Response Time Scheduling
FTMPM	-	Fast Throughput Multi-Pattern Matching
FTST	-	Fast Throughput Time Scheduling
GEPSVM	-	Generalized Eigen value Proximal Support Vector Machine
HMM	-	Hidden Markov Model
HESOA	-	Hybrid Elephant Swarm Optimization Algorithm
HPSO-LS	-	Hybrid Particle Swarm Optimization-Local Search
ISFHA	-	Improved Sheep Flock Heredity Algorithm
JMI	-	Java Meta data Interface
KNN	-	K-Nearest Neighbors
KT	-	K -tuple Table
MATLAB	-	Matrix Laboratory
MBD	-	Model-Based Design
MLMA	-	Multi-Level-Multi-Agent
MLP	-	Multi-Layer Perceptron
MPM	-	Multi-Pattern Matching
MSA	-	Mass Search Algorithm
NN	-	Neural Networks
OSPCA	-	Oversampling Principal Component Analysis
PA	-	Profile Analyzer
PCA	-	Principal Component Analysis
PSO	-	Particle Swarm Optimization
RBS	-	Run-based Scheduling
RICT	-	Resilient Identity Crime Detection
RKNNWTSVR	-	Regularized Version of the KNN-Based Weighted Twin Support Vector Regression
RKNN	-	Reverse K-Nearest Neighbors
SD	-	Spike Detection
SLPSO	-	Social Learning PSO
SODRNN	-	Stream Outlier Detection Reverse K-Nearest Neighbors

SOMNN	-	Self organizing Map Neural network
SRPSO	-	Self-Regulating Particle Swarm Optimization
SVM	-	Support Vector Machine
TWSVM	-	Twin Support Vector Machine
VFML	-	Very Fast Machine Learner
WBS	-	Web-Based Simulation
WL	-	WhiteList
WSVR	-	Weighted Support Vector Regression

LIST OF SYMBOLS

$\mathcal{A}_1$	-	Acceleration of Elephant
y_{iter}	-	Coefficients of Acceleration
ζ	-	Constant
CV	-	Correct Value
$Euc_{ij}(t)$	-	Euclidean distance
μ	-	Gaussian width
Ai	-	Number of Attributes
$iter$	-	Number of Iterations
PAi	-	Previous Attribute Value
R	-	Random Number between 0 to 1 range
Δ	-	Target Value
$Mass_{active,j}(t)$	-	The Active Mass
$Force_{ij}^{d}$	-	The attraction force for each iteration
$E_{best_{iter}}$	-	The Best Path
$Esol_{best_{iter}}$	-	The Best Solution for the Elephant Swarm
$bestFit(t)$	-	The Best Fitness Value
$G(t)$	-	The Constant of Gravity
P_{iter}^{t}	-	The Current Position for Particular Iteration
G_{init}	-	The Initial Gravity Constant
$\mathcal{A}_i^d$(t)	-	The Law of Motion Acceleration
$Mass_{Passive\ i}(t)$	-	The Passive Mass
$weak(t)$	-	The Weak Fitness
$\mathcal{V}$	-	Velocity
$\mathcal{W}$	-	Weighting Factor

CHAPTER 1

INTRODUCTION

Data mining is a powerful technology that has made its way into many fields like science, engineering, commerce and industry for dealing with massive datasets. It can be used to support a wide range of business intelligence applications such as customer profiling, target marketing, workflow management, store layout, and fraud detection. It allows users to analyze data from many different dimensions. Technically, data mining is the process of finding correlations or patterns among a large number of fields in large relational databases.

Data mining is primarily used today by companies with a strong consumer focus-retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics. It enables them to determine the impact on sales, customer satisfaction and corporate profits. Finally, it enables them to "drill down" into summary information to view detail transaction data. With data mining, a retailer could use a point-of-sale record of customer purchases to send targeted promotions based on an individual's purchase history. By mining demographic data from comment or warranty cards, the retailer could develop products and promotions to appeal to specific customer segments. Various disciplines such as database technology, statistics, machine learning algorithm, pattern recognition, neural networks, data visualization and information retrieval are integrated by the data mining technique. Major components of the data mining: data warehouse, database and other information repository; based on the client request, server are responsible for fetching the relevant data. The interestingness of

resulting patterns is evaluated by knowledge base, to focus the search towards the interesting patterns are interacted by pattern evaluation module.

1.1 Categories of Data Mining

Data mining techniques are categories into many ways, which is given below:

Class description: It can be used to describe individual classes and concepts in summarized, concise, and yet precise terms.

Association analysis: It can be used to find of association rules for displaying attribute-value conditions which obtain frequently together in a given set of data.

Cluster analysis: Clustering analyses data objects to make a group, then find the cluster head. The objects are clustered or grouped based on the principle of maximizing the intra-class similarity and minimizing the inter-class similarity.

Evolution analysis: It describes and model trends for objects whose behaviors changes over time. It normally includes time-series data analysis, sequence or periodicity pattern matching and similarity-based data analysis.

Classification: It is the process used to classify the data into distinguishes data classes or concepts, for the purpose of being able to use the model to predict the unknown class of objects. The derived model is based on the analysis of a set of training data, and can be represented in forms like classification rules, decision trees, etc.

Anomaly detection: It can be used to identify the unusual data records, that might be interesting or data errors that require further investigation.

Outlier analysis: Outliers may be detected using statistical tests or using distance measures to fulfill with the general behavior of the data object.

Regression: It can be used to attempt to find a function which models the data with the least error.

Sequential pattern mining: It can be used to find a set of data items that occur together frequently in some sequences. It can extract frequent sub-sequences from a sequence database. It can be used for the basis of many applications, such as: DNA sequence analysis, stock trend prediction, finding language or linguistic patterns from natural language texts etc.

1.3 Various Application of Data Mining

Data mining tools and techniques can be used to take advantage of the historical data. Warehoused information can be filtered with the help of pattern recognition technologies, statistical and mathematical techniques in it. It can be used to find the relationships, trends, patterns, exceptions and anomalies that might otherwise go unnoticed. It can be used in the business to discover relationships and pattern in the data in order to make better decisions. It can be used to predict customer loyalty and develop the smarter marketing promotions. Specific uses of data mining include:

Market detachment– It is used to identify the common characteristics of customers who buy the particular products regularly.

Customer churn– It is used to predict the customer mentality who likes to leave your company and go to a competitor.

Fraud detection– It is used to identify the fraud based on the unusual behavior.

Direct marketing – It is used to identify prospects which should be included to obtain the highest response rate.

Interactive marketing– It is used to predict each customer's mentality based on accessing a Web site which is frequently accessed.

Market basket analysis– It is used to account what kinds of products or services are commonly purchased together.

1.4 Classification

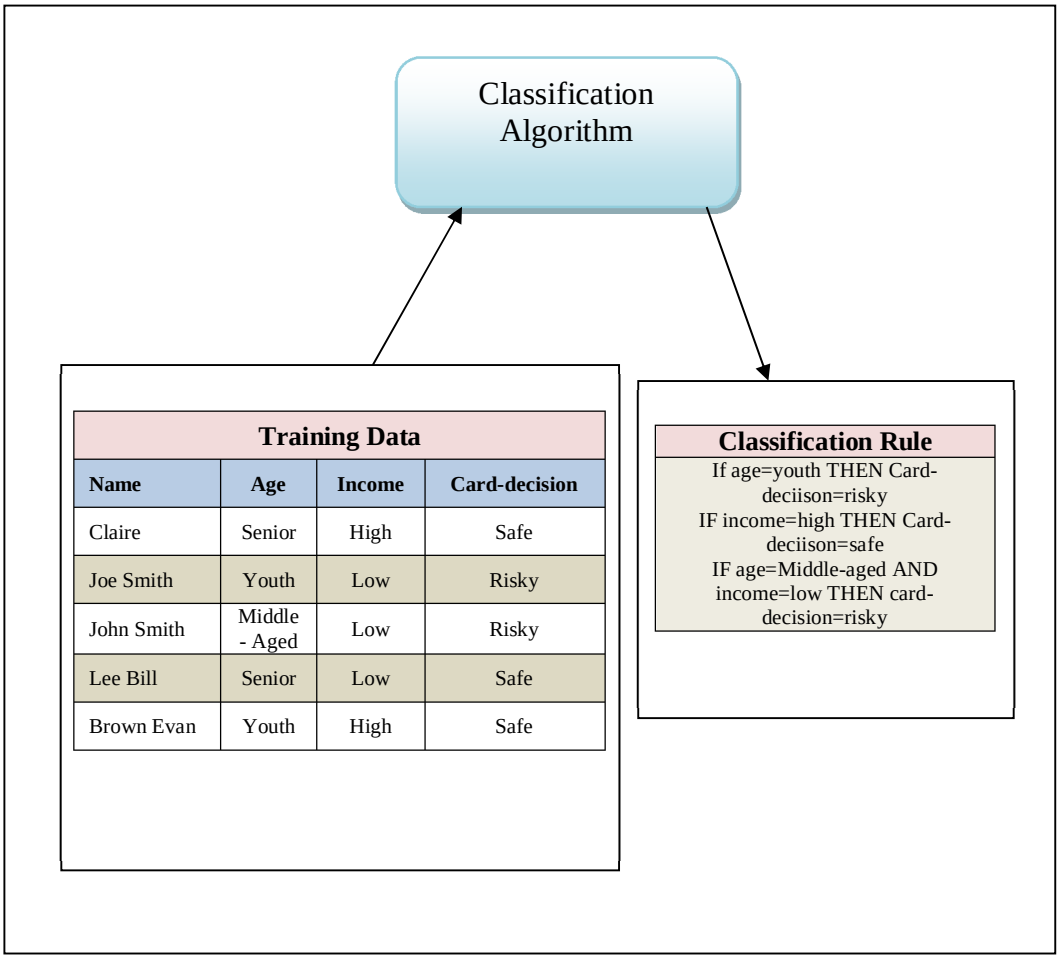

Figure 1.1 Learning Step

Classification is the process of data analysis that extracts models describing important data classes, which applies in a several areas, such as text classification, web page classification, and so on. A bank credit card officer needs to analysis of their data to learn which credit card applicants are “safe” and which are “risky” for the bank. Data classification is classified into two steps; learning step and classification step. Learning step is used to construct the classification model as shown in figure 1.1.

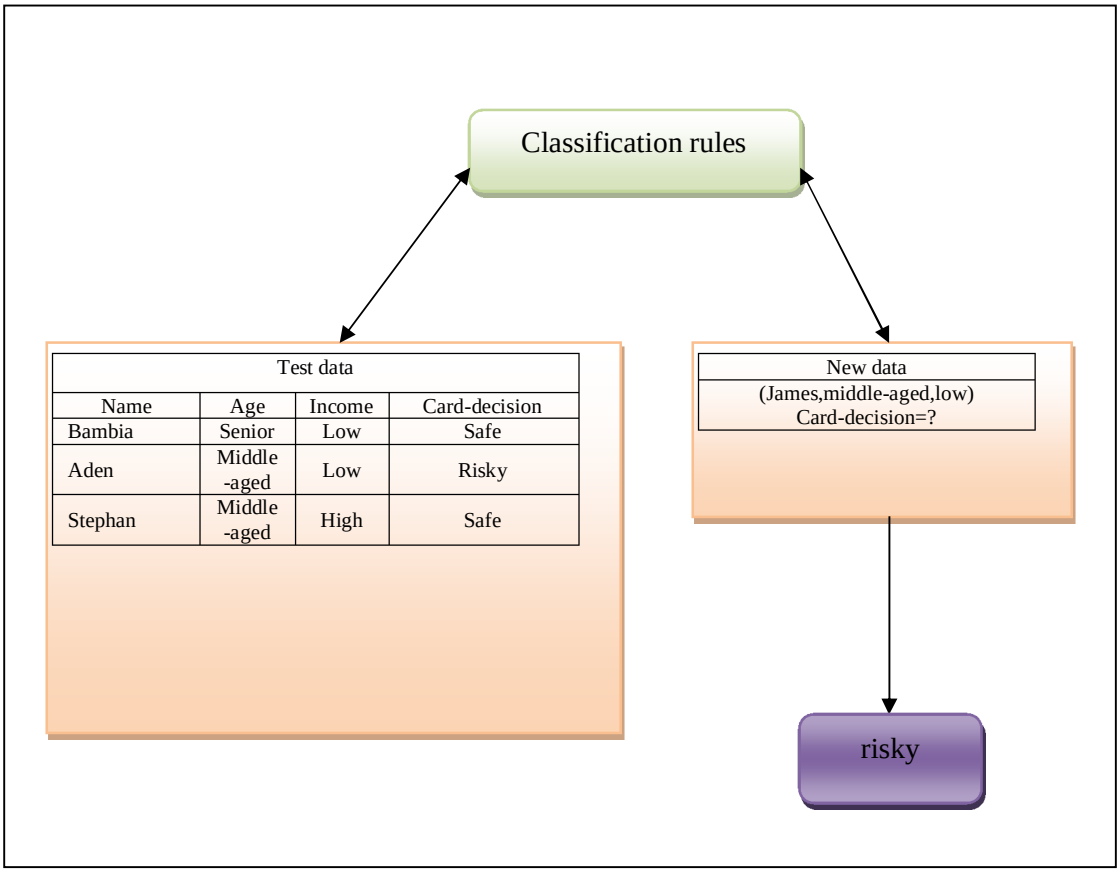

Figure 1.2 Classification Step

All the raw data are classificated into the useful information according to the specification. The classification is the one of the best ways to categorize the data into the useful format to make the decision as well as a prediction. So the classification steps are very necessity to make decisions and prediction process in the data mining. If the raw data are classified once, then the computation is very easy to make the result according to the classified

data. The classification step is used to predict class labels for giving data as shown in figure 1.2.

1.5 Classification Techniques

The machine learning terminology has been classified into the following categories given below in the figure 1.3.

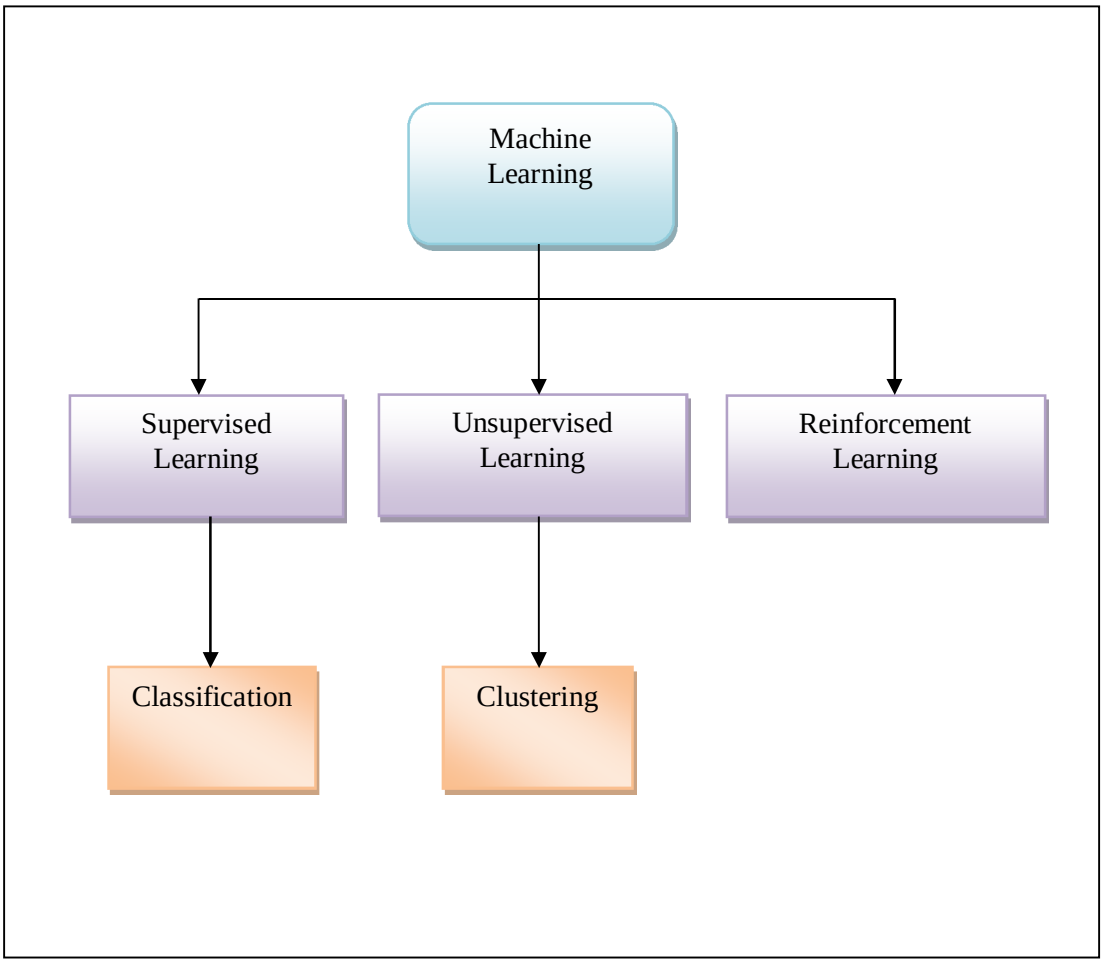

Figure 1.3 Machine Learning Classification

A supervised learning investigates the training data and generates the desired output, which can be used for mapping given data.

Unsupervised learning is the machine learning process of inferring a function to describe the unknown structure from unlabeled data.

Reinforcement learning is the machine learning for taking decision by behaviorist psychology.

Classification has numerous applications such as fraud detection, medical diagnosis and performance prediction so on. Evaluating and comparing the performance of the classification techniques are generally classified into major four techniques according to its applications as shown in the figure 1.4.

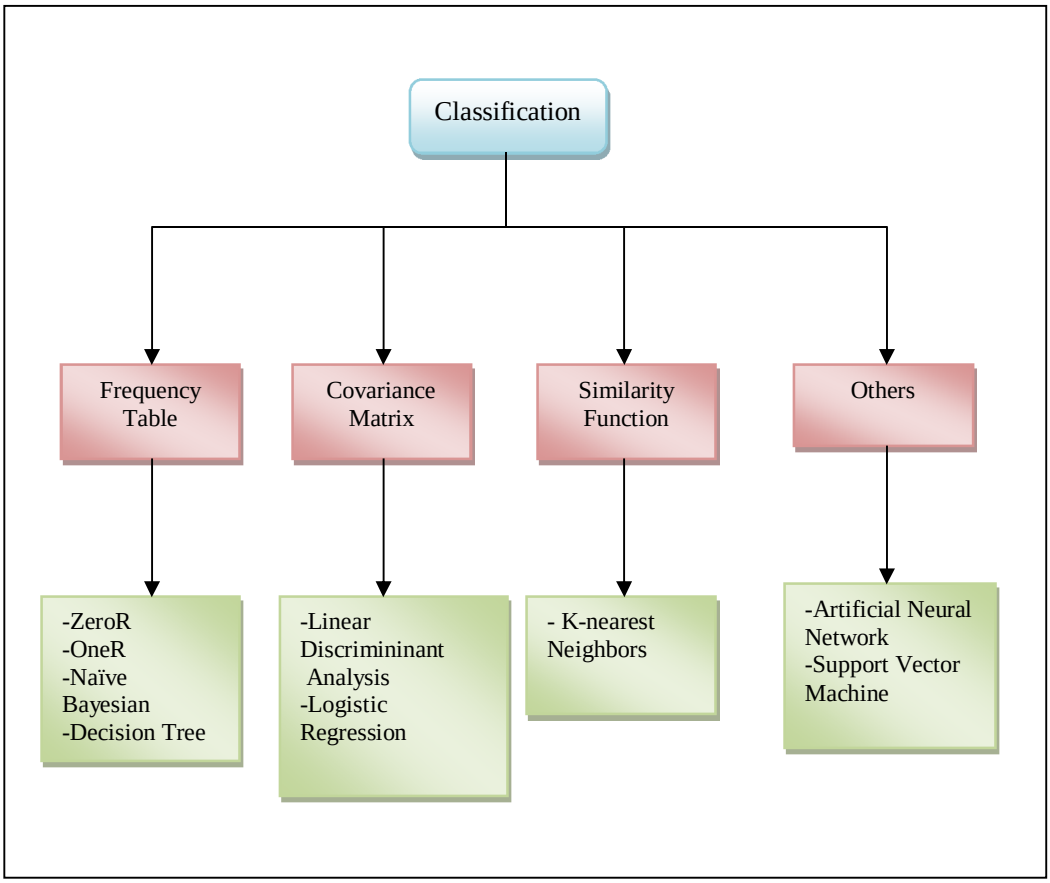

Figure 1.4 Categories of Classification Techniques

ZeroR: It is the simplest classification method which relies on the target and ignores all predictors. ZeroR classifier simply predicts the majority category (class). Although there is no predictability power in ZeroR, it is useful for determining a baseline performance as a benchmark for other classification method.

OneR: It generates one rule for each predictor in the data, and then selects the rule with the smallest total error as its “one rule”.

Naive Bayesian: The supervised learning method as well as a statistical method for classification is denoted by Naive Bayesian. It is based on Baye’s theorem with independence assumptions between predictions. It is easy to build with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Normally, it is used to provide the practical learning algorithm and prior knowledge about the problem. It is the easiest approach, which is used to understand and estimate the several machine learning algorithms.

Decision tree: It is like a tree structure. It consists of nodes. Attributes are represented by internal nodes, each node outcome is passed to next node by the branch and class labels are represented by terminal nodes. The decision tree is used to examine data and induce the tree and its rules will be used to make decisions. The output of the decision tree can be understood by the user as well as non technical person.

Linear Discriminant Analysis: It is a method which is used to find the linear combination of the features to differentiate two or more events in statistics, pattern recognition and machine learning.

Logistic Regression: Several statistical models are solved by data mining techniques. Logistic regression is the one of the standard statistical approaches to make the statistical model. It can be used for modelling binary data. The coefficients of the logistic regression can be used to estimate probability ratios for each of the independent variables in the model and it is applicable to a broader range of research situations than feature analysis.

K-nearest Neighbor: It is a non parametric method because estimation parameters are not involved in it. It includes two steps such as

inductive steps and deductive steps. **Inductive steps** can be used to construct a classification model from given data. **Deductive steps** can be used to test the model. Several anomaly detection is obtained by the concept of K-nearest neighbor. The K-nearest neighbor is the one of the best method that is used in the credit card fraud detection. It is used to support supervised learning algorithm. Three main factors are used to manipulate the performance of K-nearest neighbor; the nearest neighbors are located by distance metric, classification is obtained by it through rules and the new model is classified by the number of neighbors. High performance has obtained always by its rule. It can be improved the performance through a genetic algorithm. If the value of K is too large then it will reduce the noise data set. It consists of training data, such as genuine and fraudulent.

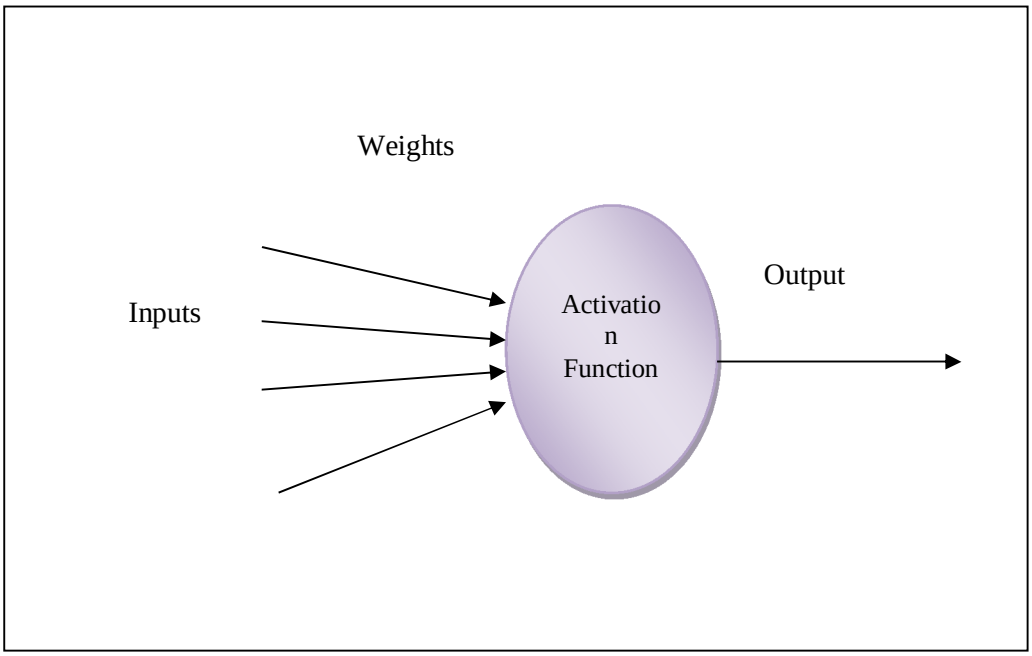

Figure 1.5 ANN

Artificial Neural Network: Biological neural system simulates to do the performance through the machine by Artificial Neural Network. The ANN can be used to compose of an interconnected assembly of nodes and directed link. It includes structure, learning ability and processing method of the human brain. It is used to process the information based on the combination of the

artificial neurons. The neuron computations are differentiated based on the weights. If the higher weight of neurons makes the input stronger. It is used in the several areas, namely forecasting, data compression, pattern recognition and so on. The general architecture of ANN is shown in the figure 1.5.

Support Vector Machine: Both linear and nonlinear data classification can be obtained through the Support Vector Machine. It belongs to class of supervised learning algorithm. It has hyperplane to separate two classes. It produces highly accurate. It can be used for numeric prediction as well as classification. It belongs to class of supervised learning algorithm. It is considered in the research work due to the nature of separating the two classes, which is used to detect and develop complex patterns in the given input. It is used in the several areas, namely text, handwriting, image recognition, Bioinformatics and so on.

1.6 Motivation

Credit card application fraud detection has been an extensive research area in the past two decades and still continues to be the area of interest in various sectors of data mining. Various detection algorithms and techniques have been proposed for fraud detection producing results with varying efficiency and accuracy. The algorithms range from a simple to varying levels of implementation and structural complexity, each employing their own detection techniques. Many techniques are found to be efficient to detect only a particular type of fraud. Further, what these techniques have in common is that they all aim at analyzing a single revision of software. Researchers are also in need to focus on fraud detection maintenance to assist the programmers.

The motivation behind this research is the fraud detection through various hybrid techniques. Moreover, the research focuses on detecting fraud through refactoring techniques and consistent modification of fraud.

1.7 Problem Statement

Many of fraud detection techniques have been proposed in literature. Most of the algorithms implemented using these techniques require other tools. Hence, they do not satisfy the detection challenges such as precision, scalability, adaptivity and quality data. For an instance, if one attribute is satisfied, then the other attribute fails. Also getting on to the basis of detection several techniques employs different forms of code representation and comparison granularities which are incomprehensible. The comprehension is the most important because they must be able to analyze, understand and appreciate the detection process along with the results obtained using these tools. These credit card application fraud detection techniques have some issues, like time constraint and extreme imbalanced class.These limitations in existing methods should overcome by three different approaches used in this research to locate the false information and place it into the boycott.

1.8 Objective of the Research

Having motivated by the limitations of the existing fraud detection techniques and algorithms, this dissertation presents hybrid algorithms aimed at detecting the fraud in the application stage of the credit card:

To implement the hybrid Fast Throughput Multi-Pattern Matching algorithm (FTMPM) that can detect the fraud in the application for the credit card along with CD and SD algorithm.

To implement an Improved Sheep Flock Heredity Algorithm (ISFHA) that can detect the fraud in the application for credit card with the Multi-Level-Multi-Agent (MLMA).

To obtain the optimal solution for the credit card application fraud detection by the Hybrid Elephant Swarm Optimization Algorithm (HESOA) to

solve the optimization and the classification, kernel based support vector machine algorithm is used. It gives the best accuracy results and shows the data handling capacity for large databases. Results show the accuracy of 99.32% in terms of detection which is comparatively improved compared to other existing methods.

1.9 Methodology

The objective of this research work is to develop and implement a hybrid fraud detection approach and to provide a support for detecting the fraud in the application of the credit card. In order to fulfill the desired objective, the research work has been carried out in three separate algorithms.

The hybrid Fast Throughput Multi-Pattern Matching algorithm (FTMPM) has proposed to match the large amount of attributes, in order to predict the fraudulent applicants with an appropriate time constraints. Together with the CD and SD algorithm that removes the redundant attributes and generates the credit score for the CIBIL list or black list. The CIBIL score varies about 300 to 900, it has been recorded in the credit history and by considering its range the credits are provided to the lender or customer, and they are added to the white list. This approach of credit card application fraud detection uses another database of Whitelist which stores the innocent applicants of credit card. Whitelisting uses real social relationships on a set of attributes. The experimental results of the proposed approach are compared with the existing approach results to compute the performance evaluation.

An Improved Sheep Flock Heredity Algorithm (ISFHA) has proposed to improve the efficiency of the credit card application fraud detection method by verifying and validating the optimized parameters, such as single and multiple attributes. The attributes of every application [offline/online] are verified by using a newly developed procedure MLMA-

[Multi-Level-Multi-Agent] and it is verified all the attribute values are best one or not. For optimizing the attributes the ISFH-[Improved Sheep Flock Heredity] algorithm is used and those attributes are validated according to the time and response with optimal value. The experimental results of the proposed approach are compared with the existing approach results to compute the performance evaluation.

To increase the speed of fraud detection, the hybrid elephant swarm optimization algorithm (HESOA) is proposed. The main aim of the work is to detect and prevent the credit card application fraud. In this method based on the attributes of credit card application validation is performed, the prevention of fraud transaction and for analyzing the system performance, the uses of hybrid elephant swarm optimization are proposed based on the heuristic search algorithm for the credit card fraud detection. The search algorithm is used to find the similarity among the neighboring attributes and elephant swarm optimization is used for finding the optimal path and best fitness. Proposed system gives the best accuracy results and shows the data handling capacity for large databases. Results show the accuracy of 99.32% in terms of detection which is comparatively improved compared to other existing methods.

1.10 Dissertation Overview

This thesis is organized in six chapters. Chapter 1 gives a preface to the proposed research work, motivation that led to the research, problem statement and a brief description of the basic concepts underlying this research work.

Chapter 2 depicts the review of state of the art works related to credit card application fraud detection research which are mainly classified as fraud detection. This chapter also presents the comparative analysis on various fraud detection and fraud prevention. The preliminary works that served as the main motive for most of our proposals have been narrated in this chapter.

Chapter 3 describes the design of the scheduling algorithm to detect the fraud in the credit card application. This algorithm takes the principle of fast response of pattern matching for scheduling. The results are compared with earlier algorithm and give the better achievement of the performance.

Chapter 4 describes the design of an Improved Sheep Flock Heredity Algorithm Based Prevention of Credit Card Fraud Detection for Online and Offline Transaction approach. The results achieved through the proposed algorithm are inferred and compared with the existing algorithm's results.

Chapter 5 describes the design the one more algorithm called a Hybrid Swarm Optimization Algorithm for preventing the fraud on credit card application, which requires less time for detecting the fraud. The results have been compared and analyzed in this chapter.

Chapter 6 presents the conclusion of this research work and scope for the further study has also been highlighted.

CHAPTER 2

LITERATURE SURVERY

An extensive literature associated with fraud detection in the credit card application research has been critically reviewed and presented in this chapter. A comprehensive review of literature on the classification of fraud, detect the fraud in the application and prevent the application fraud is presented. A comparative analysis of credit card application fraud and detection techniques is also presented. This chapter discusses the methods for identifying the application fraud in software development, maintenance and evolution. In addition, the summary of the review of literature is furnished to justify the scope of the present work.

2.1 Survey on the Credit Card Fraud Detection

All the online transactions are done by the computer in the name of E-commerce. E-commerce is the one the best way of selling or purchasing the particular product. The purchasing obtains in the E-commerce market by using either credit card or debit card. At the same time occurrence of the fraud also increased in the credit card usage. The credit card fraud is classified into two categories, namely application fraud and behavioral fraud. The fraudulent makes the deception, when applying the credit card is called as application fraud. The fraudulent makes the deception, when making the transaction is called as behavioral fraud. So some of the methodology should use to detect the fraudulent activity. Data mining technique is mainly used for credit card fraud detection. The increasing demand of credit card increases the complexity to detect the fraud in credit card. Conventional data mining methods are not much efficient to achieve better accuracy, which mainly relates to classification. Accuracy of classification is a key parameter in credit card fraud detection. To improve the accuracy, genetic algorithm is proposed for

credit card fraud detection. Since this is an intelligent algorithm, which helps to predict and improve the accuracy on the accuracy by optimizing the problem. Earlier frameworks for credit card fraud detection are based on the rules. These rules are defined by banks to detect the fraud transaction, but these methods induce trade-off in accuracy for a large number of transactions. Filippovet. al. (2008) was presented the Naive Bayesian Classifier with a cluster based model was utilized to perform the classification to overcome the problem. This system presents the credit card fraud detection method for online and offline applications. Limitation of the data is a key challenge in this area, because there is only one transaction is invalid out of thousands of transactions, which shows the probability of the fraud transaction. This issue can be solved by using intelligent techniques for data mining.

Chun Hua Ju, et al. (2009) proposed a new method for credit card detection by using the outlier method. The main approach of this work is to find the similarity in the data by using a coefficient sum method, but this method suffers from accuracy issues for large datasets. Very Fast Decision Tree learner (VFDT) method was proposed by Minegishi, T et al. (2009), since performance evaluation credit card data are considered. In this method by using a VFDT algorithm for the credit card data, they presented the results for the credit card fraud detection. Imbalanced data also considered for the performance evaluation. Rule based classification is the main drawback of these methods. To overcome these issues, an intelligent algorithm with genetic algorithm is proposed to detect the fraud in credit card transactions. There are numerous survey papers portraying the diverse sorts of fakes and distinctive misrepresentation identification strategies. Is a neural system based Visa misrepresentation identification which prepares a neural system with the previous information about the specific client spending conduct and the creators tried their product on the artificially produced information and tobuildup a multilayer feed forward neural system based misrepresentation location model for Mellon Bank.

Duman and Ozcelik (2010) proposed a mix of hereditary algorithm and diffuse quest for enhancing the credit card extortion recognition and the exploratory results demonstrate an execution change of 200%. Seyedhossein and Hashemi (2010) proposed a fraud recognition system which removes the natural example of a charge card time arrangement and uses this example for prior misrepresentation discovery. The work's majority found in the writing takes a shot at client spending conduct examination and some of them utilize some determined characteristics too. Be that as it may, we couldn't discover any exploration performed on the mysterious Visa exchange dataset where the determined characteristic idea falls flat. Hence, the target of this exploration was to build up a MasterCard misrepresentation identification model which can adequately recognize fakes from an imbalanced and mysterious dataset. So as to handle unknown information, which is the way of information by and large banks give because of security reasons, the proposed misrepresentation identification model considered every property just as without offering inclination to any trait in the dataset. Additionally, the proposed misrepresentation recognition model makes separate lawful exchange design and extortion exchange design for every client and in this manner changed over the imbalanced Visa exchange dataset into an adjusted one to take care of the issue of unevenness.

2.1.1 AIS Technique

Arunabha Mukhopadhyay et al. (2011) proposed the approach called as Fraud Detection System for Credit Card (FDSCC) that detects the online credit card frauds by An Artificial Immune System (AIS). The FDSCC is used to take the input in the form of binary string for detecting fraud by usingthe r-contiguous bit algorithm. FDSCC is based on the Biological Immune System (BIS) for detection and response.

The basic nature of the BIS interacts with the body to detect and eliminate a bacterium. Even the BIS has a perfect memory response for

identifying the structure of the disease. The FDSCC is extract the all the features from the BIS for detecting the credit card fraud. This approach also uses a similar form of memory based detection. The result of the FDSCC approach is more effective and efficient way of detecting the credit card fraud in the online.

2.1.2 FDS Techniques

The finance is purely based on the money. The financial status of a country only described about the wealth of that country. But many fraudulent behaviors happen in the financial department now a day. It is increasing day by day, so some of the remedial activities should be taken to avoiding the fraud, like either prevention measures or detection measures. The detection mechanism should be strong when the prevention mechanism fails.

Apparao, G. et al. (2009) proposed the approach called as a Generic framework for DM -Based FFD. Basically the FFD can be divided into four main categories, such as data distribution, learning types, detection approach and detection algorithm.

- In the **data distribution**, data usually formed into two groups, like fraud & non fraud company data and auditor data. Generally, data mining is used to follow two types of **learning process**; supervised learning method and unsupervised learning method. The classification is the best way to classify the fraudulent behavior of the supervised learning method.

- In the **detection approach** and **detection algorithm**, any one of the best classification technique, such as SVM, ANN, Decision Tree and Bayesian is used to detect the fraud in the finance to avoid the financial fraud, when the class label used. No class

label is there, then another method, such as K-mean, Genetic algorithm is used to detect the financial fraud. Future work of this approach is to identify the fraud in the various sectors like spam, intrusion, terrorism and so on.

2.1.3 RICT Techniques

The bank is doing one of the biggest tasks is fraud to avoid the unnecessary fund transfer. It is used to obtain the fraud detection in the two main areas, such as “application” and “transaction”, because most of the fraudulent activities happen in this area. General concept is that prevention is better than detection.

The credit card application consists of a large number of the set of attributes. Generally, the fraudster does the fraud by stealing the information from the unknown individual. They will use this particular information for applying the credit card. To avoid such kind of the activity, Clifton Phua et al. (2009) and (2012) proposed the approach to calculate the suspicion scoring to prevent in the fraud in the credit card application. This approach is called as communal analysis suspicion scoring for finding the suspicion score from the given application. It uses a technique called as pairwise matching. This technique is used to find the suspicion score for all incoming new applications for credit card. It is used to match the new application with existing application which is in the window. This approach describes three kinds of testing for detecting the fraud in the credit card application, such as Blacklist testing, Whitelist testing and Anomalous application-pairs testing. This approach is used to perform the testing in step by step testing.

In the first step, it will perform the blacklist testing, then in the second step, it will perform the whitelist testing. Following the whitelist

testing, it will perform the anomalous application-pair testing according to the new application.

Step 1: In the Blacklist test, a new application attributes are compared with the set of attributes in the blacklist to find the suspicion scoreby the equation (2.1),

$$\text{Suspicion score} = \begin{cases} 1 & \text{if } V_i \xrightarrow{\text{fraud}} V_j, \text{where } V_i = V_j \\ 0 & \text{otherwise} \end{cases} \quad (2.1)$$

Where the suspicion score is equal to 1, if a new application attribute is matched with blacklist window attributes, such as fraudulent application. Where the suspicion score is equal to 0, then it will go to step 2.

Step 2: In the Whitelist test, a new application attributes are compared with the set of attributes in the whitelist to find the suspicion score by the equation (2.2),

$$\text{Suspicion score} = \begin{cases} 1 & \text{if } V_i \xrightarrow{\text{normal}} V_j, \text{where } V_i = V_j \\ 0 & \text{otherwise} \end{cases} \quad (2.2)$$

Where the suspicion score is equal to 1, if a new application attribute is matched with whitelist window attributes, such as a normal application. Where the suspicion score is equal to 0, then it will go to step 3.

Step 3: If the new application does not match either the blacklist window or the whitelist window, then it passes to the anomalous application-pair testing to find the suspicion score by the equation (2.3),

$$\text{Suspicion score} = \begin{cases} 1 & \text{if } V_i \xrightarrow{\text{anomalous}} V_j, \text{where } V_i \neq V_j \\ 0 & \text{otherwise} \end{cases} \quad (2.3)$$

If a new application is a valid application, which is verified by one of the manual verification methods, then the suspicion score is equal to 1, otherwise it is 0.

This approach is very useful technique for the bank to avoid the risk and protect the unwanted fund transfer.

2.1.4 SOMNN Technique

Data mining is giving the compact fraud detection applications. It is used to provide the many well defined procedure for detecting the fraud in the many fields. It also gives the best procedure for detecting the fraud in the banking domain. Francisca Nonyelum Ogwueleka (2011) proposed the approach for detecting the fraud in credit card transaction based on the data mining application. This approach is based on unsupervised method on the Self organizing Map Neural network (SOMNN).

Generally, the Neural networks are trained by both the genuine transaction and fraudulent transaction to determine the accurate prediction. It consists of four clusters to differentiate the credit card transaction, named as low, high, risky and high-risk.When the transaction happens, it will fall into any one of these clusters. These clusters will analysis the transaction according to the behavior. Then it will decide whether the transaction is fraudulent or genuine.

In this approach, each transaction enters into the database for future security purpose. Because this approach is used to analysis the current transaction with the previous transaction behavior, which is available in the database.This approach makes some transaction for further more investigation if it is suspected as a fraudulent transaction. The main advantage of this approach gives more accuracy for detecting the fraudulent transaction.

2.1.5 CD Technique

Clifton Phua et al. (2006) proposed the communal detection (CD) algorithm to detect the real community of interest. This algorithm is used to identify the applicant against the trained dataset within a set window and to produce the suspicion score for categories the applicant in the manner of black, white and anomalous. This algorithm used to extract the input from the database which is in unsorted form. Then it should be sorted by descending order of arrival date and time. After collecting the input, the algorithm works in 5 steps to produce the output as given below:

Step 1: determine the single link attribute vector for each current trained dataset within a set window.

Step 2: determine single-link approximate communal score.

Step 3: determine the single-link current weight

Step 4: find out single link average previous suspicion score.

Step 5: by attaching the current weight and average trained scores of all single links to determine the multiple link suspicion score for each current applicant.

This algorithm has measured the performance through the FME curves, because the FME curve has some unique quality, such as it will work under multiple-values. It will perform under different threshold. But this algorithm has some limitation such as unnecessary usage of temporal weights.

2.1.6 Hybrid BLHA Fraud Detection Technique

The banks are ready to issue the credit card with a lot of security. Even the fraudulent makes malicious activity to break the security. By this action, the fraudulent cheats the bank as well as the customer. Amlan Kundu, et al. (2009) proposed the algorithm for credit card fraud detection called as Hybrid BLAST-SSAHA for Credit Card Fraud Detection, which is used to

obtain by two –stage sequence alignment, such as Profile Analyzer (PA) and Deviation Analyzer (DA). The Profile Analyzer is used to verify similarity of the given credit card holder regarding the past transaction. If any unusual transaction obtains, then it passes to the Deviation Analyzer for identifying the deviation.

The hybrid BLHA is the combination of the BLAST and SSAHA for credit card fraud detection to improve the efficiency and accuracy.

The BLAST is working in three steps,

Step1: collect the list of high-scoring words.

Step 2: compare the each high-scoring word with the database to record the hit.

Step 3: set the threshold value to find the similarity score, such as if it is equal or greater than then retain it, otherwise stop it.

The SSAHA is working in two steps,

Step 1: to construct the hash table, this is main memory.

Step 2: search from the hash table by the query word.

But the SSAHA has some limitation for getting accuracy, which is solved by combining the BLAST and SSAHA is called as Hybrid BLAH for credit card fraud detection. The BLAH consists of a two-stage algorithm, such as

Stage 1: to create the Clustered K-tuple.

Stage 2: to find the similarity.

The BLAH fraud detection system consists of the components to detect the fraud in the credit card, which is given below,

Profile analyzer (PA): To find the time sequence for analyzing the resemblance of the incoming transaction with a profile database of the cardholder is called as Profile Analyzer (PA). It is used to evaluate the resemblance score between Time analyzer and customer profile database is known as Profile Score (PS).

Deviation Analyzer (DA): To find the time sequence for analyzing the resemblance of the deviated transaction with the fraud history database is called as Deviation Analyzer (DA). It is used to evaluate the resemblance score between deviated sequence and fraud history database is known as Deviation Score (DS).

Amount Sequence (A): To calculate a sequence of transaction amounts correlated with the last few transactions on the particular card is called as Amount sequence (A).

Time Sequence (T): To calculate a sequence of transaction times associated with the last few transactions on the particular card is called as Time Sequence (T).

Time-amount Sequence (TA): Combination of the Amount Sequence (A) and the Time Sequence (T) is called as Time-amount Sequence (TA).

Deviated Sequence (V): Deviated sequence (V) is used to evaluate the deviation transactions from the cardholder's profile database, when the fraudulent transaction executes.

K -tuple **Table** ***(KT):*** The history database has two kind of the information, such as sequence-index and sequence-offset, which is available in the K-tuple Table (KT). The history database information is further classified

into two types, like genuine and fraud. The CKT holds the genuine cardholder's information. The FKT holds the fraudulent information.

Final Decision Maker (FDM): The final decision is taken from the nature of the transaction by the final decision maker based on the profile sequence and deviation sequence.

The hybrid BLAH Fraud detection system is used to detect the fraud in the transaction very fast compared to the other detection system. It is used to provide more accuracy. It is also used to detect the fraud in the many areas, such as telecommunication fraud detection and banking fraud detection.

2.1.7 DST and BL Techniques

Suvasini Panigrahi et al. (2009) proposed the fraud detection system for detecting the fraud in the credit card usage. The fraud detection system has four major parameters to detect the credit card fraud, such as rule-based filter, Dempster–Shafer adder, transaction history database and Bayesian learner.

- The first level of fraud detection is encountered by the rule-based filter. It is used to avoid the fraudulent transaction. It consists of the filter component, which filter the mismatched transaction and find the suspicion level. The result of the rule based filter is given to the next level parameter.

- The second level of fraud detection is encountered by the Dempster-Shafer adder. It is purely based on the mathematic theory of belief function. After taking the input, it is used to determine the overall faith in the each incoming transaction. Each transaction is unique transaction, so the rule based filter

generates the various suspicion levels for each transaction. The Dempster-Shafer adder takes the result of the rule based filter and provides the security of each transaction independently.

- The third level of fraud detection obtains in the transaction history database. This database has two kinds of record, one is belongs to normal transaction is known as good transaction history and another one belongs to a fraudulent transaction is known as fraud transaction history. Each transaction has some set of attributes to obtain the transaction. Each attribute has the information, such as card number, OPT number, time of purchase, amount of purchase and so on. The database stores the all the transaction's attributes for future reference. If any transaction may deviate from the normal process, then it will be treated as fraudulent.

- The next level of fraud detection obtains with a machine learning tool called as Bayesian learner, which is used to find the optimal decision. When each transaction occurs, Bayesian learning is used to update the suspicion score immediately. The main objective of the Bayesian learning makes the probability for each transaction to update the suspicion score.

For example, once the particular transaction declares as fraudulent, then it will be inserted into the fraud transaction history. But this approach will not do any action according to the particular card until the next transaction arises on the particular card. When the next transaction arises on the same card, then it will be blocked by this approach. This approach is very flexible for accepting any new technique, framed for the rule-base component, so that it improves the performance effectively and accurately. At the same time,

compare to other approaches, it will eliminate the false alarm. The simulation result of this approach obtains like 98% True Positive and 10% False Positive.

2.1.8 FRST Techniques

The machine learning algorithm is used to compare the current data with the trained data. This comparison takes more time to find the match. Ying Yan et al. (2010) proposed the FRST approach to find the pattern matching with less response time. This approach is used to find the response time by detecting, analyzing, and responding, because the response time is a more important parameter for analyzing the efficiency of an algorithm. This approach is the hybrid of RBS and EBS for making the dynamic scheduling to reduce the average response time.

This approach is used to obtain the pattern matching to the huge numbers of concurrent queries with the rapid response time. The EBS is the process, which is based on the event occur during the execution. Before the process will enter into the new event, it completes the test with the previous event for possible state transition. All the events are examined by the multiple runtime patterns in this approach. The RBS is the process, which is used to give a result either accepted or rejected. According to the result, it will take next runtime pattern to process. All the events are in the buffer in this approach. It will be revisited multiple times for the execution. Each event is examined by runtime pattern. When current event is examined by multiple runtime patterns, the EBS gives the rapid response time for current event before appearing the next event. Even this EBS is faster than the RBS. When the high rate streams occur, then the RBS is more effective. This approach uses both the scheduling methods for improving the performance to reduce the response time.

The result of this approach has more effective and efficient, because it obtains the scheduling dynamically. This approach is very useful for many

real time applications, such as credit card fraud detection, RFID tracking application and so on.

2.1.9 SODRNN Technique

The bank revenue is very important for a country's growth. But now a day, banks may face more problems for handling fraudster, especially in the credit card division. The fraudster may use many ways to commit the fraud. Venkata Ratnam Ganji et al. (2013) proposed the approach that detects the credit card fraud using anti-K nearest neighbor algorithm. In general, the fraudster may use many methods to commit fraud. But this approach describes the two common techniques for committing fraud is given, namely the Traditional Techniques and the Modern Techniques, which are described below,

Traditional Techniques: Traditional Techniques is denoted the "Application Fraud". When an individual uses another individual's personal detail for applying a credit card to commit fraud is called the application fraud.

Modern Techniques: When an individual uses another individual's card for some reason while the owner of the card are not aware of that.

This approach adopts the unsupervised learning algorithm, because the unsupervised learning algorithm gives better result than the supervised learning algorithm. The unsupervised learning algorithm is used to predict the fraud without getting the prior knowledge about the fraudulent and non fraudulent transactions in historical database. But in the case of, the supervised learning algorithm is used to identify the fraud by differentiating between fraudulent and non-fraudulent behavior. The supervised learning algorithm requires the fraudulent transaction in the historical database to get accurate results. This approach used the algorithm named by the Stream Outlier

Detection based on Reverse K- Nearest Neighbors (SODRNN). It consists of two methods, such as the Stream Manager and the Query Manager.

- **The stream Manager** consists of the three operations, like insert, update and delete. When inserting the new data stream, it will scan the current window by one pass and update to the KNN-list and RKNN-list of the influenced objects. While updating also, the data stream is to update to the KNN-list and RKNN-list of the influenced objects. But when deleting, the data stream deletes either from the KNN-list or from RKNN-list of the influenced objects.

- **The Query Manager** is used to demand the query of the top 'n' outlier, when it will do the scan of the current window and return the 'n' objects.

This approach takes both real as well as synthetic data set for showing the efficient and effective results for detecting the credit card fraud.

2.1.10 CEM for EFD Technique

In the recent years, online trading has become more popular. It also makes more profit. If the trading participants are trusty, then it has made good revenue. But auction cannot expect that all is legitimate. So improve the security and detect the fraud in the online auction, Jau-Shien Chang, et al. (2012) proposed a cost effective method for early fraud detection. The cost effective detection method consists of set of detection process without prolonged computation and difficult data downloading. It maintains the detection accuracy, which is helpful for developing the early fraud detection system. It consists two methods; **Measured Attributes Reduction**, **Late-Profiling for Constructing Detection Models**.

- **Measured Attributes Reduction:** Measure Attributes Reduction is used to build the detection model using a Machine learning algorithm for a given attribute. This method is used to refine the attributes by using the Principal Components analysis. The Principal Components analysis selects the main attributes from a given set of attributes. For filtering attribute noise, this Principal Components Analysis is used to select the eigen vectors. If the attributes have worst eigen vector, then that attribute is eliminated.

- **Late-Profiling for Constructing Detection Models:** The Late-profiling method is used to characterize the behavior of the fraudsters. In general, the Late-profiling measures use the partial history to detect the fraud. Normally, to detect the theft is too hard and difficult. But the Late-profiling measures give the best way to identify the thief with partial transaction histories.

The cost effective detection method is used to update the detection model with the fixed interval for increasing the high accuracy of detection by identifying the thief, who did the fraudulent activity during his first attempt.

2.1.11 Survey on Data Mining for Credit Card Fraud

Siddhartha Bhattacharyya et al. (2010) proposed the comparative study for credit card fraud. This approach describes two data mining techniques to detect the fraud in the credit card operations, namely SVM and RF work along with LR. LR is used for categories the fraud by using the binary choice model.

SVM is the one of the statistical technique available in the data mining., which is used to classify the problem according to the task. It has two important properties for making the classification, namely margin

optimization and kernel representation. RF is the one of the trendy methods for making decisions in the data mining technology, which consists of the trees. Those trees are built indenpent for each other. The large number of trees is grouped together in it. Association between the trees and the strength of each tree in the group decribes about the error rate of the RF. RF provides better performance regarding the prediction than the SVM. Recently, the RF is used in several fields, namely image classification, several bio-medical problems, banking domains and so on. This approach also has some limitations as given below:

- If the fraudulent and legitimate transaction occurs simultaneously, then the approach doesnot able to produce the proper prediction regarding the fraud.

- The non-availability of exact time stamp data beyond the date of credit card transactions are the other limitation.

2.1.12 HMM

Generally, the anomaly detection as well as the intrusion detection system includes the HMM for improving the performance and modeling time. Abhinav Srivastava et al. (2008) proposed the approach to detect the fraud, when the multi-stage network attack obtains in the credit card transaction. This approach is used to train the dataset based on the behavior of the applicant. If the applicant behavior is mismatched that it, reject it, otherwise it considers that the applicant is genuine. This approach has more effective than the other approaches. This approach uses the standard metric for detecting the fraudulent activities, namely TP and FP. The fraction of fraudulent transactions is denoted by TP. At the same time, the fraction of genuine transactions is denoted by FP. The difference between the TP and FP is

denoted by TP-FP spread, which is showing the performance efficiency. Accuracy of this approach is near to 80 % compared with other approaches.

2.1.13 AGA

Due to the fraudulent activities, many countries loose the economical growth. Rinky et al. (2013) proposed the Artificial genetic algorithm (AGA) for detecting and preventing the credit card fraud. Generally, genetic algorithm is the one of the optimization methods in the data mining. It is used to select the optimal solution and reject the unwanted one for a given problem. It consists of three operators for making the optimal solution, namely selection, recombination and mutation.

Given problem consists of several solutions, but selection is used to select the best solution. Recombination is also called as cross over, which is used to combine two or more parental solutions to form a new optimal solution. While making the recombination, mutation performs the random toddle in the surroundings of the chromosomes. The objective of an artificial genetic algorithm is to improve the better solution for the given problem by making the decision.

The working principle of the AGA is given below:

Step1: to receive the input attributes and keeps the data set confidentially.

Step2: to calculate the critical values, the CC usage frequency count, CC usage location, CC overdraft, current bank balance, average daily spending.

Step3: to implement the detection mining based on the critical values to detect the fraud transaction.

AGA aims to analyze the feasibility of credit card fraud detection based on technique, which is used to develop the security for money transfer that can detect whether the transaction is fraudulent or not.

2.2 Survey on HPSO-LS

Parham Moradi et al. (2016) proposed that a novel hybrid feature selection algorithm based on particle swarm optimization for feature selection. HPSO-LS were utilized a subset size determination for feature selection with reduced size. It was comparable with four filters based method such as minimal redundancy maximal relevance, fisher scores term variance and information gain respectively and wrapper-based method such as particle swarm optimization, genetic algorithm, ant colony optimization and simulated annealing respectively.

According to the Comparison result, this method was improved the classification accuracy by using the following steps is given below:

Step 1: determine size of the feature

Step 2: grouping

Step 3: initialize the particles

Step 4: update the particle position

Step 5: local particle search

Step 6: fitness determination

Step 7: upload the best solution

Step 8: stopping criterion. Repeat from step 4 when the fitness is not upto the level, otherwise goto step 9

Step 9: report the feature set.

On the classification task, feature selection takes an important part with the minimum computational cost and improve the classifiers ability. HPSO-LS combined with new local search operations with the global search

process of PSO. The combination of the local search operations and global search strategy provided the better classification result. It gave the better classification accuracy, less execution time and size of subset of selected features.

HPSO-LS method has more advantage for finding accuracy of the classification. Even though it has some limitations that is given below: To enhance the classification accuracy as well as reduce the number of selected features to integrate with the multi-objective feature selection optimization framework. By clustering the features into several groups to improve the local search operations.

2.3 Survey on RKNNWTSVR

One of the powerful kernel-based machine learning tools for pattern classification and regression problem is the support vector machine (SVM), which is the most used method for the pattern recognition and classification. SVM can be used to implement a wide variety of applications. But it takes more time to do the computation. Various algorithms are proposed to reduce the computational time of SVM, such as Weighted Support Vector Regression (WSVR), Generalized Eigenvalue Proximal Support Vector Machine (GEPSVM) and Twin Support Vector Machine (TWSVM). Weighted Support Vector Regression (WSVR) can be used to combine the weights to the slack variables in the objective function. This process can be reduced the size of the coefficient of each surveillance in the estimated functions, and thus it is widely used for minimizing the influence of outliers.

Twin support machine consists of the two hyperplanes that two hyperplanes are non-parallel proximal hyperplanes. Each hyperplane is nearer to one of the classes and away from other class. So that TWSVM produced more effectiveness than SVM and GEPSVM.

All the above algorithms such as TSVR, WTSVRAND and KNNWTVR have primal problem in experimental risk. To eliminate the primal problem, Tanveer M, et al. (2015) proposed the regularized version of the KNN-based weighted twin support vector regression (RKNNWTSVR) to improve the efficiency and effectiveness. In this algorithm, this algorithm replaced the 1-norm of the vector of sack variable by 2-norm to make objective functions strongly convex. The outcome of this algorithm is used to improve the performance, reduce the computational cost for simple system's linear equation. It consists of several parameters for producing the effectiveness and efficiency. But it has some limitation, such as it didn't describe about the parameter selection. Because selection for an exact parameter K is very important for the performance improvement. At the same time, the few parameters also miss during the evaluation.

2.4 Survey on the PSO

The real time problem has more complexity of obtaining the optimal solution. Getting optimal solution is the biggest research target for a number of decades. Many researches are used to develop the algorithm for getting the solution which satisfies following parameters like accuracy, time constraints, space constraints, effectively and efficiently, which kind of the solution is known as the optimal solution. The number of numerical techniques is used to find the optimum. Several optimization algorithms are available in the recent years. One of the best algorithm is called as the PSO for making the optimization, because it is a very simple and take less computation. The problems are solved in the PSO based on the mathematical methods. It is a population based heuristic search. It is very useful for several real time complex problems. Each attribute of the problem are represented by the particle. It is used to update the search pattern by two ways, either the investigation of the search space or the utilization of the better solutions found. Many researches are used to develop the optimization algorithm based on the principle of the PSO, such as FIPS, DMSPSO, CLPSO, SLPSO and so

on. The particles are self regulated by the PSO is known as SRPSO, which provides the best solution for poor performance particles. This also has some limitation, Tanweer M R et al. (2016) proposed the hybrid with the directionally driven method, which overcome the existing approach limitation, is called as DD-SRPSO. This approach has two technologies, such as a DRS and a RIS. Generally, the particles are grouped into two categories, namely good performance particles and poor performance particles. The poor performance particles are grouped together, then it is regulated to make directional updates by the good performance particles. All the remaining all the good particles are randomly chosen to make direction either by DRS or RIS.

The result of this approach is compared with several earlier PSO, namely CLPSO, SLPSO and so on. This approach gives better performance and more accurately. This approach also provides more effectiveness as well as efficiency. The speed of the approach also too faster than all other existing approaches.

2.5 Survey on OSPCA

Anomaly detection is the one of the important technique to detect the deviated data instances in the applications such as credit card application and so on. The following technologies are used to detect or identify the deviation from the applications:

PCA is used to obtain the principal directions to detect the deviated data instances from the given instance. Normally it is used to do two operations for identifying the deviated data instances, that is constructed and calculation; construction operation is used to construct the data covariance matrix. Calculation operation is used to determine its eigen vectors. Because the eigen vectors are the most informative vector in the given data space. So

this approach is more expensive to calculate the eigen vector and covariance matrix. Memory usage is also more in it.

Decremental PCA is used to avoid the limitations of the PCA, one more method to do the anomaly detection for the given data space is called decremental PCA. It is used to detect the deviated data instance for the moderate dataset. But current scenario has a large amount of data. So this approach is not suited for it.

OSPCA is used to overcome the decremental PCA drawback,Yuh-Jye Lee et al. (2013) proposed a new approach that is Oversampling Principal Component Analysis. This approach is replica the target instance multiple times, because to intensify the result of different behavior rather than the normal data for the large dataset. It doesn't calculate the multiple eigenvector. Instead, it extracts the principal direction through online. Even it doesn't construct a covariance matrix. The result of this approach is reducing the computational costs and memory space.

The benefits of this approach work on online large size dataset. Handling high dimensional data is too hard by many machine learning algorithms. But it is able to handle the high dimensional data without computing the covariance matrix.

2.6 Survey on Comparison with Parametric Optimization

Manoel Fernando Alonso Gadiet al. (2008) proposed a comparative study about the credit card fraud detection. In this approach, the credit card fraud detected by the five classification methods, such as Decision tree (DT), Artificial Immune Systems (AIS), Naive Bayes (NB), Bayesian Nets (BN) and Neural Nets(NN). The fair comparison is obtained by two steps,

- In the first step, adjust the parameters for each algorithm either through an exhaustive search algorithm or through the Genetic algorithm.

- In the second step, make the comparison with two modes, such as a cost sensitive training mode and a plain training mode.

In this work, tested resultsshow that Bayesian Nets (BN) are better than Neural Nets (NN) and Decision tree (DT), Artificial Immune Systems (AIS) also better than Bayesian Nets (BN) and Neural Nets (NN). The best results are obtained with Decision tree (DT), the second best results are obtained with Artificial Immune Systems (AIS), next Bayesian Nets (BN) and Neural Nets (NN), follows equally, and finally Naive Bayes (NB).One of the limitations of this approach is not implemented the Support Vector Machine (SVM) for getting a better relationship between the values of each parameter, the imbalance of the data and cost sensitiveness.

2.7 Survey on Application of Classification Models

Usage of the credit card is increased in recent days. At the same time, the credit card fraud also increased out of control. To avoid the fraudulent activity, Aihua Shen, et al. (2007) proposed the statistical report on the classification method for the credit card fraud detection. This approach described three classification methods for the credit card fraud detection, such as Decision Tree (DT), Neural Networks (NN) and Logistic Regression. With the strategy, the decision tree divides the complex problem into many sub-problems and conquers the sub-problem through repeatedly using. Even though, decision tree has many merits, like highly flexible, good haleness. Rumelhart (1986) focused that to form the Neural Network architectures by organizing nodes into layers and linking these layers of neurons with modifiable weighted interconnections. The neural networks are the one of the best supervised machine learning methods. It can produce the summary of

internal principles data from the given data without knowing the potential principle data ahead. Even though it has some limitation, like difficult to form the structure, excessive training, and so on.

Logistic regression is similar to linear regression. This method is very useful to predict where the presence or absence of a characteristic based on values of a set of predictor variables. This approach describes the overall performance comparison result in the table 2.1 given above.

Table 2.1 Comparison Performance Results of the Classification Methods

Deciles	Lift value by Deciles			Cumulative Lift by Deciles			Base line
	NN (sec)	Logit (sec)	DT (sec)	NN (sec)	Logit (sec)	DT (sec)	
0.1	5.88	5.84	3.89	5.88	5.84	3.89	1
0.2	1.33	1.02	1.95	3.61	3.43	2.92	1
0.3	0.66	0.66	0.97	2.63	2.51	2.27	1
0.4	0.66	0.53	0.93	2.13	2.01	1.94	1
0.5	0.44	0.58	0.27	1.80	1.73	1.60	1
0.6	0.49	0.62	0.58	1.58	1.54	1.43	1
0.7	0.22	0.18	0.71	1.38	1.35	1.33	1
0.8	0.31	0.49	0.71	1.25	1.24	1.25	1
0.9	0.00	0.09	0.00	1.11	1.11	1.11	1
1.0	0.00	0.00	0.00	1.00	1.00	1.00	1

The summary of this approach provides the clear idea about the classification method for the credit card fraud detection. The neural network and Logistic regression are performing better than the decision tree. This approach finds the fraud using the past transaction behavior of the fraudulent. This approach is very useful for the bank to avoid the risk regarding the fraud and improve revenue of the bank.

2.8 Survey on Natural Neighbor

Data mining handles the pattern matching by Pattern-recognition techniques, which are handled by the two popular approaches, such as K-nearest neighbor (KNN) and reverse k-nearest neighbor (RKNN). Those two approaches provide better performance for pattern recognition. Both of these approaches have the one of the best characteristic is simplicity. Due to this characteristic, which are often used and produce the effective results. But both of the approaches have one common difficulty, namely determine the parameter K. Depending on the parameter K, those approaches produce the effective and efficient results for pattern-recognition.

Suppose, if the K value is large, then it leads to short-circuit errors. If the K value is small, then it leads to reduce the association of neighborhood. For this above reason, these approaches have the biggest task, like to identify the optimal value for parameter K. This identification makes big overhead for the research, which is overcome by QingshengZhu, et al. (2016) proposed a new approach is called as A self-adaptive neighborhood method without parameter K. This approach is also known as a natural neighbor (NaN), not only to estimate the parameter k, but also to discover a new way of nearest neighbor method without the parameter of k.

The main aim of this approach deals with the friendship between the human. Even it is a scale free nearest neighbor technique. Here the current state is mentioned by three contributions:

- The NaNE is used in this approach instead of the parameter K. The NaNE is vigorously selected for different data sets, which retains its permanence after repeated trials of the data set of different scale. These experimental results exhibit the strength of the proposed approach.

- The dynamic numbering is used to identify the each point, where the range between 0 and NaNE, which is also supple.

- The NaN technique can generate an applicable neighbor- hood graph based on the local characteristics of various data sets.

The NaN is the best approach for classification and outlier detection without selection of parameter K value, which is used in the several areas like image segmentation, face recognition and so on. It also has some drawback due to the complexity in the computation.

2.9 Survey on CS-LDM

The machine learning approach has several numbers of the classification techniques. The SVM is the one of the best classification methods for supervised learning technique. The role of SVM is used to maximize the minimum margin to achieve the best performance. If the maximize the minimum margin leads to weak generalization performance based on the Breiman's approach. This was overcome by Reyzin, et al., proposed the maximizes the minimum margin. But it suffered from the weak margin distribution.

To avoid the above mentioned problem by Zhang, et al. (2010) proposed the LDM, which minimizes the margin variance and maximizes the margin mean. The LDM is the better approach than the SVM for the classification problem. But this approach is not suitable when training data is imbalanced. The imbalanced class data are the one of the major overhead of the researchers in the detection techniques. Normally, LDM is used to obtain the balanced classification, which is compared with SVM to provide a better detection rate. The lower detection rate occurs, when the LDM absences.

Fanyong Cheng, et al. (2016) proposed the new approach called CS-LDM, which improve the detection rate of the minority class. This approach consists of cost-sensitive margin mean and cost-sensitive penalty, This approach is used to obtain the balanced classifier, at the same time it increases the detection rate with increasing of the cost parameter. The CS-LDM produces the result with increasing the detection rate when the parameter is inthe imbalanced class.

2.10 Summary

The credit card application fraud detection is an active research area for two decades, initially on fraud detection and analysis and later on slowly migrated to credit card application fraud detection, which in turn made to focus on the prevention of the application fraud.

In this chapter first, a comparison and analysis of all currently available fraud detection techniques and tools are presented. Most of the techniques denote the fraud detectionin the transaction time. In fraud detection, recently the researchers focused on hybrid techniques, to increase the efficiency and speed of detection.

Second, the different credit card application fraud detection techniques and tools which are currently available are discussed. However, there are no common parameters for the comparison, to analyze the techniques in different aspects. Therefore, the comparison of the techniques is made with different parameters which are appropriate for its subdivisions. In the credit card application fraud detection, until now research is on for newer explorations.

CHAPTER 3

FAST THROUGHPUT SCHEDULING MULTI-PATTERN MATCHING

3.1 Introduction

Generally, Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. This information can be used to increase the revenue, cut costs and both. Data mining software is one of the analytical tools that allow data from various categories to make a feature. Data mining in-turns is the process of finding patterns or correlations among the enormous data.

In recent years, the Data mining concerns with the fraud detection process since the problem here approached is the credit card fraud. This system identifies the fraudsters and does not give a chance in the credit card application. The credit card fraud becomes more prominent, because there are large amount of data similar to other data. The fraudsters can easily make a fake account in order to get a credit card application. There may be two ways of data used by the fraudsters one refers to be plausible, but identical data of another customer which is effortless to create but more difficult to apply successfully. The other is the real identity theft that is illegal use of innocent's data.

The credit applications are paper-based forms or the online application form to request by the potential customers for credit card, mortgage loans, and personal loans. In this case of credit card fraud, count of fraudster's increases that are highly experienced, organized and sophisticated. Their visible patterns can be different to each other and constantly change.

They are persistent, due to higher financial rewards the risk and effort involved are minimal. Based on the anecdotal observation of experienced credit card application investigation fraudsters can use software automation to manipulate particular values within an application and increase frequency of successful values.

The duplicate data of the fraudsters may refer to the applicants with common value. There are two types of duplicate data, one is an exact duplicate which have all data similar to other data and another duplicate been the near duplicate (i.e.) approximately similar data with some alteration in the spellings. This system of credit card fraud detection argues that each successful credit application fraud pattern is represented by a sudden and sharp spike in duplicate within a short time.

The applicant who is new to the credit card application, but has the facility of other credit service such as personal loan, mortgage loan, but committed a crime are listed towards the blacklist or the CIBIL list which consists of the fraudsters data. Hence, these kind of persons is checked for the CIBIL score that varies 300 to 900 that are generated by the SD algorithm are consider whether to provide further service or not.

Due to a rapid advancement in the electronic commerce technology, the use of credit cards has dramatically increased. As credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of fraud associated with it are also rising. Credit card fraud is the specific crime in the banking system. The credit card crime has been growing rapidly for the last few years. The process of making profit through credit card on the economy has been decreased about 8.2 crores annually in India. To avoid and predict the fraudulent activities on credit card application, in this research a method of detecting the fraud over credit card on behalf of the CIBIL scores. As data mining provides various ways to retrieve an appropriate

data from the storage, here in proposed system an efficient way of matching the data provided by the applicants of the credit card along with the CIBIL list to predict the fraudsters.

The existing process of fraud detection has the drawbacks of effectiveness and scalability for multiple variants of data, the Scheduling for Fast Throughput Multi-Pattern Matching algorithm used to match the large amount of attributes. In order to predict the fraudulent applicants with an appropriate time constraints. Together with the communal detection (CD) and spike detection (SD) algorithm that removes the redundant attributes and generates the credit score for the CIBIL list or black list. The CIBIL score varies about 300 to 900, it has been recorded in the credit history and by considering its range the credits are provided to the lender or customer, and they are added to the white list.

This approach of credit card application fraud detection uses another database of Whitelist which stores the innocent applicants of credit card. Whitelisting uses real social relationships on a set of attributes.

3.1.1 Main Challenges in Fraud Detection System

Generally, the unconstitutional activity taking place in various applications is called as the fraud and the method of recognizing the unconstitutional individual is known as the fraud detection. The detection system needs more powerful techniques to identify the different types of attacks. It adopts the new technologies for improving the security in the credit card application. The earlier approaches had few drawbacks in the fraud detection, such as less effective, less scalability, high response time, less efficiency, imbalance of data, false predictions, etc. The main challenges in fraud detection need to avoid the limitations of the above mentioned drawbacks and get fast throughput during the execution of the approach.

The main objective of this research is to detect the credit card fraud detection in the very initial stage of credit card application. This system uses the efficient approach for prediction of general frauds and crime activities. In order to achieve the identification of fraud, this work proposes two layers to complement the existing system are the CD and SD along **with two static scheduling algorithms: EBS and RBS,** then comes up with a hybrid method called Fast Response Time Scheduling (FRTS) to dynamically manage the scheduling in order to further reduce the average response time.

The processes involved in this research have the contribution of CIBIL score that are assigned to the applicants. Higher the score lowers the risk in providing the credit card. Initially the score is assigned to be 900 which may be varied by the transaction and other crime history of the customer. The CIBIL score is about 300 to 900, the lower the score results more risk of providing the credit card.

3.2 Methodology

The main contribution of this work is to achieve the challenges of the existing drawbacks of effectiveness, scalability, high response time, efficiency, imbalance of data, false predictions, etc. Being the first stage of the credit life cycle the fraudulent applicant is detected and hence the further transaction crimes will be prevented.

To schedule process for finding the fraud in the credit card application is obtained by the algorithm, namely Fast Throughput Time Scheduling (FTST) algorithm. The FTST consists of three methods together to improve the performance of the approach against the fraud, namely Multi-Pattern Matching, CD and SD. After receiving the input like application, each attribute of the application are stored in the data storage for future reference. At the same time, the application’s attributes are passed to the Multi-Pattern

Matching (MPM) method. MPM takes the attribute for comparison purpose of detecting the fraudulent applicant with the help of the White list and then with Black list database.

The performance of the whitelist and blacklist are described in two steps which is given below:

Step 1: The Whitelist database consists of the genuine individual's information. When a new application attributes are compared with the set of attributes in the whitelist to find the suspicion score by the equation (3.1),

$$\text{Suspicion score} = \begin{cases} 1 & \text{if } V_i \xrightarrow{\text{normal}} V_j, \text{where } V_i = V_j \\ 0 & \text{otherwise} \end{cases} \qquad (3.1)$$

Where the suspicion score is equal to 1, if a new application attribute is matched with whitelist window attributes, such as a normal application. The white list computes the suspicion score, which is higher for the particular application, then that application belongs to a genuine applicant. Where the suspicion score is equal to 0, if a new application attribute is not matched with the whitelist window attributes, then the application may be either new or fraudulent one.

Step 2: The Blacklist database consists of the known fraud information. When a new application attributes are compared with the set of attributes in the blacklist to find the suspicion score by the equation (3.2),

$$\text{Suspicion score} = \begin{cases} 1 & \text{if } V_i \xrightarrow{\text{fraud}} V_j, \text{where } V_i = V_j \\ 0 & \text{otherwise} \end{cases} \qquad (3.2)$$

When a new application attributes is matched with blacklist window attributes, then the suspicion score is equal to 1, the application is fraudulent. Otherwise the application is a new without any fraud.

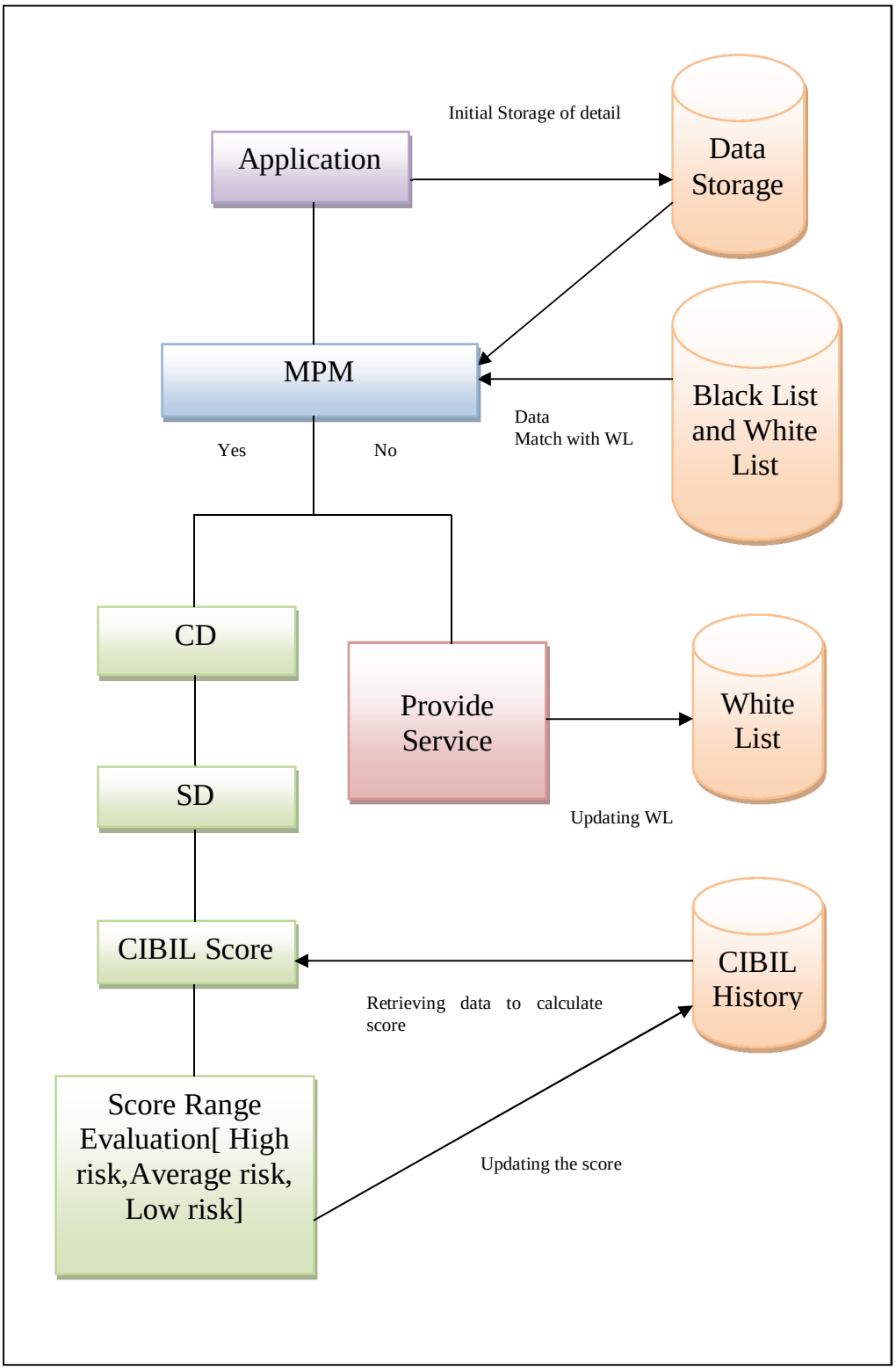

Figure 3.1 General Architecture of FTST

Based on the result obtained from each process, the further results are processed. The CIBIL history holds the data of the entire user and their transaction history that can be viewed and updated by the Administrator as shown in figure 3.1.

Multi- Pattern matching algorithm is the combination of both RBS and EBS algorithm is called as the hybrid method, which is the fast response of time scheduling used to extract the results for a specific time period. In the matching stage of both the algorithm the Multi - Pattern Matching has been applied to reduce the time complexity, by executing one or more query simultaneously the enormous amount of data can be matched and the result will be produced with less time constraints.

3.2.1 Multi-Pattern Matching Algorithm

This Multi-Pattern Matching is the initial process of the identification of fraud in the credit card application. This method consists of two basic methods of scheduling approaches, **EBS and RBS.**

- **EBS:** The EBS is the process, which is based on the event occur during the execution. Before the process will enter into the new event, it completes the test with the previous event for possible state transition. All the events are examined by the multiple runtime patterns in this approach.

- **RBS:** The RBS is the process, which is used to give a result either accepted or rejected. According to the result, it will take next runtime pattern to process. All the events are in the buffer in this approach. It will be revisited multiple times for the execution. Each event is examined by runtime pattern.

When current event is examined by multiple runtime patterns, the EBS gives the rapid throughput time for current event before appearing the

next event. Even this EBS is faster than the RBS. When the high rate streams occur, then the RBS is more effective. This approach uses both the scheduling methods for improving the performance to reduce the response time. The result of this approach has more effective and efficient, because it obtains the scheduling dynamically. Each attribute of the credit card application is treated as the event for this approach.

3.2.2 Algorithm Description

According to the above inference, the proposed approach designs the algorithm based on the FTMPM. The algorithm procedure is described step by step as follows.

Algorithm 3.1 Fast Throughput Multi-Pattern Matching

Input:

Attributes of the credit card application (A_1 , A_2 ……A_{21}), which is treated like an event.

1. Iterate **For** accessing each event **Then** update the datastorage.
2. Check the Event along with the Blacklist and whitelist database Find the suspicion score.
3. **If** the events are matched with whitelist,**Then**
   Suspicion score has more; accepted the application and update in the Whitelist database;
4. **Else If** the events are matched with Balcklist, **Then**
   Suspicion score is more; rejecting the application, then goto the next step.
5. Verify with CIBIL History;
6. Calculate the new CIBIL Score for new entries; and update in the CIBIL History database for future verification.

Output:

The application is accepted or reject according to the suspicion score.

3.3 Experimental Result and Discussions

The proposed Fast throughput Multi-pattern Matching (FTMPM) approach is validated through the performance of the fraud detection, which is evaluated for detecting the fraud function in a many number of datasets. The experimental study is carried out the following research directions:

- **Validation of the fraud detection approach:** The dataset creation is too hard, because the bank does not do. The dataset is generated through the simulator in the form of genuine and fraudulent application. At the same time, the fraud detection data of 50,000 credit card applications are chosen from https://sites.google.com/site/cliftonphua/communal–fraud-scoring data.zip for validating the fraud detection.

- **Performance comparison of the FTMPM with the existing tool:** The performance attributes, namely accuracy, efficiency, scalability and imbalanced class of attributes are used for comparison of the results between FTMPM and other existing approaches namely CD and SD.

3.3.1 Experimental Details

To experiment and verify the performance of this proposed approach, it is implemented in Java software and initially the raw data is taken which is converted into a MATLAB compatibility format. The experimental analysis has been carried out for a varying range of datasets. Developing a

Credit card application is very difficult due to some reasons, such as: 1) the companies do not share their databases, 2) the size of the database is growing day by day and 3) banking and companies are periodically changing the GUI on their applications. The synthetic dataset is created from https://sites.google.com/site/cliftonphua/communal-fraud-scoring-data.zip for this research. The bank will receive more than 10,000 applications per day. There are about 21 raw attributes such as name of the person (first,middle and last), DOB, Gender, addresses, telephone numbers, driver license numbers (or Aadhar card no), and other identifying attributes. Only 14 of the most essential identity attributes (1 to 14) are chosen. The attributes are in the form of alphanumeric, in which every numeric attributes are treated as string attributes, because which is implemented in the Java language. For preserving the privacy, few of these identifying attributes, are encrypted by homomorphic encryption.

Table 3.1 Sample Credit Card Application with Seven Attributes

S. no	First name	Last name	DOB	Gender	Address	Landline no
1	Harsha	Jeeva	26/8/1990	F	MG Road	8932121257
2	Harsha	Jeeva	6/8/1980	M	MG Road	8932121257
3	Prasanna	Elumalai	23/11/1980	M	Car street	9922121210
4	Balaji	Elumalai	27/9/1987	M	Car street	9922121210
5	Sunita	Patil	27/9/1987	F	Car street	9922121210
6	Sumita	Patil	27/9/1987	F	Car street	9922121210

Table 3.1 provides how social relationships are extracted from the applications, such as Harsha and Harsha both are husband and wife. They are

living in the same place and sharing the same phone number. Sunita and Sumita are twins. Prasanna and Balaji are brothers. They are the room mate of Sumita and Sunita. Balaji, Sunita and Sumita have common data of birth.

3.3.2 Experimental Results

Table 3.2 shows the performance evaluation of time taken by the various algorithms. It is most visible that the proposed FTMPM using CD and SD to verify and validate the attributes, and produce the pattern matching according to the time and rapid response. The performance values are depicted in Table 3.2.

Table 3.2 Performance Evaluation of Time Taken with Various Algorithms

Data set size (KB)	VFML (sec)	DT (sec)	NB (sec)	CD & SD (sec)	FTMPM (sec)
105737292	235	197	182	154	75
116716292	239	205	187	158	78
119397330	247	211	194	168	82
129397220	254	219	199	172	85
135357329	257	225	206	176	88
142139734	264	230	214	179	92
151356379	275	236	219	188	107
169397541	285	243	223	193	110

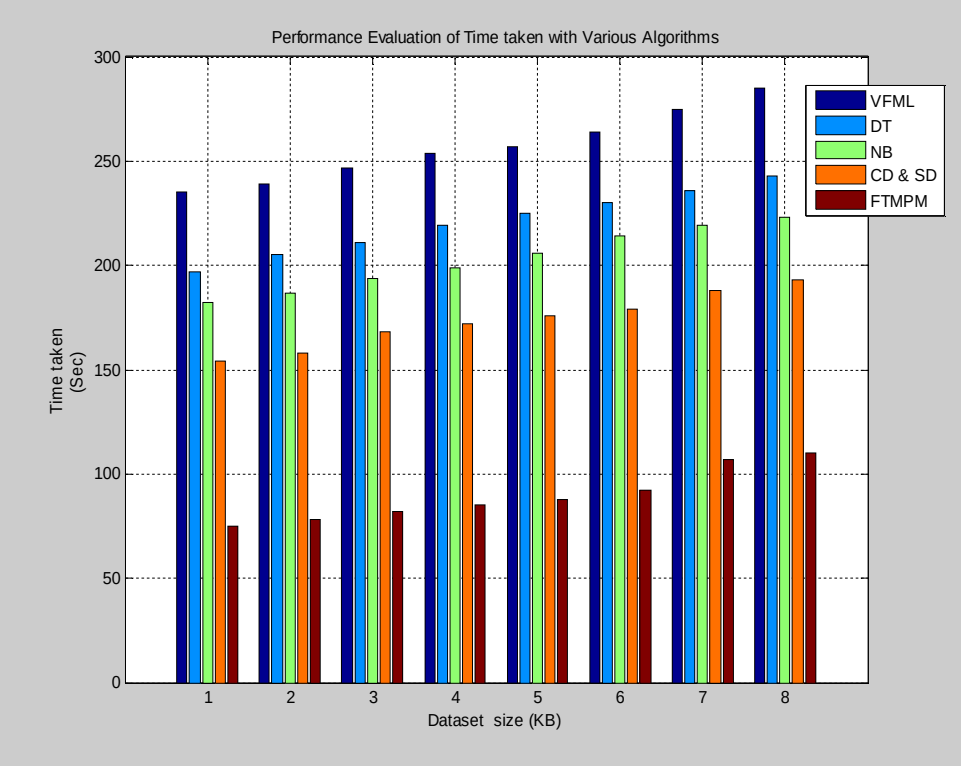

Figure 3.2 Performance Evaluation of Time Taken with the Various Algorithms

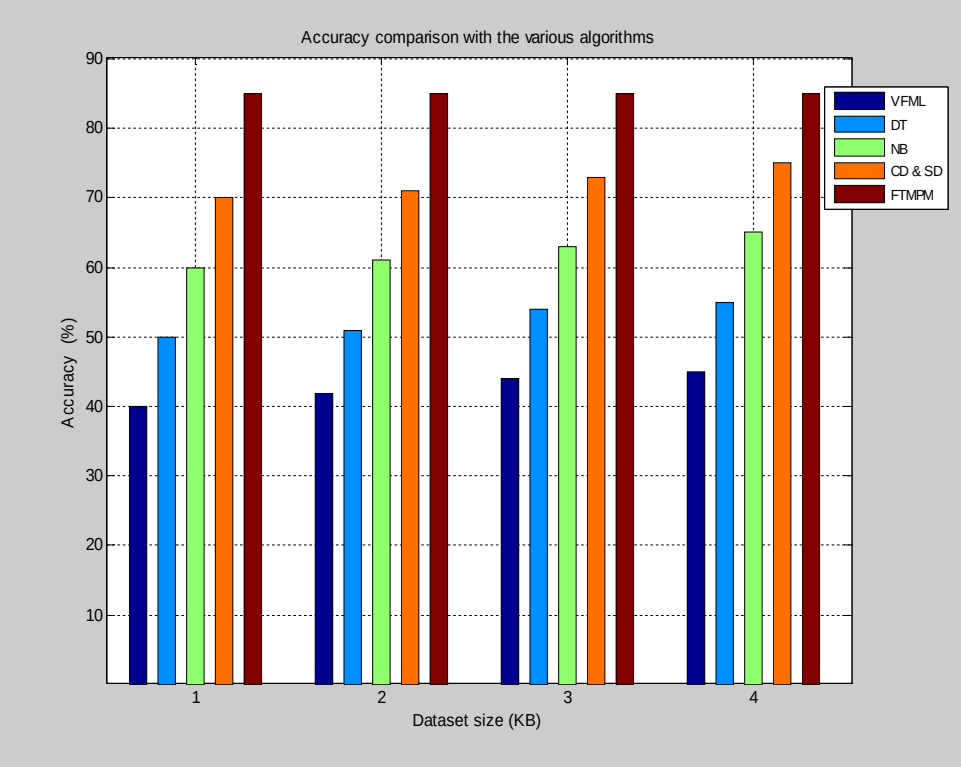

Figure 3.3 Accuracy Comparison with the Various Algorithms

The performance evaluation of rapid throughput for fraud detection in the credit card application is shown in the figure 3.2. It is evident that the proposed FTMPM approach surpasses other existing approaches in terms of the fraud detection in the credit card application.

Figure 3.3 presents the accuracy comparison with the various algorithms along with the proposed algorithm. It is more understandable that the proposed FTMPM approach surpasses other approaches.

3.4 Summary

The main target that focused on this chapter is to safeguard the credit application in the initial stage of the credit life cycle. The implementation of the Multi pattern matching algorithm in order to compare the attributes makes the identification process reliable with less time complexity. Here, the two main challenges are time constraints and accuracy have been achieved with balanced data load.

This work has been proposed with the efficiency in scalability by updating the evaluation of data. This FTMPM is not achieved the optimization level, because it worked along with the CD and SD approaches. To improve the performance in terms of the optimization, the next chapter introduces the new approach called as an **Improved Sheep Flock Heredity Algorithm** in which the attributes of every application [offline/online] are verified by using a newly developed procedure **MLMA [Multi-Level-Multi-Agent]** and it is verified all the attribute values are best one or not.

For optimizing the attributes the **ISFH-[Improved Sheep Flock Heredity]** has proposed to improve the efficiency of the credit card application fraud detection method by verifying and validating the optimized parameters, such as single and multiple attributes. The attributes of every

application [offline/online] are verified by using a newly developed procedure **MLMA**-[**Multi-Level-Multi-Agent**] and it is verified all the attribute values are best one or not. The experimental results of the proposed approach are compared with the existing approach results to compute the performance evaluation.

CHAPTER 4

IMPROVED SHEEP FLOCK HEREDITY ALGORITHM

4.1 Introduction

One of the Intrusion Detection System is credit card fraud detection in data mining. The existing approaches validate the fraud occurrence by computing a communal analysis suspicion score of the credit applications. The key challenge of this chapter is to improve the efficiency of the credit card fraud detection method by verifying and validating the optimized parameters, such as single and multiple attributes. The attributes of every application [offline/online] are verified by using a newly developed procedure **MLMA [Multi-Level-Multi-Agent]** and it is verified all the attribute values are best one or not. For optimizing the attributes the **ISFH-[Improved Sheep Flock Heredity]** algorithm is used and those attributes are validated according to the time and response with optimal value. The experimental results of the proposed approach are compared with the existing approach results to compute the performance evaluation.

4.2 Background Study

One of the dishonor criminal acts in online banking is credit card fraud. One of the expensive identity crime applications is the credit card application. From the ethical point of view the action taken against credit card fraud in banks and credit card companies. But, the software companies try to provide a solution on behalf of the banks and the customer. The early systems are having limitations in terms of score computation and applying rules. To overcome these limitations, the spike detection and communal detection are the two main processes applied to the existing chapter. The Communal

detection methodology finds out the social relationships among the data inputs. The Spike detection finds out the duplicates in the input data to find out the attacks.

In this chapter, the application refers to detecting the fraud in the credit card application. The identity verification in this application is synthetic and real identity verification. This Credit card application fraud detection needs to deploy in various generalized-distributed applications like insurance, telecommunication, online transactions and so on. In abroad countries, all these kinds of applications use a registered secret number as identity. It is well known that 75% of the World Bank's run their business on mainframes. So there are chances of hacking the systems on online. IBM introduced some of the following techniques, where card application fraud detection can follow for improving the detection accuracy such as:

- Identifying Vulnerabilities
- Transaction Detection
- Workloads are Evaluated
- Remediation conductance
- Process Appeals

Several approaches and techniques were applied in the earlier researches due to improve the detection accuracy and to provide high security. Sam Karl, et al. (2002) proposed Bayesian classification with neural network based approach for CFD. These approaches use the learning models and find out the fraud.

Kim and Kim (2002) proposed an analysis where it combines both fraud and legitimate transactions to improve the comparison detection. Few scholars in the early stage used clustering methods for grouping fraud and fraud-less data, grouping patterns which are visible and invisible. In general, the clustering approach clusters the parameters in regions.

Foster and Stine (2004) proposed a model which predicts the personal information from the bank data for the user, one who are using credit card alone. It verifies the non-linearity, missing values, standard errors, time taken for transaction and more transaction based destination addresses in the database. It is well known that the credit card based transaction is growing in the internet sales and purchases. In this case, the fraudsters make use of manipulating the credit card data in a charge-back method. Shilesh, et al. (2012) proposed the Utilized the hidden Markov model for analyzing the hidden entires of the credit card transaction in online payments. Ethics of banking is strongly provided for fraud detection. Anderson, R. (2007) proposed the various types of credit card fraud in financial industries with the appropriate remedies. According to the survey given by Linda Delamaire et al. (2009) and the Euromonitor International (2006) focused that 120 million numbers of credit cards were used in transaction in Germany alone. Due to increased number of credit card usage the fraud transaction is also getting increased. To detect the credit card fraud, the research scholars are proposing various approaches. In this chapter, an optimized attribute based application approval is introduced for credit card delivery to a customer.

4.3 Existing Approach

One of the security service company MaxMind, calculating a riskScore can determine the fraudulent. It uses the statistical analysis on IP-address, Devices, email-address, Geolocation verification, proxy detection. Bank ID, and compare with the minFraud network. It verifies the likelihood ratio of the available data and the input data of the card request. The range of the risk score is from 0.01 to 100. For example, if the request order has 20.00 riskScore is being a fraudulent.

Clifton Phua et al. (2012) proposed a multi layered detection approach for detecting credit card application fraud. Various existing

approaches are non-data mining approaches compares the business rules and scorecards with the known fraud limitations. But, Clifton Phua et al. (2012) utilized Communal Detection and Spike Detection based fraud detection. The Spike detection finds out the duplicate, fraud data in dynamic attributes and the Communal detection find out the duplicate, fraud data in static attributes of the credit card applications.

The combined CD and SD detect various attacks by comparing the input data with the persisted probe data. Since, the dynamic data may change according to the bank rules and increasing e-business, the detection rate and comparison time are poor. To improve the detection accuracy within a stipulated time a MLMA approach is proposed in this chapter.

4.4 Materials and Methods

A real time data set is chosen for experimenting our approach to improve the efficiency in terms of detecting most recent fraud people applied for a credit card. The data are taken from FSTC :http://www.fstc.org/, is a synthetic data having 50,000 credit card application information. In this data set most of the social attributes are very similar and twisted. The interval between the applications is in milliseconds. 75% of the attributes are treated as string attributes and other attributes are numeric. Some of the attributes are encrypted for privacy purpose.

Each attribute of every application is filtered and verified to detect frauds. The number of fields in each record is 30 and the size of the data is 140 bytes. In the overall data, 20% of the data are showing fraud entries. Our proposed approach can evaluate the entire data ad classify the score to predict the normal and fraud data. The following section describes about the optimization process.

4.5 Proposed Approach

In this research work, MLMA model is taken for analyzing the Credit card application fraud detection methods. There are two ways to apply credit card, such as online and offline. Both online and offline application data are fed into a software, which can directly convert the data into separate fields in a table. MLMA approach is proposed in this work in order to reduce the time complexity, improve the attribute verification accuracy and fraud detection accuracy. There are three agents IAgent, BAgent and CCAgent are integrated in our approach. All the input data [attributes of the credit card application form] is read and separated into three categories as Personal, social and Official using online/offline software.

IAgent verifies the Personal information, BAgent verifies the Social information and the CCAgent verifies the official information obtained from the application form. It is assumed that, the three agents can behave as CIBIL software by comparing the input data [attributes] with the available data, such as training data and real time data. The agents are also having permission to verify the fetched data formats and the values. The application is also having more credentials to be filled to get approved credit card for delivery.

Most of the fraud detection is obtained mostly from BAgent and CCAgent not from IAgent. IAgent always reads the encrypted data like DOB, Sex, Address, ID, and etc. But the BAgent compares the social information with the available real time information and finds the matching score. Similarly, the CCAgent compares official information with the available real time information and finds the matching score. Less cases, IAgent also give a less number of matching score, where the same person may have a credit card with other banks or from other credit card companies. The meta-information, history, IP address of the system from where the input comes, the interval between the applications, amount of credit are mainly focused and verified

from the DB and a score is assigned for the analysis. After verification and scores assigned by the agents, all the attributes and relevant data values are fed into ISFH algorithm, and finds the optimal score for credit card approval. Also, the MLMA model used in this chapter is shown in Figure 4.1.

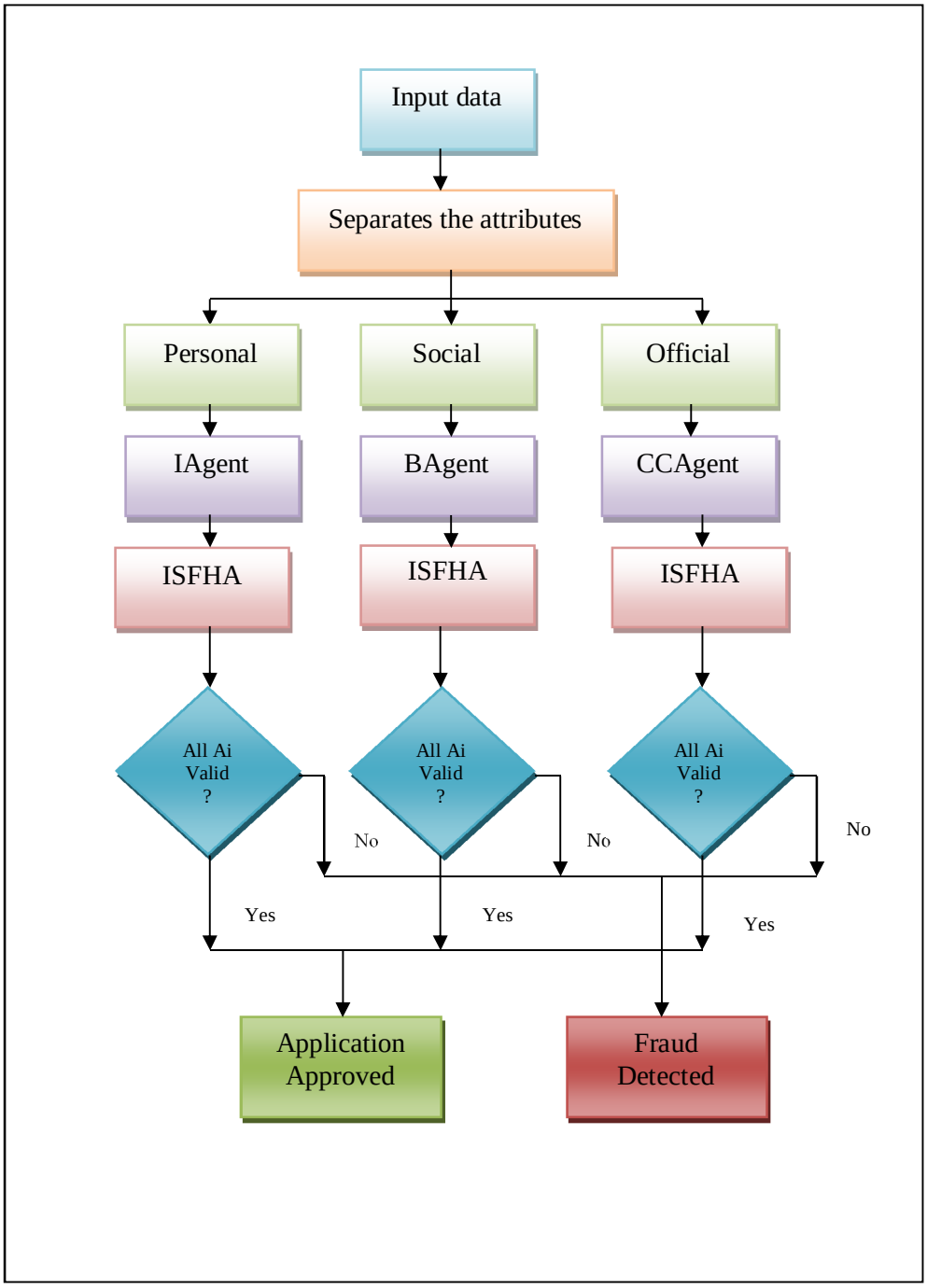

Figure 4.1 Multi Level Multi Agent Model for Analyzing the CFD Analysis

IAgent fetches the personal data from the user application, BAgent fetches the bank information from the user as well as from the relevant bank, and CCAgent fetches the credit card data from the user as well as from the credit card company for verification. All the fetched, verified data is optimized using ISFH algorithm and compare it with the training data. If the matched score is high, then the Application is treated as genuine, else it is treated as fraud. ISFHA has a set of states where each state is connected with an optimized probability distribution. The set of probabilities is called as application probabilities which are calculated on top of the training data. Since, the number of state and observation are internal as well as external it should be optimized in each credit card request and which is compared to the recorded early applications. Since the optimization technique is used in the detection rate of the false positive and false negative is reduced, because the maliciousness is decided according to the FPR and FNR values, where the ISFH algorithm helps to predict the FPR and FNR.

In this chapter, the CD can be obtained by comparing the present input values with the correct value to be derived as genuine values through the equation (4.1),

$$\begin{Bmatrix} \boldsymbol{score = 1} & \sum_{i=1}^{i=k} \boldsymbol{if\ A_i == CV} \\ \boldsymbol{score = 0} & \boldsymbol{if\ A_i \neq CV} \end{Bmatrix} \quad (4.1)$$

Where, score=1 denotes that the value of the attributes [A_i], matches with the correct value [CV] and score=0 denotes that the attributes are not meeting the perfectness.

The SD can be obtained by comparing the input values with the previous application values. The genuine values and the previous values are provided in the form of database to be compared to finding the matching score from the below equation (4.2),

$$\begin{cases} \boldsymbol{score = 1} & \boldsymbol{\sum_{i=1}^{i=k} if\ A_i \geq PA_i} \\ \boldsymbol{score = 0} & \boldsymbol{if\ A_i < PA_i} \end{cases} \quad (4.2)$$

Where score=1 denotes it scored a value, if the attribute value [A_i] is greater than the previous attribute value [PA_i]. In the second case, score=0 denotes, if the attribute value [A_i] is less than previous attribute value [PA_i].

Developing Credit card application fraud detection is very difficult due to some reasons, such as: 1) the companies do not share their databases, 2) the size of the database is growing day by day and 3) banking and companies are periodically changing the GUI on their applications. To address these issues, the models and the Metadata, and the software agents working intermediate are verified every time and grant permission only for the optimized entries. This chapter proposes a Credit card application Fraud Detection method with ISFH algorithm. ISFH algorithms are heuristic algorithms can provide better solutions within a stipulated time. Therefore, even though several correctly classification methods are available, ISFH algorithm reduces the time and improves the accuracy in finding the fraud by optimizing the parameters. The list of parameters used in this chapter is given in the following Table 4.1.

Table 4.1 Attributes Taken in Sample Data

Attributes	Description
A1	First Name
A2	Middle Name
A3	Last Name
A4	Date of Birth
A5	Gender
A6	Qualification
A7	ID Proof

A8	Email-ID
A9	Mobile
A10	Address Line-1
A11	Address Line-2
A12	City
A13	Pin Code
A14	Res. Number
A15	Occupation Type
A16	Bank of Credit Card
A17	Last Transaction Date
A18	No of Siblings
A19	Name of the Siblings
A20	Existing Bank Details
A21	Annual Income

A set of sample personal, social and official attributes are given in the Table 4.1. The present values of the attributes are determined and a comparison between the dataset and the critical values based parameters is obtained for increasing the number of true alerts. To find out a better solution, ISFH algorithm is applied continuously to compute the critical values, frequency of the particular individual's credit card application etc.

4.5.1 Improved Sheep Flock Heredity Algorithm

ISFH algorithm is an evolutionary algorithm and it aims to obtain the best solution within a short time. This process is also applied in data mining for feature selection methods. In this chapter, we are finding a solution to solve the classification problem using only improved sheep flock heredity algorithm. Improved sheep flock heredity algorithm is basically used for evaluating the natural evolution of sheep in a flock.

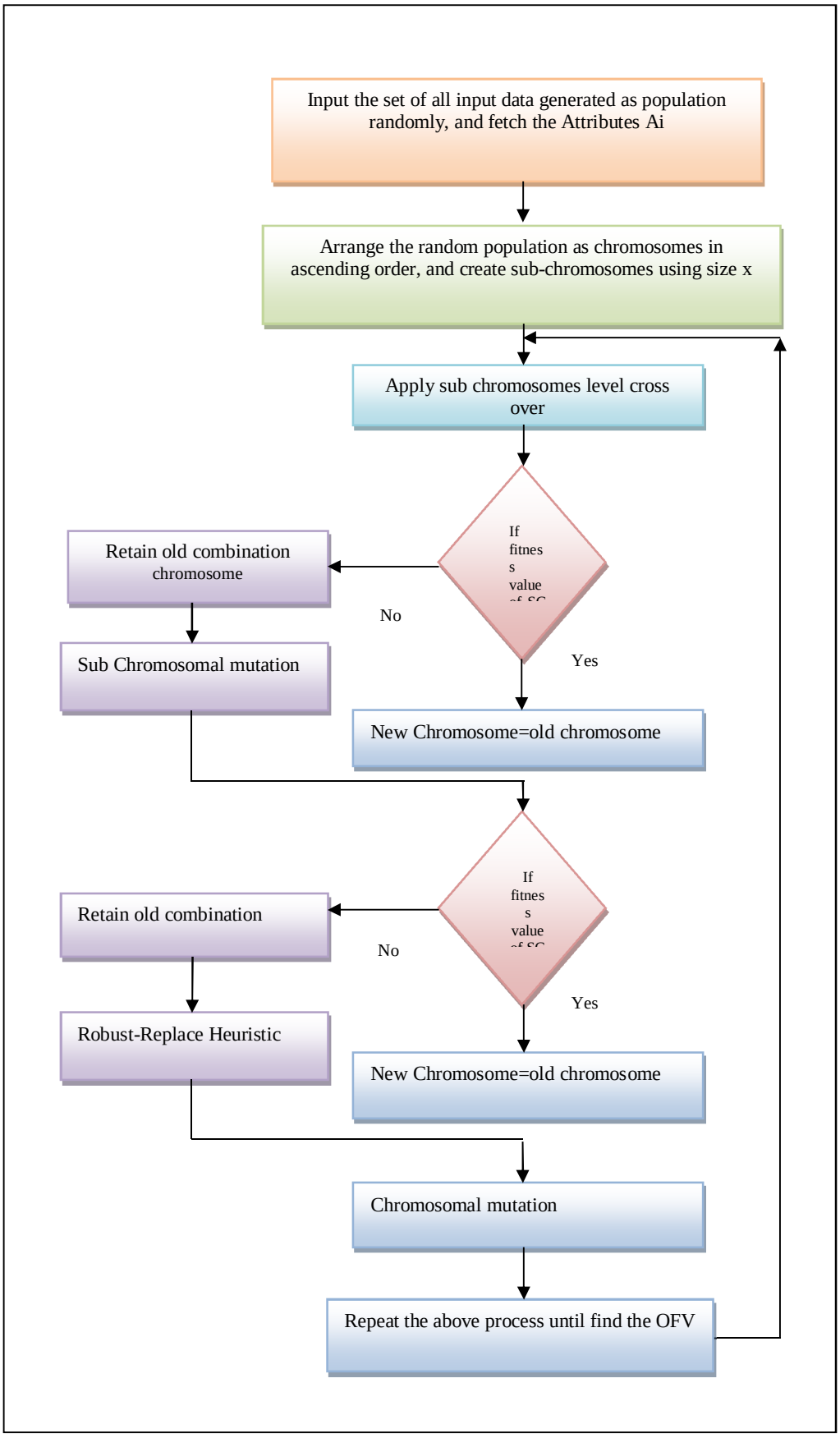

Figure 4.2 ISFH Algorithm Flow Diagram

ISFH algorithm simulates heredity of sheep flock in the lowland. Sheep in every flock are controlled by a shepherd. So that, the inheritance of the genetics can affect only the other sheep within the flock. Only, some special characteristics affect the sheep within the flock as well as in the nearest flocks. Those characteristics are called as fitness characteristics which can breed in the flock. Two sheep flock may have mixedcharacteristics with other flocks. The better fitness characteristics can breed the most.

The objective function considered in this chapter is given in the below equation (4.3),

$$\boldsymbol{A_i \geq \delta \; and \; A_i == Trainingdata - value, \forall \; i = 1, 2, 3, \ldots, N} \quad (4.3)$$

4.5.2 Algorithm Depiction

According to the above inference, an algorithm and flowchart are designed based on ISFH. The flowchart is shown in the figure 4.2. The algorithm procedure describes step by step as follows.

Algorithm 4.1 ISFH Algorithm

Input: Attributes of the credit card application (A_1 , A_2 ……A_{21}).

1. Initialize and generate random population P.
2. Make Sub-chromosome using length x.
3. Set the threshold values based on the training data for all elements in P called as OFV.
4. Apply cross over on a sub-chromosome.
5. Apply Inverse Mutation on sub-chromosomes.
6. Compute OFV for mutated sub-chromosomes and compare with the parent chromosome OFV and choose the chromosome according to the best OFV.

7. Apply Single Point Mutation on sub-chromosomes.
8. Compute OFV for mutated sub-chromosomes and compare with the parent chromosome OFV and choose the chromosome according to the best OFV.
9. Compute the Roust Replace Heuristic [1/OFV].
10. Repeat the step 4 for chromosome levels until obtaining a best OFV.

Output : Find the Optimal solution

An improved sheep flock heredity Algorithm is used to obtain the optimal solution for the credit card application fraud detection. This optimization is obtained through the step by step as follows.

Algorithm 4.2 Improved Sheep Flock Heredity Algorithm Pseudo code

Step1: Read all the attributes in the application where $N=\{A_1, A_{2,.....} A_N\}$, using IAgent. Read all the Trained data with N number of attributes where $N=\{BA_1, BA_{2,.....} BA_N\}$, using BAgent. Read all the Credit Card data with N number of attributes where $N=\{CCA_1, CCA_{2,.....} CCA_N\}$, using CCAgent.

Step2: The critical values are calculated by comparing the Ai values with the training data.

Step3: For all the generations the A_i values are verified as a Fraud or Normal using ISFH algorithm

Step4: Generate fraud transactions using ISFH algorithm and repeat step-1 to step-3. This process is used to analyze the feasibility of the Credit card Application Fraud Dection.

The implementation of the algorithm for detecting the fraud is given in the step by step as follows.

Algorithm 4.3 Implementation

1. Initialize Population P = $\{A_1\ to\ A_n\}$

   Check the conditions

   $$\begin{Bmatrix} score = i & if\ (A_i\ meets\ conditional\ value) \\ score = 0 & else \end{Bmatrix}$$

   According to the Credit Card application Frequency

   $$CCFreq = Total\ Number\ of\ Card\ applied(CU)/CCage$$

   If CCFreq is less than 0.5 then If $(\text{A1.valid} = \text{true})\&\&(\text{A2} = \text{Accept})\&\&(\text{A3} \geq \text{CC} - \text{amount})$

   $\&\&(\text{A4} < BankBalance)\&\&(\text{A5} \leq \text{dueDate})\&\&(\text{A6} \leq \text{CardExpiry Date})\&\&(\text{A9} \leq \text{curTime} + 10\text{s})\&\&(\text{A16} \leq \text{Exsiting Bank of the card})\&\&(\text{A7} = \text{rec.history})$

   then

   $$score = 1$$

   else

   $$score = 0$$

2. Make Sub-chromosomes as = $\{\{A_1\ to\ A_m\}, \{A_m\ to\ A_k\}, \ldots \{A_l\ to\ A_n\}\}$
3. Apply cross over and compute the critical score and compare it with the DB values and registered values
4. Apply inverse mutation and compute the critical score and compare it with the DB values and registered values.
5. Apply the single point mutation and compute the critical score and compare it with the DB values and registered values.
6. Repeat the above steps until all attributes satisfy all the conditions, then take that application as genuine application, and provide acceptance notification for further proceedings, else, reject the corresponding application as fraud.

4.6 Experimental Result and Discussions

To validate the proposed Improved Sheep Flock Heredity Algorithm (ISFH) approach, the performance of the fraud detection is evaluated for detecting the fraud function in a number of datasets using Multi-Level-Multi-Agent **(MLMA)**. Each agent performs the operation according to its responsibilities. The result is based on the combined results of all three agents.

The experimental study is carried out the following research directions:

- **Validation of the fraud detection approach:** The dataset creation is too hard, because the bank doesn't do. The dataset isgenerated through the simulator in the form of genuine and fraudulent application. At the same time, the fraud detection data of 50,000 credit card applications are chosen https://sites.google.com/site/cliftonphua/communal–fraud-scoring data.zip for validating the fraud detection.

- **Performance comparison of the ISFH-MLMA with the Existing tool:** The performance attributes, namely accuracy, efficiency, scalability and imbalanced class of attributes are used for comparison of the results between ISFH-MLMA and other existing approaches namely FTMPM.

4.6.1 Experimental Details

To experiment and verify the performance of this proposed approach, it is implemented in Java software and three systems were used for that. One system is assumed as server, where the registration data are stored, the other system is treated as the bank and the bank database is stored, in the

third system the credit card company data and card database is stored. Since, all the systems are connected in WiFi network, all the systems and software are distributed via MATLAB software, due to its interoperability.

- In online application, all the credentials are verified element by element. If the elements are valid, the next level of application can be permitted else, it gets rejected. For example, the PAN number entered in the application form is a not valid one, so it makes the customer to enter the valid PAN card number compulsory, as they should cancel themselves. This is the initial level of prevention applied to Credit card application Fraud Detection.

- In the offline application, all the credential values are fed into a database and each element values are fed into ISFH algorithm and verify all the attributes are best values or not. If the values are the best values, then that application form will be approved else it will get rejected. The Table 4.2 shows the sample attributes details of the application, which is in the terms of the intermediate computation where the highlighted entries are fraudulent.

Table 4.2 Experimented Intermediate Data

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	Score
22221	1	4232	4	10000	4	12000	12	5	0	0	0	0	1
22222	2	4342	45	4500	30	23000	23	6	9	2	0	9000	1
22223	3	5445	52	5300	5	12000	34	3	8	2	1	15000	1

22224	4	6453	64	5000	65	34000	45	7	7	14	0	85000	1
22225	5	6423	74	5000	23	9000	65	3	6	7	0	12000	1
22226	6	7634	34	6000	2	8000	54	3	3	0	0	12000	0
22227	7	9876	45	30000	5	120000	43	3	8	0	1	0	1
22228	8	1234	65	6500	7	10000	34	8	9	4	0	1900	1
22229	9	4325	76	54000	65	65000	32	54	8	2	0	16000	1
22230	10	5464	56	95000	65	100000	21	7	6	10	1	16000	1
22231	11	6536	34	10000	43	23000	11	4	7	2	0	11000	1
22232	12	7647	54	34000	5	45000	15	7	6	11	0	0	1
22233	13	9856	34	45000	43	56000	16	3	3	11	0	0	0
22234	14	3846	35	50000	7	60000	76	6	5	12	0	14000	1
22235	15	8473	67	4000	8	5000	87	3	4	3	0	9000	1
22236	16	3836	86	56000	12	129000	19	6	5	1	1	19000	1
22237	17	8476	34	55000	32	85000	87	3	3	12	0	0	0
22238	18	7363	67	45000	45	80000	86	6	5	3	0	19000	1
22239	19	5264	34	55000	67	65000	46	3	3	12	0	0	0
22240	20	7362	43	55000	89	89000	74	4	5	0	1	7000	1

From the above table 4.2, The entries 22226, 22233, 22237 and 22239 are having similar scores in attribute number 4, attribute number 10, attribute number 12 and similarity total score are rejected and it is shown in Table-2. From the overall dataset, a set of 1000000 samples is taken and analyzed using ISFH algorithm.

4.6.2 Experimental Results

Table 4.3 shows the performance evaluation of time taken by the various algorithms. It is the most visible that the proposed **ISFH algorithm** using **MLMA** to verify and validate the attributes and produce the optimal solution according to the time and response tooptimize value.

Figure 4.3 shows the performance evaluation of rapid response for fraud detection in the credit card application. It is evident that the proposed ISFH approach surpasses other existing approaches in terms of the fraud detection in the credit card application. The performance values are depicted in Table 4.3.

The performance evaluation of the time taken with the various algorithms is shown in the figure 4.3. It provides the results of the various algorithms. So ISFH algorithm takes less time to verify and validate the various sets of attributes.

Table 4.3 Performance Evaluation of Time Taken with Various Algorithms

Data set size (KB)	VFML (sec)	DT (sec)	NB (sec)	CD & SD (sec)	FTMPM (sec)	ISFH (sec)
105737292	235	197	182	154	75	31
116716292	239	205	187	158	78	35
119397330	247	211	194	168	82	28
129397220	254	219	199	172	85	37
135357329	257	225	206	176	88	42
142139734	264	230	214	179	92	45
151356379	275	236	219	188	107	48
169397541	285	243	223	193	110	54

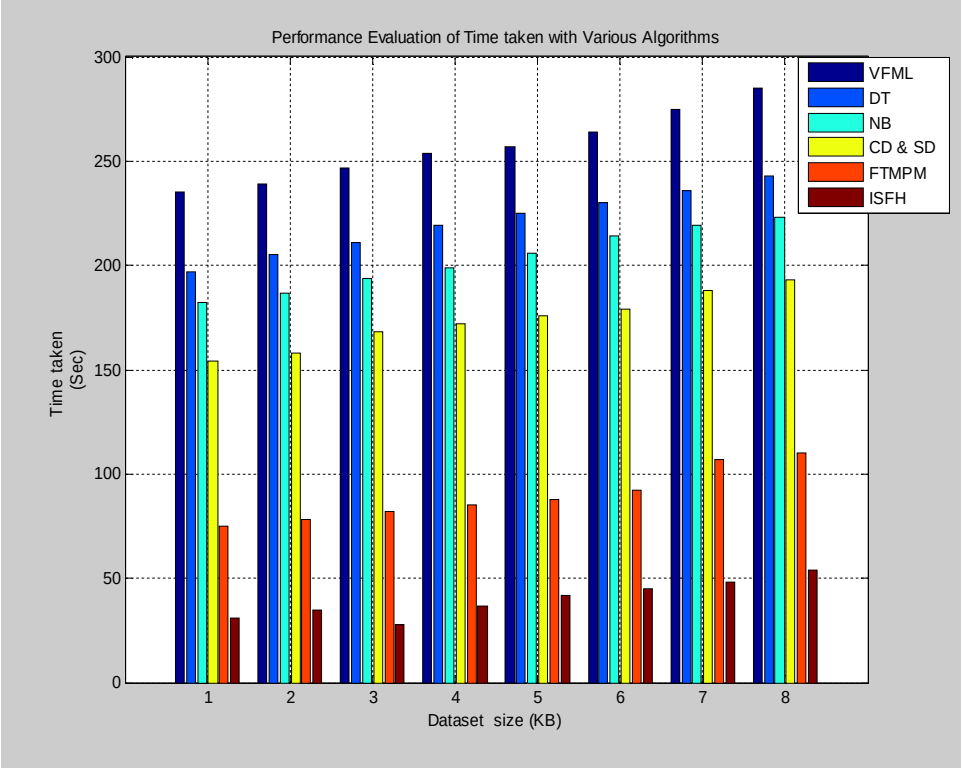

Figure 4.3 Performance Evaluation of Time Taken with the Various Algorithms

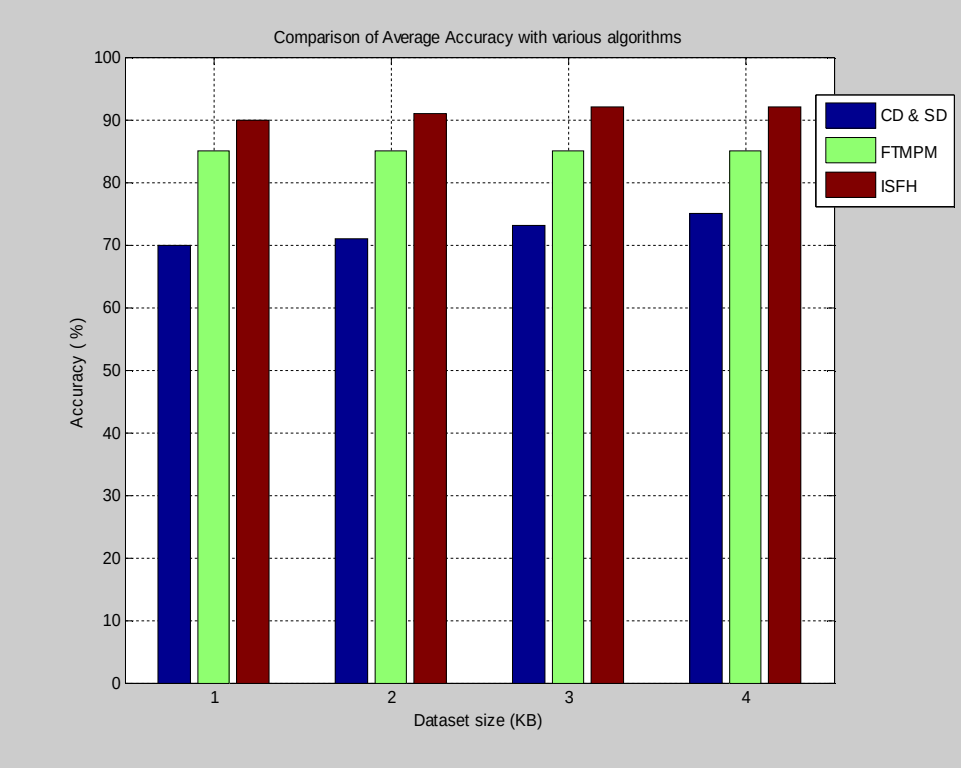

Figure 4.4 Accuracy Comparison with the Various Algorithms

Figure 4.4 presents the accuracy comparison with the various algorithms along with the proposed algorithm. It is more understandable that the proposed ISFH approach surpasses other approaches.

4.7 Summary

In this chapter, the MLMA system with ISFH algorithm based optimization is presented for finding the credit card application fraud detection and the results are examined in the applications of Credit Card. The ISFH algorithm is used to execute the credit card fraud detection by verifying all the attribute values are in a particular range or not. In this chapter, it is presented to detect the credit card fraud and the results are examined according to the principles of MLMA and ISFH algorithm. It can be seen that ISFH algorithm is applied to execute the Credit card Fraud Detection (CFD), to avoid fraud occurring in a financial institution to a customer or merchant.

ISFH algorithm is a heuristic algorithm, which is used in this chapter to obtain a better optimal solution. All the fraud data are detected by the ISFH algorithm and providing prevention for CFD. From the results and Figures, it is clear that the proposed approach is better than the other approaches can provide prevention for CFD also saves the time and un-wanted fund transfer. But it also has some limitation, such as it is used to test the limited number of problem sets. The collection of sheep in the one flock is unavoidably combined with the other flocks. This work takes huge number of dataset for verification and validation. But the ISFH has less memory capacity to store the large amount of dataset. To overcome this limitation, this research work intends to propose a new approach is called **Hybrid Swarm Optimization** Method in the next chapter.

CHAPTER 5

HYBRID ELEPHANT SWARM OPTIMIZATION METHOD

5.1 Introduction

The MasterCard's have found limitless use on account of the accommodation they offer. Credit applications are Internet or paper-based structures with formed requesting by potential customers for MasterCard's, home loan advances, and individual advances. The credit application fraud is a specific case of character fraud. The credit application fraud illustration is addressed by a sudden and sharp spike in duplicates within a brief range, as to the developmental benchmark level. Reduced plate finds real social associations with decrease the suspicion score, and modify impenetrable to make social associations. Detection of Spikes (SD) finds spikes in duplicates to extend the suspicion score and is tested safe for properties. The current framework perceives whether the applicant is extortion. In the present structure, the fraudster datum is secured in the database physically. In this proposed system, Communal Detection (CD) and Spike Detection (SD) can perceive more sorts of assaults, better record for changing legal lead and clear the overabundance credits and to store the false datum in blacklist using proposed ESS (Elephant Swarm Streamlining) estimation, because the most powerful characteristic of the elephant is its memory capacity, which survive and lead the family perfectly. The elephant makes the perfect solution for the following scenarios, namely recognition, identification and problem solving techniques with its memory capability.

Elephant Swarm Headway computerized examination using recuperation and determination to make the data more secure and to find the

false data. This proposed system makes the structure more powerful and update the security. The data mining includes various disclosure estimations. Data mining acknowledgment computation is used as a piece of the online MasterCard application. The estimations are used as a piece of this system is the spike recognizable proof, open revelation and ESO computations. These estimations are used to perceive the deception and hurl the data in the database at one of a kind data or blacklist database. This system overhauls the database physically. This system does not permit to fraudsters in control card application. The character misconduct is described as extensively as could be normal considering the present situation in this system. These can be harder to get (albeit vast volumes of some personal information are broadly accessible) yet simpler to effectively apply.

In actuality, character misconduct or fraud can be conferred with a blend of both engineered and genuine personality points of interest. Credit card applications are Internet or paper-based structures with composed solicitations by potential clients for charge cards, home loan advances and individual advances. Credit application misrepresentation is a particular instance of personality misconduct or fraud, including engineered character extortion and genuine wholesale fraud. As in this character misconduct or fraud, credit application extortion has come to a minimum amount of fraudsters who are exceptionally experienced, composed, and advanced.

There are two sorts of copies: careful (or indistinguishable) copies have the very same quality; close (or estimated) copies have some same values (or characters), some comparative qualities with somewhat modified spellings, or both. To put it plainly, the new systems depend on White-posting and Detecting spikes of comparable applications. White-posting uses genuine social connections on a settled arrangement of characteristics. This decreases false positives by bringing down some suspicion scores. Identifying spikes in copies, on a variable arrangement of qualities. This builds genuine positives

by modifying suspicion scores properly. Information mining, the extraction of concealed prescient data from huge databases, is a capable new innovation with incredible potential to offer organizations some assistance with focusing on the most vital data in the information distribution centers. Information mining devices anticipate future patterns and practices, permitting organizations to make proactive, learning driven choices. The manager confirms the given information the current datum to discover whether it is false or unique. In the event that the information is unique, it will be added to the database else it will be tossed into the boycott.

Elephant swarm improvement is currently making a huge commitment to the undertaking of extortion identification. Elephant swarm enhancement frameworks can gain from test examples of the Visa use to order new cases and this methodology additionally has the guarantee of having the capacity to adjust new examples of extortion as they arise. The ESO framework is the utilization of versatile and half breed learning frameworks. The ESO issues are already considered excessively progressive, disorderly, or complex to precisely show. This approach of credit card application fraud detection is very popular because of the vast field of the E-commerce and the use of credit card is growing very fast in online and offline shopping. There are some key challenges in this field, firstly the data limitation is the main issue. This limitation of data makes the difficulty to recognize the patterns of applications.

The main approach in order to detect the fraud application, is the analysis of the pattern of the previous usual transaction and unusual transaction. But in various cases the behavior of the transaction changes constantly which causes difficulties for fraud detection. The next issue is the benchmark data for fraud detection, which is due to the availability issues of the credit card application data. The credit card datasets are highly imbalanced

an example the credit card data contains very less fraudulent application and more genuine application.

These key issues bring up the requirement of an efficient approach to detect and prevent the fraud transaction in the credit cards. Recently, various researches have been done in this area by using the data mining techniques. The process of analyzing the patterns and relationships in the data is called the data mining process. In order to perform data mining some steps need to be followed as formulation of problem, the collection of data, thorough study of the data, preparation of the algorithm models, validation of the models and implementation of the models.

In the last decade, many methods have been proposed for the credit card fraud detection using data mining techniques. Supervised learning method is the most commonly used methods in this area. Tao Guo et al. (2008) proposed a neural network based approach for the credit card fraud detection. Based on the synthetic data sequences of the transactions a confidence -based neural network approach was proposed to achieve the optimal results for the credit card fraud detection. Receiver operating characteristics were utilized to examine the efficiency. In some of the cases, fraud is done by merchants by stealing cardholder's confidential information. In this case of fraud by merchant, two main aspects need to be analyzed first is to observe the suspicious activity of the merchant and the second is to detect the potential of the merchant to do the fraud.

Fuzzy expert system was proposed by HaratiNik, M.R., et al. (2012) to examine the first aspect and for the second aspect Fogg behavioral model was proposed to detect the potential. Another data mining techniques which are helpful to detect the fraud transaction are discussed in existing work which are Multi-Layer Perceptron (MLP), decision tree method and Chebyshev functional link artificial neural network (CFLANN). But these algorithms

suffer from the issues like time taken for processing, accuracy for large datasets and efficiency for various attribute, etc. To overcome these issues related to credit card fraud detection we propose a hybrid elephant swarm optimization algorithm (HESO) based on the mass interaction of the swarm. The support vector machine method is used to perform the classification of data. The performance of the proposed system is measured in terms of accuracy, precision, recall, sensitivity, false score, specificity, true positive rate and false positive rate.

5.1.1 Problem Statement

The online credit card application is web based structures. This framework identifies the misrepresentation candidate utilizing the information mining calculations. The current framework utilized two calculations they are spike identification and shared location calculation. These two consolidate together recognizes whether the candidate is extortion or unique. The proposed framework brushing with these calculations ESO uses to locate the false information and places it into the boycott. This framework utilizes for the match investigation from the current boycott database to make the framework productive and secure.

5.1.2 Objective of the Work

Data mining is worried with an investigation of expansive volumes of information to consequently find intriguing regularities or connections which thus prompts better comprehension of the hidden procedures. The information mining comprises of different calculations, a few calculations utilizes for the recognition of misrepresentation location in Visa. The online Visa application utilizes these calculations mutual and spike recognition uses to distinguish the numerous candidate and with the counterfeit insightful ESO calculation uses to make the false information operating at a profit list.

5.1.3 Existing System

The credit card application the framework distinguishes whether the candidate is extortion or unique. The current framework executes two information mining layers they are collective and spike identification. The Communal Detection is utilized to locate the suspicious information about the fake individuals. It additionally used to locate the public relationship that is close to mirror the family bond. It is white rundown arranged. The spike recognition, it is quality arranged. This does not distinguish the extortion but rather overhauls the framework consistently and properties frequently other than the common recognition.

5.1.4 Demerits of Existing System

The framework recognizes the whether the information is misrepresentation or unique. In the event that the framework is information is misrepresentation the procedure does not continue to the following level. The framework is property arranged that the information is upgraded in the collective recognition physically. The framework does not check from the boycott database. Through the spike location the framework overhauls the traits routinely. The framework is not secure and it recognizes the first information likewise as extortion. (For e.g. - twins applying the card are additionally identified as the false information).

5.2 Proposed Approach

This section presents the proposed approach for the credit card application fraud detection, the hybrid elephant swarm optimization algorithm to solve the optimization and the classification, kernel based support vector machine algorithm is used. HESOA is a method of evolutionary computation method which is derived from the PSO (Particle Swarm Optimization). The

proposed algorithm is inspired from the behavior of elephants. According to this approach elephants try to find out the best solution in the search space in their path. In this approach, each elephant considers the velocity, distance, best path and best solution. Let us consider E is the elephant which tries to find the best solution. In order to find the best solution first step is to achieve the velocity of particles based on the weighting factor which can be defined by the equation (5.1),

$$x_{itx_{y^2}er}{}^{iter+1} = \mathcal{W}V_{iter}^{t} \quad (5.1)$$

Where,

$\mathcal{W}$ is the weighting factor

V is velocity

$iter$ is the number of iterations

t is the representation of iteration instance

The next step is to achieve coefficients of acceleration, which can be calculated by the equation (5.2),

$$y_{iter}{}^{iter+1} = \mathcal{A}_1 \times R \times \left(E_{best_{iter}} - P_{iter}^{t}\right) \quad (5.2)$$

Where,

$\mathcal{A}_1$ is acceleration of elephant

R is random number between 0 to 1 range

$E_{best_{iter}}$ is the best path

P_{iter}^{t} presents the current position for particular iteration

The next step to achieve the best solution for the current position which can be calculated as defined in the equation (5.3),

$$\mathrm{z_{iter}}^{\mathrm{iter+1}} = \mathrm{A_2} \times \mathrm{R} \times \left(\mathrm{Esol_{best_{iter}}\text{-}P_{iter}^{t}}\right) \quad (5.3)$$

Where,

$Esol_{best_{iter}}$ is the best solution for the elephant swarm

By combining equation (5.1), (5.2) and (5.3),

$$Velocity_{iter}{}^{iter+1} = x_{iter}{}^{iter+1} + y_{iter}{}^{iter+1} + z_{iter}{}^{iter+1} \quad (5.4)$$

Finally, the positions of elephant particles can be calculated by the below equation (5.5),

$$P_{iter}^{iter+1} = P_{iter}^{t} + Velocity_{iter}{}^{iter+1} \quad (5.5)$$

The System Architecture is described as shown in figure 5.1, initially display the GUI for entering credit card application details, then the client details are accepted by the system. The client also submits a new application. Comparison of new applications is done with the existing ones. Then CD and SD algorithms are performed. After identifying whether the application is a genuine one the application is accepted.

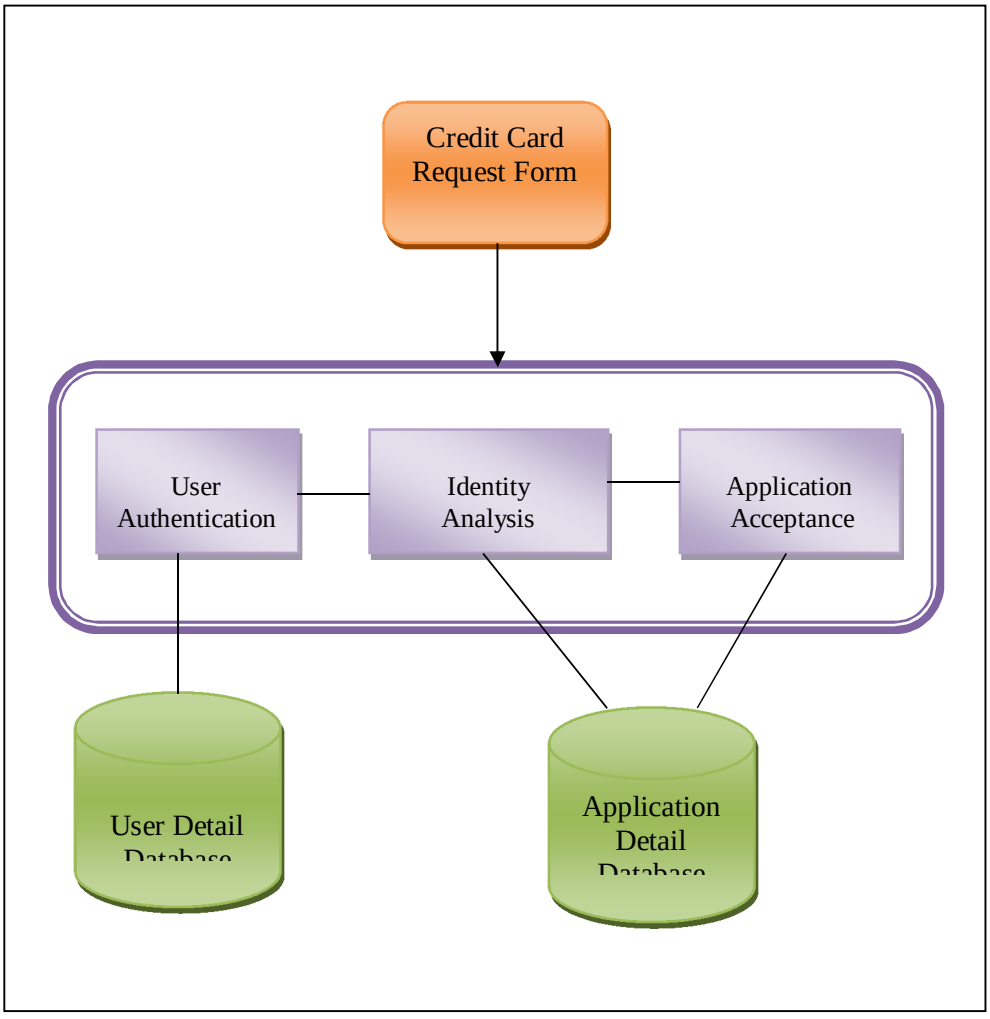

Figure 5.1 Hybrid Elephant Swarm Optimization

5.2.1 Proposed Model of Metric Indexing

To optimize the best position & the best path of elephant swarm, the research work proposed a new heuristic search algorithm based on the physical rules of the swarm. The algorithm is called mass search algorithm (MSA). This algorithm is inspired from the newton's law which says that all particles attract each other by the gravitational force. In this work, the particles (attributes) are denoted with an elephant. Each particle consists of four key properties: (i) Initial position of a particle, (ii) Inertia and mass, (iii) Gravitational mass of the particle, and (iv) Passive gravitational mass of the particle. The attraction force depends on the mass of the particle.

Let us consider a with elephant particle, the position solution can be defined by the equation (5.6),

$$P_{iter} = \left(P_{iter}^{1}, \ldots \ P_{iter}^{2}, \ldots . P_{iter}^{n}\right), i = 1,2, \ldots \ldots . . N \qquad (5.6)$$

The dimension of the problem is considered N and P_{iter}^{iter+1} is the position of the elephant in the i^{th} iteration.

Initially, all the positions of the elephants are assigned randomly. The attraction force for each iteration can be calculated by the equation (5.7),

$$Force_{ij}^{d}(t) = G(t)\frac{Mass_{Passive,i}(t) \times Mass_{active,j}(t)}{Euc_{ij}(t) + \zeta}(P_j^{\,d}(t) - P_j^{\,d}(t)) \quad (5.7)$$

Where,

$Mass_{Passive,i}(t)$ is the passive mass of i elephant at t instance of the epoch.

$Mass_{active,j}(t)$ is the active mass of j elephant at t instance of the epoch.

$G(t)$is the constant of gravity.

$Euc_{ij}(t)$is Euclidean distance.

ζconstant.

$G(t)$can be calculatedby the equation (5.8),

$$G(t) = G_{init} \times \exp(-\psi \times iter/maxiter) \quad (5.8)$$

Where,

G_{init} is the initial gravity constant.

Euclidean distance can be calculated by the equation (5.9),

$$\mathrm{Euc_{ij}(t)} = \left|\left|\mathrm{P_i(t), P_j(t)}\right|\right|_2 \quad (5.9)$$

According to the law of motion acceleration can be calculated by the equation (5.10),

$$\mathcal{A}_i^d(\mathrm{t}) = \frac{Force_i^d(t)}{Mass_{ii}(t)'} \quad (5.10)$$

The mass of the each elephant in each epoch is defined using fitness value. The particle which is having more mass compare to other particles is considered efficient, but processing is slow for that particle. For each iteration the mass is also updated which is defined by the equation (5.11),

$$mass_i(\mathrm{t}) = \frac{fitness_i(t) - weak(t)}{bestFit(t) - weak(t)} \quad (5.11)$$

Where,

$fitness_i(t)$is the fitness value and $weak(t)$is the weaker particle.

$bestFit(t)$is the best fitness value achieved.

$$bestFit(t) = \min_{j \epsilon 1..N} fitness(t) \quad (5.12)$$

$$weak(t) = \max_{j \epsilon 1..N} fitness(t) \quad (5.13)$$

By using equations (5.12) and (5.13) best and weak fitness values can be calculated for minimization problem by the equations (5.14) and (5.15),

$$bestFit(t) = \underset{j\epsilon 1..N}{max} fitness(t) \quad (5.14)$$

$$weak(t) = \underset{j\epsilon 1..N}{min} fitness(t) \quad (5.15)$$

For maximization problem the fitness values can be calculated by using equation (5.14) and (5.15).

Finally the normalized mass can be calculated by the below equation (5.16),

$$Mass_i(t) = \frac{mass_i(t)}{\sum_{j=1}^{N} m_j(t)} \quad (5.16)$$

In this approach, all particles initialized with random values and each particle contain the solution for the each swarm. The velocity and positions of particles are defined using equation 5.5 and 5.8. Masses can be calculated as equation (5.16).This section describes proposed hybrid swarm algorithm for optimization problem. This is the combination of elephant swarm optimization and heuristic search algorithm along with the SVM. In proposing algorithm, both algorithm's working is combined together to improve the performance to find the optimal solution. At the time of execution, both the algorithms work in parallel.

The overall algorithm is defined by the below equation (5.17),

$$\mathcal{V}_i(t+1) = \mathcal{W} \times \mathcal{V}_i(t) + \mathcal{A}_1{}' \times R \times Acc_i(t) + \mathcal{A}_2{}' \times R \times (E_{best_{iter}} - P_{iter}^{t}(t)) \quad (5.17)$$

According to the above inference, the flowchart is designed based on HESO. The flowchart is shown in the figure 5.2.

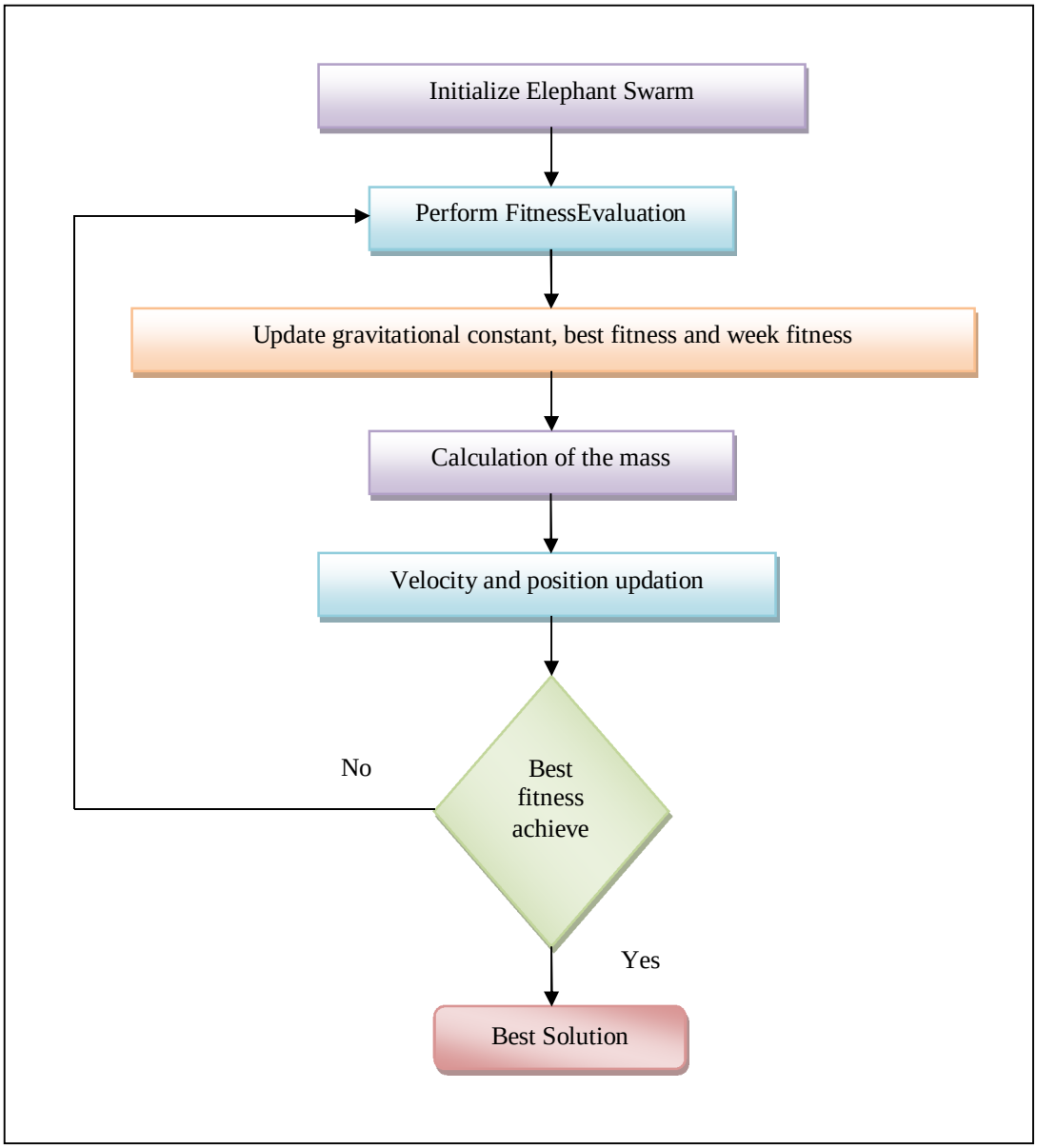

Figure 5.2 Flow Chart of the Proposed Hybrid Elephant Swarm Optimization Algorithm

The key points of proposed algorithm:

- It achieves the better fitness function because of its updating nature.
- The best solution in each iteration attracts other better solution.
- Search algorithm improves the speed of achieving the best solution faster.

These remarks show that the proposed algorithm improves the accuracy of optimizing the problem.

5.2.2 Support Vector Machine for Credit Card Application Fraud Detection

Support vector machine is the most used method for the pattern recognition and classification. In this approach it performs prediction and classification on the credit card application dataset and classifies into two classes; fraud and genuine application.

The basic function of binary classification using support vector machine is given by the equation (5.18),

$$\mathcal{F}(x) = sgn(v.w) + \alpha \quad (5.18)$$

Where,
v is the input vector;
w presentsthe weights of the input vector;
α is a constant;

The constant α is derived by maximizing the margin between two classes. This margin is calculated by utilization of hyper plane. These hyper planes are defined by the equation (5.19),

$$T: y = w.v + \alpha = 0 \quad (5.19)$$

Two classes are separated using the threshold T and two separation criteria are defined as T1 and T2.
The separation criteria are defined by the equations (5.20) and (5.21),

$$T1: y = w.v + \alpha = +1 \quad (5.20)$$

$$T2: y = w.v + \alpha = -1 \quad (5.21)$$

The above mentioned methods are utilized for linear problems of data, but as this research work have discussed in the literature survey that credit card data are imbalanced or non-linear, so to overcome this issue this work propose a kernel based approach for classification of credit card data using support vector machine.

Cristianini, N. and Shawe-Taylor, et.al. (2000) introduced kernel function of support vector machine which is defined by the below equation (5.22),

$$\langle v_i, v_j \rangle \rightarrow k(v_i, v_j) \quad (5.22)$$

By using this equation input vector can be mapped into a higher dimension in the support vector space. In this proposed approach has used RBF kernel method for vector projection in the space. This kernel is computed by the equation (5.23),

$$k(v_i, v_j) = \exp(-\mu \|v_i - v_j\|^2) \quad (5.23)$$

Where μ is the definition as Gaussian width.

In this proposed approach μ is used to tune the kernel parameter which helps to solve the imbalanced problems for classification. The performance of the SVM method depends onthe μand data fitting constant F. If data fitting constant value is small, then it induces insufficiency for training the data, and if it is more than it causes over fitting in the data.

This work is computing the overall performance by using the confusion matrix. The confusion matrix forms by using fewer parameter, namely true positive, true negative, false positive and false negative. But this

work takes two parameters to obtain the performance, such as given in the below equations (5.24) and (5.25),

$$True\ Positive\ Rate = \frac{\text{True Positive}}{True\ Positive + false\ Negative} \quad (5.24)$$

$$False\ Positive\ Rate = \frac{\text{False Positive}}{True\ Negative + False\ Positive} \quad (5.25)$$

The overall architecture of SVM classification is shown in the below given figure 5.3.

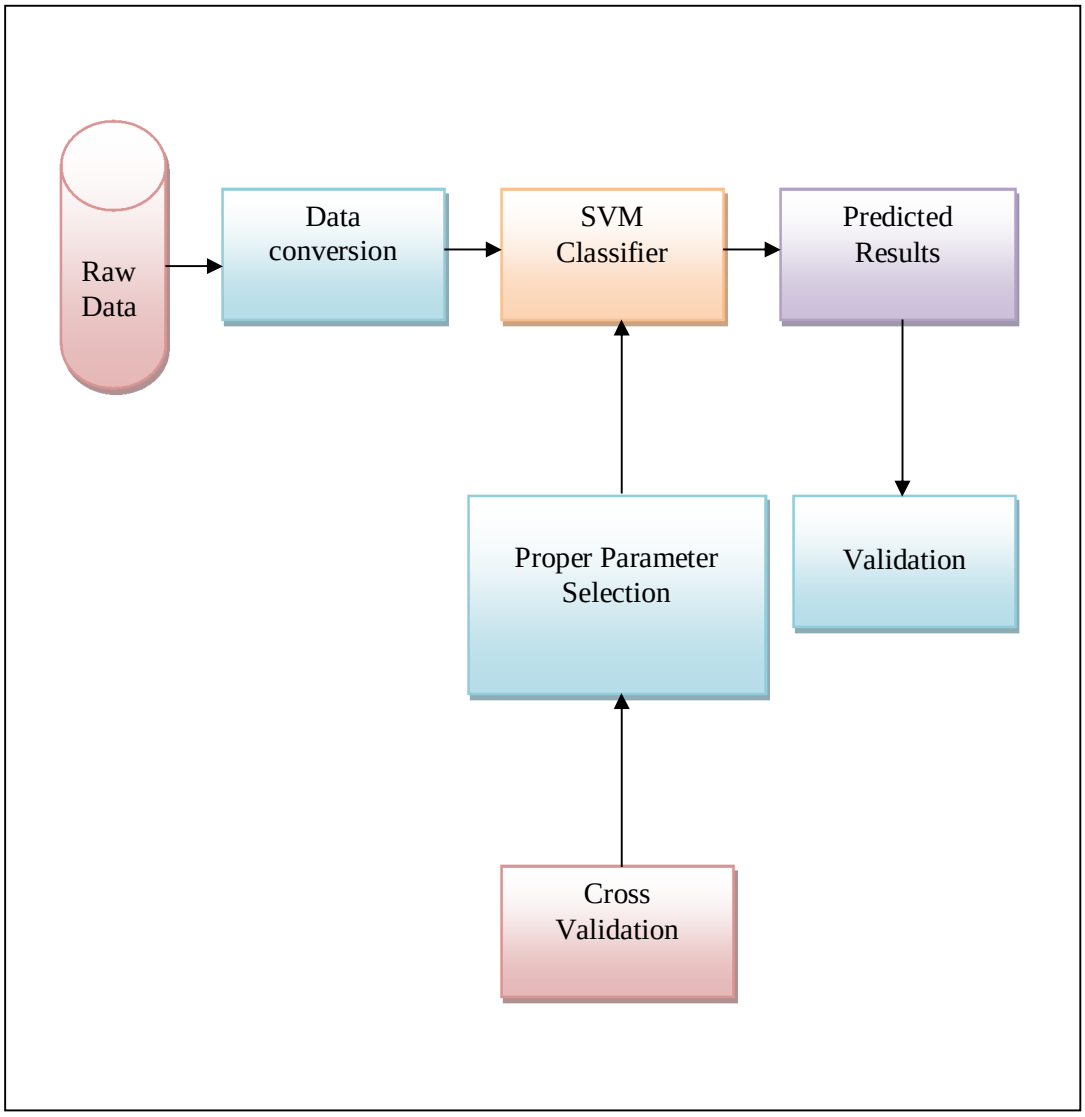

Figure 5. 3 Overall Architecture of SVM Classification for Credit Card Fraud Detection

5.2.3 Algorithm Depiction

According to the above inference, an algorithm is designed based on HESO. The algorithm procedure describes step by step as follows.

Algorithm 5.1 Hybrid Elephant Swarm Optimization

Input: Each attribute in the application treated as elephant particles

$P_{iter} = (P_{iter}{}^{1}, \ldots P_{iter}{}^{2}, \ldots . P_{iter}{}^{n}), i = 1,2, \ldots \ldots . . N$

1. Initialize elephant swarm
2. Evaluate the fitness
3. Update gravitational constant,

$$G(t) = G_{init} \times \exp(-\Psi \times iter/maxiter)$$

4. Update best fit, $bestFit(t) = \min_{j \in 1 \ldots N} fitness(t)$
5. Update weak fit, $weak(t) = \max_{j \in 1 \ldots N} fitness(t)$
6. Calculate the mass, $Mass_i(t) = \frac{mass_i(t)}{\sum_{j=1}^{N} m_j(t)}$
7. Update the velocity and position
8. If the best fit is achieved, then obtained best solution
9. Else
10. Goto step 2 to evaluate the fitness.

Output: Obtain the optimal solution.

5.3. Experimental Result and Discussions

To validate the proposed Hybrid Elephant Swarm Optimization (HESO) approach, the performance of the fraud detection is evaluated for detecting the fraud function in a number of datasets. The experimental study is carried out the following research directions:

- **Validation of the fraud detection approach:** The fraud detection data of 50,000 credit card applications are chosen from https://sites.google.com/site/cliftonphua/communal–fraud-scoring data.zip for validating the fraud detection.

- **Performance comparison of the HESO with the existing approaches:** The performance attributes, namely accuracy, efficiency, scalability and imbalanced class of attributes are used for comparison of the results between HESO and other existing approaches namely ISFHA and FTMPM.

5.3.1 Experimental Details

This section describes the proposed methodology for the classification of the data using kernel based SVM classification. Initially the raw data is taken which is converted into a MATLAB compatibility format. The next step is to perform the classification using SVM. SVM is trained using the input data; it gives the predicted output class as output of classification.

The training phase of SVM requires important features of the input data. In this chapter, the attributes are considered as key features. As this work has discussed that credit card application is an online application process under which users enter their personal information for registration.

According to this proposed model when the user feeds the data to the online application form, this data is matched to the database in the server end. Prior to this data matching process the attributes are learned by this proposed algorithm and then the dual optimization process is performed during this the attribute are divided based on their priorities.

5.3.2 Experimental Results

Table 5.1 shows the performance evaluation of time taken by the proposed algorithm along with the various existing algorithms, namely CD and SD, FTMPM, ISFH. It is most visible that the proposed **HESO algorithm** to verify and validate the attributes, and produce the optimization solution according to the time and responded with an optimal path for detecting the fraud in the credit card application.

Table 5.1 Performance Evaluation of Time Taken with Various Algorithms

Data set size (KB)	CD & SD (sec)	FSMPM (sec)	ISFH (sec)	HESO (sec)
105737292	154	75	31	10
116716292	158	78	35	15
119397330	168	82	28	19
129397220	172	85	37	20
135357329	176	88	42	22
142139734	179	92	45	25
151356379	188	107	48	30
169397541	193	110	54	32

This section presents the experimental study and result about the proposed system for the credit card fraud detection. MATLAB tool is used for the simulation. This work has taken the input like a credit card application dataset. According to proposed elephant swarm optimization method, best

fitness and best path is achieved. In the next step this data is used for training using the SVM classifier. This work has considered two classes fraud and genuine while training the data. The fraud class is represented by the 1 and the genuine class is represented by 0.

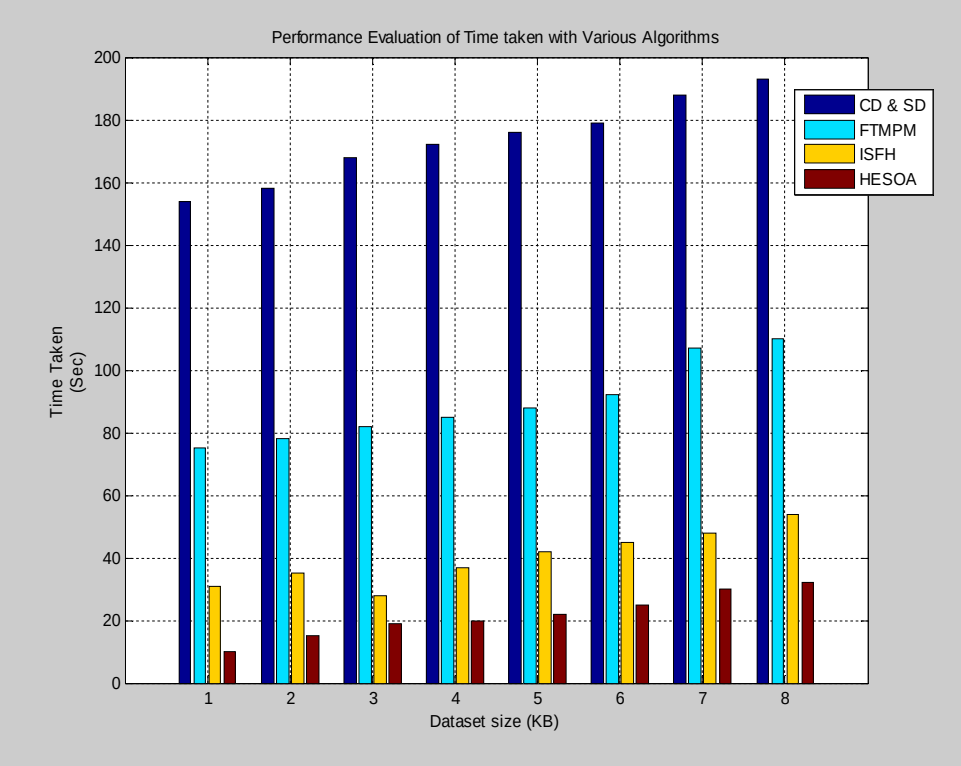

Figure 5.4 Performance Evaluation of Time Taken with the Various Algorithms

Figure 5.4 presents the performance evaluation of the time taken with the various algorithms. It is most noticeable that the proposed HESOA mechanism surpasses other mechanisms. The performance values are depicted in Table 5.1.

Figure 5.5 presents the accuracy comparison with the various algorithms along with the proposed algorithm. It is more understandable that the proposed HESO approach surpasses other approaches.

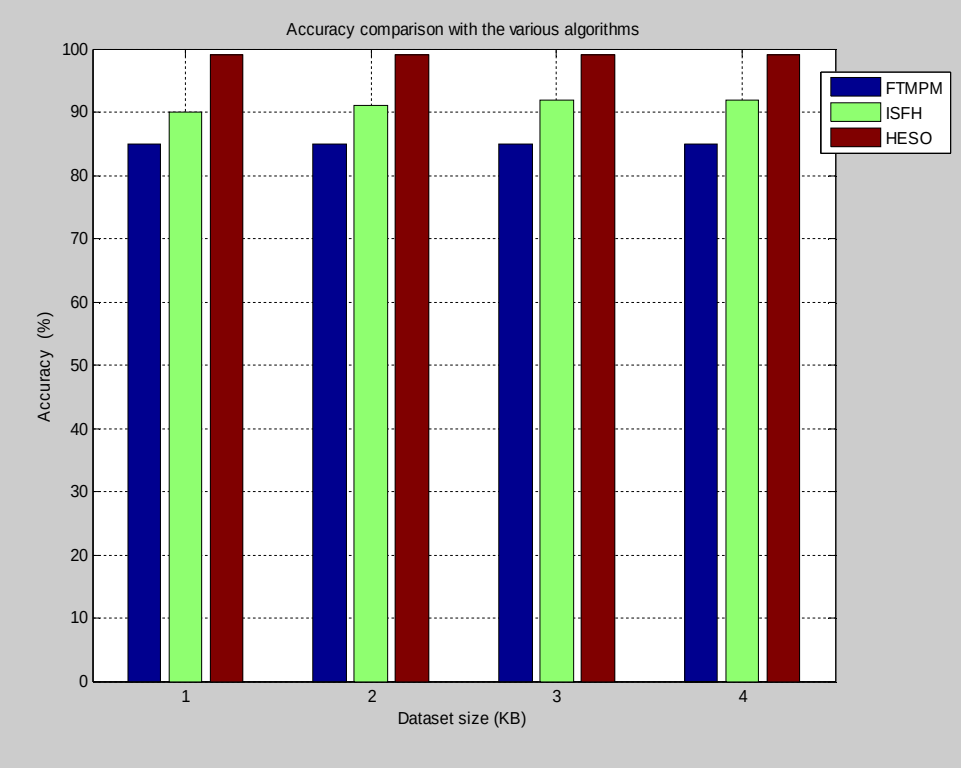

Figure 5.5 Accuracy Comparison with the Various Algorithms

5.4 Summary

The main aim of the work is to detect and prevent the credit card fraud application. In this method based on the attributes of credit card application validation is performed, the prevention of fraud transaction and for analyzing the system performance, this research work uses hybrid elephant swarm optimization is proposed based on the heuristic search algorithm for the credit card application fraud detection. The search algorithm is used to find the similarity among the neighboring attributes and elephant swarm optimization is used for finding the optimal path and best fitness. For classification of the transaction kernel based SVM method is used. Proposed system gives the best accuracy results and shows the data handling capacity for large databases. Results show the accuracy of 99.32% in terms of detection which is comparatively improved compared to other existing methods.

CHAPTER 6

CONCLUSION AND SCOPE FOR FURTHER STUDY

6.1 Conclusion

This study focused on the detection and analysis of fraud for easy maintenance. The core objective was the fraud detection using three approaches to detect the fraud which had improved the performance of the fraud detection based on the optimization.It also extended for fraud detection and prevention. Here, conclusions of this study are summarized.

i. The main target that focused on this FTMPM is to safeguard the credit application in the initial stage of the credit life cycle. The implementation of Multi pattern matching algorithm in order to compare the attributes makes the identification process reliable with less time complexity. Here the two main challenges are time constraints and accuracy have been achieved with balanced data load. This work has been proposed with the efficiency in scalability by updating the evaluation of data.This FTMPM has not achieved the optimization level, because it worked along with the CD and SD approaches. To improve the performance in terms of the optimization, the next chapter introduces the new approach called as an Improved Sheep Flock Heredity Algorithm.

ii. The MLMA system with ISFH algorithm based optimization is presented for finding the credit card application fraud detection and the results are examined in the applications of

Credit Card. The ISFH algorithm is used to execute the credit card fraud detection by verifying all the attribute values are in a particular range or not. In this chapter, it is presented to detect the credit card fraud and the results are examined according to the principles of MLMA and ISFH algorithm. It can be seen that ISFH algorithm is applied to execute the CFD, to avoid fraud occurring in a financial institution to a customer or merchant. ISFH algorithm is a heuristic algorithm, which is used in this chapter to obtain a better optimal solution. All the fraud data are detected by the ISFH algorithm and providing prevention for CFD. From the results and figures, it is clear that the proposed approach is better than the other approaches can provide prevention for CFD also saves the time and un-wanted fund transfer. But it also has some limitation, such as it is used to test the limited number of problem sets. The collection of sheep in the one flock is unavoidably combined with the other flocks. This research work takes a huge number of datasets for verification and validation of obtaining the optimal solution. But the ISFH has less memory capacity to store the large amount of datasets. To overcome this limitation, this research work intends to propose a new approach is called Hybrid Swarm Optimization Methodin the next chapter.

iii. The main aim of the work is to detect and prevent the credit card fraud application. In this method based on the attributes of credit card application validation is performed, the prevention of fraud transaction and for analysing the system performance, this research work uses hybrid elephant swarm optimization is proposed based on the heuristic search algorithm for the credit card application fraud detection. The

search algorithm is used to find the similarity among the neighboring attributes and elephant swarm optimization is used for finding the optimal path and best fitness. For classification of the transaction kernel based SVM method is used. Proposed system gives the best accuracy results and shows the data handling capacity for large databases. Results show the accuracy of 99.32% in terms of detection which is comparatively improved compared to other existing methods.

6.2 Scope for Further Study

Hybrid approach fulfills the present situations of safeguarding the Credit card application in the initial stage of the credit card life cycle. In the future, due to the increase of the data, it requires to adapt the larger database for storing the high volume of data. So it is preferred to enhance the same scenario in a cloud based system which intern to verify many data sets for BGSV (Background Screening and verification). At the same time, data can be divided as structured data (Relational), semi structured data (XML Data) and unstructured data (Word, pdf, text_media logs) where the physical verification and virtual matching of the documents are submitted. To implement in the cloud, it can get connected to third-party OLAP systems with cloud at securing level or Electronic Data Interchange (EDI) system between banks.

ANNEXURE

A. The data set chosen for the testing and estimation of the proposed credit card fraud detection approaches, are as follows

1. Oracle 10g (Database) runs on the many different operating systems and also relational database management system (RDBMS).

2. Matlab (Matrix Laboratory) is a multi-paradigm numerical computing environment and fourth-generation programming language.

3. MuPAD is a CAS. Originally developed. MuPAD offers: To manipulate formulas symbolically a CAS, in discretionary accuracy classic and verified numerical analysis, program packages for linear algebra, number theory, statistics, differential equations, and functional programming, a programming language that supports object oriented programming and functional programming, an interactive graphic system that supports animations and transparent areas in 3D via menus often used commands are accessible. Similar to the word processing systems that allows the formulation of the mathematical problems as well as graphic visualization and explanations in the formatted text MuPAD offers a notebook concept.

4. The Java code can also be embedded.It is possible to extend MuPAD with C++-routines to accelerate calculations.

5. Is a software program that allows computation over mathematical expressions in a way which is similar to the traditional manual computations of mathematicians and scientists as a computer algebra system (CAS). The development of the computer algebra systems in the second half of the 20th century is part of the

discipline of "computer algebra" or "symbolic computation", which has spurred work in algorithms over mathematical objects such as polynomials.

6. A user interface allowing to enter and display mathematical formula,a programming language and an interpreter (the result of a computation has commonly an unpredictable form and an unpredictable size; therefore user intervention is frequently needed), a simplifies, which is a rewrite system for simplifying mathematics formulas, a memory manager, including a garbage collector, needed by the huge size of the intermediate data, which may appear during a computation, an arbitrary-precision arithmetic, needed by the huge size of the integers that may occur,a large library of mathematical algorithms.

7. The main ones are Axiom, Macsyma, Magma, Maple, Mathematica and Sage. Thislarge amount of required computer capabilities explains the small number of general purpose computer algebra systems.

8. Is a type of user interface that allows users to interact with electronic devices through graphical icons and visual indicators such as secondary notation, instead of text-based user interfaces, typed command labels or text navigation a graphical user interface (GUI). GUI's were introduced in reaction to the perceived steep learning curve of command-line interfaces (CLIs),which require commands to be typed on a computer keyboard.

9. Libraries written in Perl, Java, ActiveX or .NET can be directly called from MATLAB, and many MATLAB libraries (for example XML or SQL support) are implemented as wrappers around Java or Active X libraries. Calling MATLAB from Java is more complicated, but can be done with a MATLAB toolbox which is

sold separately by Math Works, or using an undocumented mechanism called JMI (Java-to-MAT LAB Interface,which should not be confused with the unrelated Java Meta data Interface that is also called JMI).

10. Model-Based Design(MBD) is a mathematical and visual method of addressing problems associated with designing complex control,signal processing and communication systems. It is used in many motion controls, industrial equipment, aerospace and automotive applications.

11. Linux is the open source operating system.

12. MATLAB has value classes and reference classes, depending on whether the class has handle as a super-class (for reference classes) or not (for value classes). Mat Lab's support for object-oriented programming includes classes, inheritance, virtual dispatch, packages, pass-by-value semantics and pass-by-reference semantics. However, the syntax and calling conventions are significantly different from other languages.

13. Increasingly, the web is being looked upon as an environment for providing modeling and simulation applications and as such, is an emerging area of investigation within the simulation community. Web-based simulation (WBS) is the invocation of computer simulation services over the World Wide Web, specifically through a web browser.

14. A tool for modeling and visualization of physical phenomena, that automatically generates Java code from mathematical expressions Easy Java Simulations.

B. The comparative results of the fraud detection approaches are presented in the following Table, as discussed in subsection.

Table A.1 Comparative Result of Performance Evaluation of the Time Taken with the Various Algorithms, namely NB, CD and SD and FTMPM

Data set size (KB)	NB (sec)	CD & SD (sec)	FTMPM (Sec)
105737292	182	154	75
116716292	187	158	78
119397330	194	168	82
129397220	199	172	85
135357329	206	176	88
142139734	214	179	92
151356379	219	188	107
169397541	223	193	110

Table A.2 Comparative Result of Performance Evaluation of the Time Taken with the Various Algorithms, namely CD and SD, FTMPM and ISFH

Data set size (KB)	CD & SD (sec)	FTMPM (sec)	ISFH (sec)
105737292	154	75	31
116716292	158	78	35
119397330	168	82	28
129397220	172	85	37
135357329	176	88	42
142139734	179	92	45
151356379	188	107	48
169397541	193	110	54

Table A.3 Comparative Result of Performance Evaluation of the Time Taken with the Various Algorithms, namely FTMPM, ISFH and HESO

Data set size (KB)	FSMPM (sec)	ISFH (sec)	HESO (sec)
105737292	75	31	10
116716292	78	35	15
119397330	82	28	19
129397220	85	37	20
135357329	88	42	22
142139734	92	45	25
151356379	107	48	30
169397541	110	54	32

REFERENCES

1. Agrawal, J., Diao, Y., Gyllstrom, D. and Immerman, N. “Efficient pattern matching over event streams,” in ACM SIGMOD international conference on Management of data, pp. 147–160, 2008.

2. Akdere, M., Etintemel, U. C. and Tatbul, N. “Plan-based Complex Event Detection Across Distributed Sources,” in VLDB conference, pp. 66-77, 2008.

3. Aleskerov, E., Freisleben, B. and Rao, B. “CARDWATCH: a neural network based database mining system for credit card fraud detection”, in computational intelligence for financial engineering. In: Proceedings of the IEEE/IAFE, IEEE, Piscataway, NJ, pp. 220-226, 1997.

4. Anderson, R. “The Credit Scoring Toolkit: theory and practice for retail credit risk management and decision automation”, New York: Oxford University Press, 2007.

5. Avnur, R. and Hellerstein, J. M. “Eddies: Continuously adaptive query processing”, SIGMOD '00 Proceedings of the 2000 ACM SIGMOD international conference on Management of data, pp. 261-272 , 2000.

6. Apparao, G., Singh Arun, Rao, G.S., Lalitha Bhavani, B., Eswar, K. and Rajani, D. “Financial Statement Fraud Detection by Data Mining”, Int. J. of Advanced Networking and Applications, Vol. 01, pp. 159-163, 2009.

7. Ayad, A.M. and Naughton, J.F. “Static optimization of conjunctive queries with sliding windows over infinite streams,” in SIGMOD Conference. ACM, pp. 419–430, 2004.

8. Ayyub, B.M. “Elicitation of Expert Opinions for Uncertainty and Risks”, CRC Press, Boca Raton, 2001.

9. Babbitt, T., Morrell, C. and Szymanski, B. “Self-selecting reliable path routing in diverse wireless sensor network environments”, Proc. IEEE International Symposium on Computers and Communications, pp. 1-7, 2009.

10. Barbieri,R., deLima, N.B.K. “Some applications of the pso for optimization of acoustic filters”, Application Acoustics, Vol. 89, pp. 62-70, 2015.

11. Baxter, R., Christen, P. and Churches, T. “A Comparison of fast blocking methods for record linkage”, In: Proceedings of SIGKDD03 Workshop on Data Cleaning, Record Linkage, and Object Consolidation, pp.1-6, 2003.

12. Beheshti, Z. and Shamsuddin, S.M.H. “Areviewofpopulation-basedmeta-heuristic algorithms” , Int.J.Adv.SofComput.Appl., Vol. 5, pp. 1-35, 2013.

13. Bhattacharyya Siddhartha, Jha Sanjeev, Tharakunnel Kurian and Westland Christopher, J. “Data mining for credit card fraud: A comparative study”, Elsevier conference, pp. 602-613, 2010.

14. Bifet, A., Holmes, G., Kirkby, R. and Fahringer, B. “MOA: Massive Online Analysis”, Journal of Machine Learning Research,pp. 1601-1604, 2010.

15. Bilenko, M., Mooney, R., Cohen, W., Ravikumar, P. and Fienberg, S. “Adaptive name matching in information integration”, IEEE Intelligent Systems, Vol. 18, pp. 16–23, 2003.

16. Bolton, R. and Hand, D. “Unsupervised Profiling Methods for Fraud Detection”, Statistical Science, Vol. 17, pp. 235-255, 2001.

17. Boussaid, I., Lepagnot, J. and Siarry, P. “A survey on optimization meta heuristics”, Information Science, Vol. 237, pp. 82-117, 2013.

18. Brause, R., Langsdorf, T. And Hepp, M. "Neural data mining for credit card fraud detection," in Tools with Artificial Intelligence, 1999. Proceedings. 11th IEEE International Conference on, pp.103-106, 1999.

19. Brockett, P., Derrig, R., Golden, L., Levine, A. and Alpert, M. “Fraud Classification Using Principal Component Analysis of RIDITs”, The J. Risk and Insurance, Vol. 69, no. 3, pp. 341-371, 2002.

20. Campos, F. and Cavalcante, S. “An extended approach for Dempster–Shafer theory”, in: Proceedings of the IEEE International Conference on Information Reuse and Integration, pp. 338–344, 2003.

21. Carney, D., Cetintemel, U., Rasin, A., Zdonik, S., Cherniack, M. and Stonebraker, M. “Operator scheduling in a data stream manager”, In VLDB, pp. 838–849, 2003.

22. Caruana, R. and Niculescu-Mizil, A. “Data Mining in Metric Space: An Empirical Analysis of Supervised Learning Performance Criteria”, Proc. 10th ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining (KDD ‘04), 2004.

23. Caudil, S.B., Ayuso, M. and Guillen, M. “Fraud detection using a multinomial logit model with missing information”, The Journal of Risk and Insurance, Vol. 72, pp. 539-550, 2005.

24. Chan, P.K., Fan, W., Prodromidis, A.L. and Stolfo, S.J.“Distributed data mining in credit card fraud detection”, in: Proceedings of the IEEE Intelligent Systems, pp.67-74 , 1999.

25. Chandrasekaran, M., Asokan, P., Kumanan, S. and Balamurugan, T. “Sheep Flocks Heredity Model Algorithm for Solving Job Shop Scheduling Problems”, International Journal of Applied Management and Technology, Vol. 4, pp. 79-100, 2006.

26. Chang Jau-Shienand Chang Wen-His “A Cost- Effective Method for Early Fraud Detection in Online Auction”, 10th International IEEE Conference on ICT and knowledge Engineering, 2012.

27. Chapman, S. “Simmetrics–Open Source Similarity Measure Library Accessed from:http://sourceforge.net/projects/simmetrics/”, 2005.

28. Chen, R.C., Chiu, M.L., Huang, Y.L. and Chen, L.T. “Detecting credit card fraud by using questionnaire-responded transaction model based on support vector Machines”, in: Proceedings of the Fifth International Conference on Intelligent Data Engineering and Automated Learning, Vol. 3177, pp. 800–806, 2004.

29. Chen, R.C., Luo, S.T., Liang, X. and Lee, V.C.S. “Personalized approach based on SVM and ANN for detecting credit card fraud”, in: Proceedings of the IEEE International Conference on Neural Networks and Brain, pp.810–815, 2005.

30. Chen, T.M. and Venkataramanan, V. “Dempster–Shafer theory for intrusiondetection in ad hoc networks”, in: Proceedings of the IEEE Internet Computing, pp. 35–41, 2005.

31. Chiu, C. and Tsai, C. “A web services-based collaborative scheme for credit card fraud detection”, in: Proceedings of the IEEE International Conference on e-Technology, e-Commerce and e-Service, pp. 177–181, 2004.

32. Christen, P. and Goiser, K. “Quality and Complexity Measures for Data Linkage and Deduplication”, Quality Measures in Data Mining, F. Guillet and H. Hamilton, eds., Vol. 43, Springer, doi: 10.1007/978-3-540-44918-8, pp. 127-151, 2007.

33. Chun-Hua JU and Na Wang “Research on Credit Card Fraud Detection Model Based on Similar Coefficient Sum”, 2009 First International Workshop on Database Technology and Applications, DOI 10.1109/DBTA.2009.170, pp. 295- 298, 2009.

34. Cortes, C. and Pregibon, D. “Information mining platforms: an infrastructure for KDD rapid deployment”, In: Proceedings of SIGKDD99, pp. 327–331, 1999.

35. Cortes, C., Pregibon, D. and Volinsky, C. “Computational Methods for Dynamic Graphs”, J. Computational and Graphical Statistics, Vol. 12, pp. 950-970, 2003.

36. Cremer, F., Breejen, E. D. and Schutte, K. “Sensor data fusion for anti-personnel land mine detection, in: Proceedings of the International Conference on Data Fusion (EuroFusion98), pp. 55–60, 1998.

37. Cristianini, N. and Shawe-Taylor “Support Vector Machines and other Kernal-based learning Methods”, Cambridge University Press, 2000.

38. Delamaire Linda, Abdou Hussein and Pointon John “Credit card fraud and detection techniques: a review”, Linda Delamaire-Banks and Bank Systems, Vol. 4, 2009.

39. Demers, A. J., Gehrke, J., Hong, M., Riedewald, M. and White, W.M. “Towards expressive publish/subscribe systems”, in EDBT, pp.627–644, 2006.

40. Dhanapal, R. “An intelligent information retrieval agent”, Elsevier International Journal on Knowledge Based Systems 2008.

41. Dorronsoro, J.R., Ginel, F., Sanchez, C. and Cruz, C.S. “Neural fraud detection in credit card operations”, IEEE Transactions on Neural Networks 8, pp. 827– 834, 1997.

42. EbbersMike “5 Things to Know About Detecting Credit Card Fraud” – IBMRedBook, 2013.

43. Ester, M., Kriegel, H. P., Sander, J. and Xu, X. “A density-based algorithm for discovering clusters in large spatial databases with noise”, in: Proceedings ofthe Second International Conference on Knowledge Discovery and Data Mining (KDD), pp. 226–231, 1996.

44. Euro monitor International, Financial cards in Germany Available at: http ://www.euromonitor.com / Financial_ Cards _in_ Germany, 2006.

45. Experian Detect: “Application Fraud Prevention System”, Whitepaper, http://www.experian.com/products/pdf/experian_detect.pdf, 2008.

46. Fawcett,T. “An Introduction to ROC Analysis”, Pattern Recognition Letters, Vol. 27, pp. 861-874, 2006.

47. Filippov, V., Mukhanov, L. And Shchukin, B. "Credit card fraud detection system", 7th IEEE International Conference on Cybernetic Intelligent Systems, CIS, pp.1-6, 2008.

48. Foster, D. and Stine, R., “Variable Selection in Data Mining: Building a Predictive Model for Bankruptcy”, Journal of American Statistical Association- 99, pp. 303-313, 2004.

49. FSTC :http://www.fstc.org/.

50. Ganji Ratnam Venkata and Mannem Prasad Naga Siva “Credit card fraud detection using anti-k nearest neighbor Algorithm”, International Journal on Computer Science and Engineering (IJCSE), Vol. 4, pp.1035-1039, 2012.

51. Ghosh, S. and Reilly, D.L.“Credit Card Fraud Detection with a Neural-Network”, 27th Hawaii International l Conference on Information Systems, Vol. 3, pp. 621- 630, 2003.

52. Goldenberg, A., Shmueli, G., Caruana, R. and Fienberg, S. “Early Statistical Detection of Anthrax Outbreaks by Tracking Over-the-Counter Medication Sales”, Proc. Nat’l Academy of Sciences USA (PNAS ’02), Vol. 99, no. 8, pp. 5237-5240, 2002.

53. Gordon, G., Rebovich, D. and Gordon, K. “Identity Fraud Trends and Patterns: Building a Data-Based Foundation for Proactive Enforcement,” Center for Identity Management and Information Protection, Utica College, 2007.

54. Guo Tao and Li Gui-Yang "Neural data mining for credit card fraud detection", International Conference on Machine Learning and Cybernetics, Vol.7, pp.3630-3634, 2008.

55. Gyllstrom, D., Agrawal, J., Diao, Y. and Immerman, N. “On supporting kleene closure over event streams,” in ICDE, pp. 1391–1393, 2008.

56. Hand, D. “Classifier Technology and the Illusion of Progress”, Statistical Science, Vol. 21, pp. 1-15, 2006.

57. Harati Nik, M. R., Akrami, M., Khadivi, S. And Shajari, M. "FUZZGY: A hybrid model for credit card fraud detection", 2012 Sixth International Symposium on Telecommunications (IST), pp.1088-1093, 2012.

58. Head, B. “Biometrics Gets in the Picture,” Information Age, pp. 10-11, 2006.

59. Heo, J., Hong, J. and Cho, Y. “EARQ: Energy Aware Routing for Real-Time and Reliable Communication in Wireless Industrial Sensor Networks”, IEEE Transactions on Industrial Informatics, Vol.5, pp.3-11, 2009.

60. Hodge, V. and Austin, J. “A survey of outlier detection methodologies”, Journal of Artificial Intelligence Review 22 (2), pp. 85–126, 2004.

61. Https://WWW.maxmind.com/en/minfraud-services.

62. Hua Chun JU, Wang Na “Research on Credit Card Fraud Detection Model Based on Similar Coefficient Sum”, 2009 First International Workshop on Database Technology and Applications, IEEE Computer Society, DOI 10.1109/DBTA.2009.170, 2009.

63. Hutwagner, L., Thompson, W., Seeman, G. and Treadwell, T. “The Bioterrorism Preparedness and Response Early Aberration Reporting System (EARS)”, J. Urban Health, Vol. 80, pp. 89-96, 2006.

64. IDAnalytics “ID Score-Risk: Gain Greater Visibility into Individual Identity Risk,”Unpublished, 2008.

65. Jaberipour Majid and Khorram Esmaile “Two improved harmony search algorithms for solving engineering optimization problems”, Commun Nonlinear Sci. Number Simulate, Vol. 15, pp. 3316-3331, 2010.

66. Jackson, M., Baer, A., Painter, I. and Duchin, J. “A Simulation Study Comparing Aberration Detection Algorithms for Syndromic Surveillance”, BMC Medical Informatics and Decision Making, Vol. 7, no. 6, 2007.

67. Jian Liu, J. and Li, F. “An Improvement of AODV Protocol Based on Reliable Delivery in Mobile Ad hoc Networks”, Proc. IEEE Fifth International Conference on Information Assurance and Security, Vol. 1, pp. 507-510, 2009.

68. Johnson, B. and Maltz, D.A. “Dynamic Source Routing in Ad Hoc Wireless Networks”, Kluwer, 1996.

69. Jonas, J. “Non-Obvious Relationship Awareness (NORA)”, Proc. Identity Mashup, 2006.

70. Ju Chun Hua and Wang Na "Research on Credit Card Fraud Detection Model Based on Similar Coefficient Sum",2009 First International Workshop on Database Technology and Applications, pp.295-298, 2009.

71. Kalyani Rama, K. Devi Uma, D. “Fraud Detection of Credit Card Payment System by Genetic Algorithm”, International Journal of Scientific & Engineering Research, Vol. 3, pp. 1-6, 2012.

72. Kantarcioglu, M., Jiang, W. and Malin, B. “A Privacy-Preserving Framework for Integrating Person-Specific Databases”, Proc. UNESCO Chair in Data Privacy Int’l Conf. Privacy in Statistical Databases (PSD ’08), pp. 298-314, 2008.

73. Kavitha, M. And Suriakala “Fraud Detection in Current Scenario, Sophistications and Directions: A Comprehensive Survey” in International Journal of Computer Applications, Vol. 111 –No 5, pp. 35-40, 2015.

74. Kennedy, J. and Eberhart, R. “Particle swarm optimization”, in:Proceedings of IEEE International Conference on Neural Networks, pp.1942–1948, 1995.

75. Kim, M. J. and Kim, T. S. “A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection”, Proc. International Conference on Intelligent Data Engineering and Automated Learning, Lecture Notes in Computer Science, Springer Verlag, Vol. 2412, pp. 378-383, 2002.

76. Klaus-Robert Muller, Sebastian Mika, Gunnar Ratsch, Koji Tsuda and Bernhard Scholkopf “An Introduction to Kernel-Based Learning Algorithms”, IEEE Transactions on Neural Networks, Vol. 12, No. 2, 2001.

77. Kleinberg, J. “Temporal Dynamics of On-Line Information Streams”, Data Stream Management: Processing High-Speed Data Streams, M. Garofalakis, J. Gehrke, and R. Rastogi, eds., Springer, pp. 221-239, 2005.

78. Knight, R.“Fraudsters favour brandy and one-way tickets”, Financial Times, UK, 2007.

79. Kubica, J., Moore, A., Cohn, D. and Schneider, J. “Finding underlying connections: a fast graph-based method for link analysis and collaboration queries”, In: Proceedings of the ICML, pp. 392–399, 2003.

80. Kursun, O., Koufakou, A., Chen, B., Georgiopoulos, M., Reynolds, K. and Eaglin, R. “A Dictionary-Based Approach to Fast andAccurate Name Matching in Large Law Enforcement Databases”, Proc. IEEE Int’l Conf. Intelligence and Security Informatics (ISI ’06),pp. 72-82, 2006.

81. Lee Yuh-Jye, Yeh Yi-Ren and Wang Frank Yu-Chiang “Anomaly Detection Via Online Oversampling Principal Component Analysis”, IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 7, pp. 1460 – 1470, 2013.

82. Levitt, S. and Dubner, S. “Freakonomics: A Rogue Economist Explores the Hidden Side of Everything”, Penguin Books, Great Britain. Liberty Financial, 2004, Fears over Holiday Credit Binge”, 2005.

83. Li, F. Y, Xu, Y. X. and Li, G. X. “Optimization Design of the Wind Turbine Gearbox Based onGenetic Algorithm Metho”, Advances in Materials Manufacturing Science and Technology XIV, Vol. 697-698, pp.697-700, 2011.

84. Li, Y., Zhan, Z., Lin, S., Zhang, J. and Luo, X. “Competitive and cooperative particle swarm optimization with information sharing mechanism for global optimization problems”, Information Sciences, Vol.293, pp. 370–382, 2015.

85. Li, Y. and Zhang, X., “Securing credit card transactions with one-time payment scheme”, Journal of Electronic Commerce Research and Applications ,Vol. 4, pp. 413–426, 2005.

86. Liu, P. and Li, L. “A Game-Theoretic Approach for Attack Prediction”, Technical Report, PSU-S2-2002-01, Penn State University, 2002.

87. Liu, M., Li, M., Golovnya, D., Rundensteiner, E. and Claypool, K “Sequence pattern query processing over out-of-order event streams”, IEEE 25th International Conference on Data Engineering, pp. 784 - 795, 2009.

88. Macskassy, S. and Provost, F. “Suspicion scoring based on guilt-by-association, collective inference, and focused data access”, In: Proceedings of the International Conference on Intelligence Analysis, 2005.

89. Maes Sam, Tuyls Karl, Schoenwinkel Van Bram and Manderick Bernard “Credit Card Fraud Detection Using Bayesian and Neural Networks”, First International NAISO CongressonNeuro Fuzzy Technologies, Havana, Cuba, 2002.

90. Mahmood, M.A. and Seah Winston, K.G. “Event Reliability in Wireless SensorNetworks”, Proc.Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISS-NIP), Adelaide, Australia, pp. 1-6, 2011.

91. Minegishi, T. and Niimi, A. "Detection of fraud use of credit card by extended VFDT," 2011 World Congress on Internet Security (WorldCIS), pp.152-159, 2011.

92. Mishra, M. K. and Dash, R. "A Comparative Study of Chebyshev Functional Link Artificial Neural Network, Multi-layer Perceptron and Decision Tree for Credit Card Fraud Detection", 2014 International Conference on Information Technology (ICIT), pp.228-233, 2014.

93. Mohajerzadeh, A. and Yaghmaee, M. H. “An efficient energy aware routing protocol for real time traffic in wireless sensor networks”, Proc. IEEE International Conference on Ultra-Modern Telecommunications & Workshops, 2009 (ICUMT'09),pp. 1-9, 2009.

94. Molyneaux, D. “Two case study scenarios in banking: a commentary on The Hutton Prize for Professional Ethics, 2004 and 2005”, Business Ethics: A European Review, 16:4, pp. 372-386, 2007.

95. Moradi Parham and Gholampour Mozhgan “A hybrid particle swarm optimization for feature subset selection byintegrating a novel local search strategy”, Applied Soft Computing, Vol. 43, pp. 117-130, 2016.

96. Mukherjee Rajarshi, Chakraborty Shankar and Samanta Suman “Selection of wire electrical discharge machining process parameters using non-traditional optimization algorithms”, Applied Soft Computing, Vol. 12, pp. 2506–2516, 2012.

97. Mukhopadhyay Arunabha, Mukherjee Sayali, and Mahanti Ambuj “Artificial Immune System for detecting online credit card frauds”, CSI Communications, 2011.

98. Muller Klaus-Robert, Mika Sebastian, Ratsch Gunnar, Tsuda Koji, and Scholkopf Bernhard “An Introduction to Kernel-Based LearningAlgorithms”, IEEE Transactions on Neural Networks, Vol. 12, No. 2, 2001.

99. Nara Koichi, Takeyama Tomomi and Kim Hyunchul “A New Evolutionary Algorithm Based on SheepFlocks Heredity Model and Its Application to Scheduling Problem”, IEEE SMC '99 Conference Proceedings, Vol. 6, pp. 503-508, 1999.

100. Neville, J., Simsek, O., Jensen, D., Komoroske, J.,Palmer, K. and Goldberg, H. “Using Relational Knowledge Discovery to Prevent Securities Fraud”, Proc. 11th ACM SIGKDD Int’l Conf. Knowledge Discovery in Data Mining (KDD ’05), doi: 10.1145/1081870.1081922, 2005.

101. Ogwueleka Nonyelum Francisca “Data Mining Application in Credit Card Fraud Detection System”, Journal of Engineering Science and Technology, Vol. 6, No. 3, 2011.

102. Oscherwtiz, T. “Sythetic Identity Fraud: Unseen Identity Challenge,” Bank Security News, Vol.3, p.7, 2005.

103. Ozcelik, M. H., Duman, E. And Cevik, T. "Improving a credit card fraud detection system using genetic algorithm",2010 International Conference on Networking and Information Technology (ICNIT), pp.436-440, 2010.

104. Padmanabhan, S., Chandrasekaran, M., Asokan, P. and Srinivasa Raman,V. “Optimal Solution for GearDrive Design Using Population Based Algorithm”, Intl. Journal of Review of Mechanical Engineering, Vol. 5,pp. 802-806, 2013.

105. Pago-Report- “The development of E-commerce sectors”, ©PagoeTransaction Services GmbH, 2005.

106. Patel, K., Chern, L. J., Bleakley, C. J. and Vanderbauwhede, W. “MAW: A reliable lightweight multi-hop wireless sensor network routing protocol”, Proc. IEEE International Conference on Computational Science and Engineering, CSE'09. Vol. 2, pp. 487-493, 2009.

107. Perkins, E. and Royer, E., “Ad-hoc On-Demand Distance Vector Routing” in Proc. of the Second IEEE Workshop on Mobile Computing Systems and Applications, New Orleans, LA, USA, pp.90-100.1999.

108. Phua Clifton, Gayler Ross, Lee Vincent and Smith-Miles Kate “Resilient Identity Crime Detection”, IEEE Transactions On Knowledge And Data Engineering, Vol. 24, No. 3, 2012.

109. Phua Clifton, Gayler Ross, Lee Vincent and Smith-Miles Kate “On the communal analysis suspicion scoring for identity crimein streaming credit applications”, Elsevier, 2009.

110. Phua Clifton, Gayler Ross, Lee Vincent and Smith-Miles Kate “ Communal Detetcion of Implicit Personal Identity Streams”, sixth IEEE conference on Data Mining – Workshop (ICDMW’06) ,2006.

111. Phua Clifton, Gayler Ross, Lee Vincent and Smith-Miles Kate “A comprehensive survey of data mining-based fraud detection research”, Clayton School of Information Technology, Monash University, 2005.

112. Prodromidis, A. L. and Stolfo, S. J. “Agent-based Distributed Learning Applied to Fraud Detection”, Technical Report CUCS-014-99, 1999.

113. Rada-Vilela, J., Johnston, M. and Zhang, M. “Population statistics for particle swarm optimization: hybrid methods in noisy optimization problems, Swarm Evolut.”, Comput. 22, 2015.

114. Ramkumar, E. and Kavitha, P. “Online Credit Card Application and Identity Crime Detection”, International Journal of Engg. Research and Technology (IJERT) Vol. 2, 2013.

115. Ramya, G. and Chandrasekaran, M. “ Solving Job Shop Scheduling Problem Based on Employee Availability Constraint”, Applied Mechanics and Materials, Vol. 376, pp.197-206, 2013.

116. Rao, R. V. and Savsani, V. J. “Mechanical Design Optimization Using Advanced Optimization Techniques”, Springer,2012.

117. Rinky Patel, D. and Singh Kumar Dheeraj “Credit Card Fraud Detection & Prevention of Fraud Using Genetic Algorithm”, International Journal of Soft Computing and Engineering (IJSCE), Vol. 2, pp. 292-294, 2013.

118. Roberds, W. “The impact of fraud on new methods of retail payment”, Federal Reserve Bank of Atlanta Economic Review, First Quarter, pp. 42–52, 1998.

119. Romanosky, S., Sharp, R. and Acquisti, A. “Data Breaches and Identity Theft: When Is Mandatory Disclosure Optimal?”, Proc. Ninth Workshop Economics of Information Security (WEIS), 2010.

120. Saihood, A. A. and Kumar, R. “Enhanced Location Based Energy-Efficient Reliable Routing Protocol for Wireless Sensor Networks”, International Journal of Inventive Engineering and Sciences (IJIES), Vol. 1, 2013.

121. Sanchez Juan, J. A., Marin-Perez, R. and Ruiz, P. M. “Boss: Beacon-Less on Demand Strategy for Geographic Routing In wireless Sensor Networks”, Proc. IEEE International Conference onMobile Ad hoc and Sensor Systems. MASS, pp. 1-10, 2007.

122. Sahin, Y. and Duman, E. "Detecting credit card fraud by ANN and logistic regression", in Innovations in Intelligent Systems and Applications (INISTA), 2011 International Symposium on , pp.315-319, 2011.

123. Savsani, V., Rao, R. V. and Vakharia, D. P. “Optimal weight design of a gear train using particle swarm optimization and simulated annealing algorithms”, Journal: Mechanism and Machine Theory, Vol. 45, pp. 531-541, 2010.

124. Schneier, B. “Schneier on Security”, Wiley, 2008.

125. Schneier, B. “Beyond Fear: Thinking Sensibly about Security in an Uncertain World”, Copernicus, 2003.

126. Shafer,G. “A Mathematical Theory of Evidence”, Princeton University Press, Princeton, 1976.

127. Shailesh Dhok, S. “Credit Card Fraud Detection Using Hidden Markov Model”, IJSCE, Vol. 2, PP. 88-92, 2012.

128. Sentz, K. “Combination of Evidence in Dempster–Shafer Theory”, Sandia National Laboratories, US Department of Energy, 2007.

129. Seyedhossein Leila and Hashemi Reza Mahmoud “Mining Information from Credit Card Time Series for Timelier Fraud Detection”, 2010 5th International Symposium on Telecommunications (IST'2010), 2010.

130. Srivastava, A., Kundu, A., Sural, S. And Majumdar, A.K. "Credit Card Fraud Detection Using Hidden Markov Model", in Dependable and Secure Computing, IEEE Transactions on , Vol.5, no.1, pp.37-48, 2008.

131. Stolfo, S. J., Fan, D. W., Lee, W. and Prodromidis, A. L. “Credit card fraud detection using meta-learning: issues and initial results”, in: Proceedings of the Workshop on AI Methods in Fraud and Risk Management, pp. 83–90 , 1997.

132. Suvasini Panigrahi and Kundu Amlan “Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning” Int’l Journal on Information fusion, Elsevier, vol. 10, pp.354–363, 2009.

133. Syeda, M., Zhang, Y. Q. and Pan, Y. “Parallel granular neural networks for fast credit card fraud detection”, in: Proceedings of the IEEE International Conference on Fuzzy Systems, Vol. 1, pp. 572–577, 2002.

134. Tanaka Hideaki, Tsukao Shigeyuki, Yamashita Daiki, Niimura Takahide and Yokoyama Ryuichi “Multiple Criteria Assessment of Substation Conditions by Pair-Wise Comparison of Analytic Hierarchy Process”, IEEE Transactions On Power Delivery, Vol.25, No. 4, pp. 3017 – 3023,2010.

135. Tanveera, M., Shubhamb, K., Aldhaifallahc, M. and Ho, S. S. “An efficient regularized K-nearest neighbor based weighted twin support vector regression”, Knowledge-Based Systems, Elsevier, Vol. 94, pp. 70-87, 2016.

136. Tanweer, M. R. and Suresh, S. “Human cognition inspired particle swarm optimization algorithm”, in:IEEE9th International Conference on Intelligent Sensors, Sensor Networks and Information Processing, pp.1–6, 2014.

137. Tanweer, M. R., Auditya, R., Suresh, S., Sundararajan, N. and Srikanth, N. “Directionally Driven Self-Regulating Particle Swarm Optimization algorithm”, Swarm and Evolutionary Computation, Elsevier, Vol. 28, pp. 98-116, 2016.

138. Thomas Binu and Raju “A Novel Fuzzy Clustering Method for Outlier Detection in Data Mining”, International Journal of Recent Trends in Engineering, Vol.1, No.2, pp. 161-165, 2009.

139. Wang “The study on gear transmission multi objective optimum design based on SQP algorithm”, Proc. Fourth International Conference on Machine Vision, Vol. 8350, doi:10.1117/12.920249, 2012.

140. Wang, Y., Yang, H., Wang, X. and Zhang, R. “Distributed intrusion detection system based on data fusion method”, in: Proceedings of the Fifth World Congress on Intelligent Control and Automation, Vol. 5, pp. 4331–4334, 2004.

141. Wasserman, S., Faust, K. “Social Network Analysis: Methods and Applications”, Cambridge University Press, New York, 1994.

142. Witten, I. and Frank, E. “Data Mining: Practical Machine Learning Tools and Techniques”, Morgan Kaufmann, San Francisco, 2005.

143. Wong, W. “Data Mining for Early Disease Outbreak Detection,” PhD thesis, Carnegie Mellon Univ., 2004.

144. Xu, X., Tang, Y., Li, J., Hua, C. and Guan, X. “Dynamic multi-swarm particle swarm optimizer with cooperative learning strategy”, Applied Soft Computing, Elsevier, Vol.29, pp. 169-183, 2015.

145. Yan Ying and Zhang Jin “Scheduling for Fast Response Multi-pattern Matching over Streaming Events”, IEEE 26th International Conference on Data Engineering (ICDE 2010), pp. 89-100, 2010.

146. Yi, Z., Khing, H. Y., Seng, C. C. and Wei, Z. X. “Multi-ultrasonic sensor fusion for mobile robots”, in: Proceedings of the IEEE Intelligent Vehicles Symposium, pp.387–391, 2000.

147. Yongbin Zhang, Fucheng You and Hua Qun Liu "Behavior-Based Credit Card Fraud Detecting Model", Fifth International Joint Conference on INC, IMS, IDC and NCM '09, pp.855-858, 2009.

148. Yusup Norfadzlan, Zain Mohd Azlan and Siti Zaiton Mohd Hashim “Evolutionary techniques in optimizing machining parameters: Review and recent applications (2007–2011)”, An International Journal Expert Systems with Applications, Vol. 39, pp. 9909-9927, 2012.

149. Zhu Qingsheng, Feng Jiand Huang Jinlong “Natural neighbor: A self-adaptive neighborhood method without parameter K”, Pattern Recognition Letters, Elsevier conference, Vol. 80, pp. 30-36, 2016.

PUBLICATIONS

1. Mareeswari, V. and Gunasekaran, G. “Improving identity crime detection using Scheduling for Fast Response Multi-pattern Matching over Streaming Events”, International Journal of Advanced Research in Computer Science Engineering and Information Technology, Vol. 3, No. 2321-3337, pp. 375-385, 2014.

2. Mareeswari, V. and Gunasekaran, G. “Improved Sheep Flock Heredity Algorithm Based Prevention of Credit Card Fraud Detection for Online and Offline Transaction”, International Journal of Applied Engineering Research, Vol. 10, No. 0973-4562, pp. 14285-14300, 2015.

3. Mareeswari, V. and Gunasekaran, G. “ Improved Traditionalist Flock Inheritance Algorithm Based Prevention of Credit Card Fraud Detection for Online and Offline Transaction”, International Journal of Recent Advances in Science Engineering, Vol. 1, pp.14-22, 2015.

4. Mareeswari, V. and Gunasekaran, G. “Fraud Prevention on Credit Card Application Using Hybrid Swarm Optimization Method”, International Journal of Emerging Research in Management &Technology,Vol. 5, No. 2278-9359, pp. 103-108, 2016.