wjert, 2020, Vol. 6, Issue 4, 393-425. Review Article ISSN 2454-695X

World Journal of Engineering Research and Technology

WJERT

www.wjert.org

SJIF Impact Factor: 5.924

A SURVEY ON MULTIPLE CLASSIFIERS SYSTEM FOR ANOMALY DETECTION IN CREDIT CARD DATA WITH UNBALANCED AND OVERLAPPED CLASSES

***[1]Mukesh Kumar Mandal and [2]Dr. Avinash Sharma**

[1]M.Tech. Scholar, CSE MITS Bhopal.

[2]A.P. & Head. CSE MITS Bhopal.

Article Received on 26/06/2020 **Article Revised on 16/07/2020** **Article Accepted on 06/08/2020**

***Corresponding Author**
Mukesh Kumar Mandal
M.Tech. Scholar, CSE
MITS Bhopal.

ABSTRACT

Credit card plays a very important rule in today's economy. It becomes an unavoidable part of household, business and global activities. Although using credit cards provides enormous benefits when used carefully and responsibly, significant credit and financial damages may be caused by fraudulent activities. Many techniques have been proposed to confront the growth in credit card fraud. However, all of these techniques have the same goal of avoiding the credit card fraud; each one has its own drawbacks, advantages and characteristics. The widened uses of Internet credit cards in e-banking systems are currently prone to credit card fraud. Data imbalance also poses a significant difficulty in the method of fraud detection. The efficiency of the existing fraud detection systems is only in question because it detects fraudulent action after the suspect transaction has been completed. In this study, a Multiple Classifiers System (MCS) has been used on two data sets: (i) credit card frauds (CCF), and (ii) credit card default payments (CCDP). The MCS employs a sequential decision combination strategy to produce accurate anomaly detection. The empirical studies show that the MCS outperforms the existing research, particularly in detecting the anomalies that are minorities in these two credit card data sets.

KEYWORDS: Credit Card, e-banking, Credit Card Frauds (CCF), Credit Card Default Payments (CCDP), Multiple Classifiers System (MCS), Anomalies Detection.

1. INTRODUCTION

1.1 Overview

At the current state of the world, financial organizations expand the availability of financial facilities by employing of innovative services such as credit cards, Automated Teller Machines (ATM), internet and mobile banking services. Besides, along with the rapid advances of e-commerce, the use of credit card has become a convenience and necessary part of financial life. Credit card is a payment card supplied to customers as a system of payment. There are lots of advantages in using credit cards such as.

Ease of purchase

Credit cards can make life easier. They allow customers to purchase on credit in arbitrary time, location and amount, without carrying the cash. Provide a convenient payment method for purchases made on the internet, over the telephone, through ATMs, etc.

Keep customer credit history

Having a good credit history is often important in detecting loyal customers. This history is valuable not only for credit cards, but also for other financial services like loans, rental applications, or even some jobs. Lenders and issuers of credit mortgage companies, credit card companies, retail stores, and utility companies can review customer credit score and history to see how punctual and responsible customers are in paying back their debts.

Protection of Purchases

Credit cards may also offer customers, additional protection if the purchased merchandise becomes lost, damaged, or stolen. Both the buyer's credit card statement and company can confirm that the customer has bought if the original receipt is lost or stolen. In addition, some credit card companies provide insurance for large purchases.

In spite of all mentioned advantages, the problem of fraud is a serious issue in e-banking services that threaten credit card transactions especially. Fraud is an intentional deception with the purpose of obtaining financial gain or causing loss by implicit or explicit trick. Fraud is a public law violation in which the fraudster gains an unlawful advantage or causes unlawful damage. The estimation of amount of damage made by fraud activities indicates that fraud costs a very considerable sum of money. Credit card fraud is increasing significantly with the development of modern technology resulting in the loss of billions of dollars worldwide each year. Statistics from the Internet Crime Complaint Center show that there has

been a significant rising in reported fraud in last decade.

Fraud detection involves identifying scarce fraud activities among numerous legitimate transactions as quickly as possible. Fraud detection methods are developing rapidly in order to adapt with new incoming fraudulent strategies across the world. But, development of new fraud detection techniques becomes more difficult due to the severe limitation of the ideas exchange in fraud detection. On the other hand, fraud detection is essentially a rare event problem, which has been variously called outlier analysis, anomaly detection, exception mining, mining rare classes, mining imbalanced data etc. The number of fraudulent transactions is usually a very low fraction of the total transactions. Hence the task of detecting fraud transactions in an accurate and efficient manner is fairly difficult and challengeable. Therefore, development of efficient methods which can distinguish rare fraud activities from billions of legitimate transaction seems essential.

Although, credit card fraud detection has gained attention and extensive study especially in recent years and there are lots of surveys about this kind of fraud that neither classify all credit card fraud detection techniques with analysis of datasets and attributes. Therefore in this paper, we attempt to collect and integrate a complete set of researches of literature and analyze them from various aspects.

1.2 Credit card fraud

Illegal use of credit card or its information without the knowledge of the owner is referred to as credit card fraud. Different credit card fraud tricks belong mainly to two groups of application and behavioral fraud.[3] Application fraud takes place when, fraudsters apply new cards from bank or issuing companies using false or other's information. Multiple applications may be submitted by one user with one set of user details (called duplication fraud) or different user with identical details (called identity fraud). Behavioral fraud, on the other hand, has four principal types: stolen/lost card, mail theft, counterfeit card and "card holder not present" fraud. Stolen/lost card fraud occurs when fraudsters steal a credit card or get access to a lost card. Mail theft fraud occurs when the fraudster get a credit card in mail or personal information from bank before reaching to actual cardholder.[3] In both counterfeit and "card holder not present" frauds, credit card details are obtained without the knowledge of card holders. In the former, remote transactions can be conducted using card details through mail, phone, or the Internet. In the latter, counterfeit cards are made based on card information.

Based on statistical data stated in[1] in 2012, the high risk countries facing credit card fraud threat is illustrated in Fig.1. Ukraine has the most fraud rate with staggering 19%, which is closely followed by Indonesia at 18.3% fraud rate. After these two, Yugoslavia with the rate of 17.8% is the most risky country. The next highest fraud rate belongs to Malaysia (5.9%), Turkey (9%) and finally United States. Other countries that are prune to credit card fraud with the rate below than 1% are not demonstrated in figure 1.

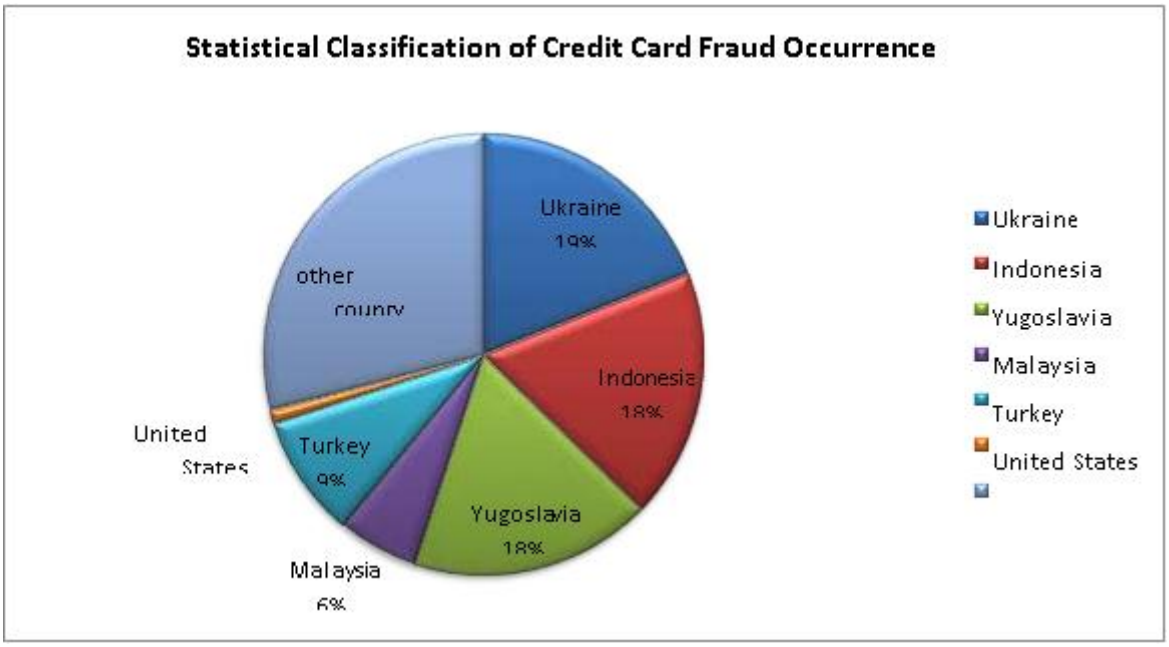

Fig. 1: High risk countries facing credit card fraud threat.

1.3 Difficulties of Credit Card Fraud Detection

Fraud detection systems are prune to several difficulties and challenges enumerated bellow. An effective fraud detection technique should have abilities to address these difficulties in order to achieve best performance.

Imbalanced data: The credit card fraud detection data has imbalanced nature. It means thatvery small percentages of all credit card transactions are fraudulent. This cause the detection of fraud transactions very difficult and imprecise.

Different misclassification importance: in fraud detection task, different misclassification errors have different importance. Misclassification of a normal transaction asfraud is not as harmful as detecting a fraud transaction as normal. Because in the first case the mistake in classification will be identified in further investigations.

Overlapping data: many transactions may be considered fraudulent, while actually they are normal (false positive) and reversely, a fraudulent transaction may also seem to be legitimate (false negative). Hence obtaining low rate of false positive and false negative is a key challenge of fraud detection systems.[4,5,6]

Lack of adaptability: classification algorithms are usually faced with the problem of detecting new types of normal or fraudulent patterns. The supervised and unsupervised fraud detection systems are inefficient in detecting new patterns of normal and fraud behaviors, respectively.

Fraud detection cost: The system should take into account both the cost of fraudulent behavior that is detected and the cost of preventing it. For example, no revenue is obtained by stopping a fraudulent transaction of a few dollars.[5,7]

Lack of standard metrics: there is no standard evaluation criterion for assessing and comparing the results of fraud detection systems.

2. LITERATURE REVIEW

2.1 Credit Card Fraud Detection Techniques

The credit card fraud detection techniques are classified in two general categories:

(1) Fraud analysis (misuse detection) and

(2) User behavior analysis (anomaly detection).

The first group of techniques deals with supervised classification task in transaction level. In these methods, transactions are labeled as fraudulent or normal based on previous historical data. This dataset is then used to create classification models which can predict the state (normal or fraud) of new records. There are numerous model creation methods for a typical two class classification task such as rule induction,[1] decision trees,[2] and neural networks.[3] This approach is proven to reliably detect most fraud tricks which have been observed before,[4] it also known as misuse detection.

The second approach deals with unsupervised methodologies which are based on account behavior. In this method a transaction is detected fraudulent if it is in contrast with user‟s normal behavior. This is because we don’t expect fraudsters behave the same as the account owner or be aware of the behavior model of the owner.[5] To this aim, we need to extract the legitimate user behavioral model (e.. user profile)for each account and then detect fraudulent

activities according to it. Comparing new behaviors with this model, different enough activities are distinguished as frauds. The profiles may contain the activity information of the account; such as merchant types, amount, location and time of transactions,[6] This method is also known as anomaly detection.

It is important to highlight the key differences between user behavior analysis and fraud analysis approaches. The fraud analysis method can detect known fraud tricks, with a low false positive rate. These systems extract the signature and model of fraud tricks presented in oracle dataset and can then easily determine exactly which frauds, the system is currently experiencing. If the test data does not contain any fraud signatures, no alarm is raised. Thus, the false positive rate can be reduced extremely. However, since learning of a fraud analysis system (i.e. classifier) is based on limited and specific fraud records, It cannot detect novel frauds. As a result, the false negatives rate may be extremely high depending on how ingenious are the fraudsters. User behavior analysis, on the other hand, greatly addresses the problem of detecting novel frauds. These methods do not search for specific fraud patterns, but rather compare incoming activities with the constructed model of legitimate user behavior. Any activity that is enough different from the model will be considered as a possible fraud. Though, user behavior analysis approaches are powerful in detecting innovative frauds, they really suffer from high rates of false alarm. Moreover, if a fraud occurs during the training phase, this fraudulent behavior will be entered in baseline mode and is assumed to be normal in further analysis.[7] In this section we will briefly introduce some current fraud detection techniques which are applied to credit card fraud detection tasks, also main advantage and disadvantage of each approach will be discussed.

2.2 Artificial Neural Network

An artificial neural network (ANN) is a set of interconnected nodes designed to imitate the functioning of the human brain.[9] Each node has a weighted connection to several other nodes in adjacent layers. Individual nodes take the input received from connected nodes and use the weights together with a simple function to compute output values. Neural networks come in many shapes and architectures. The Neural network architecture, including the number of hidden layers, the number of nodes within a specific hidden layer and their connectivity, most be specified by user based on the complexity of the problem. ANNs can be configured by supervised, unsupervised or hybrid learning methods.

2.3 Supervised techniques

In supervised learning, samples of both fraudulent and non-fraudulent records, associated with their labels are used to create models. These techniques are often used in fraud analysis approach. One of the most popular supervised neural networks is back propagation network (BPN). It minimizes the objective function using a multi-stage dynamic optimization methodthat is a generalization of the delta rule. The back propagation method is often useful for feed-forward network with no feedback. The BPN algorithm is usually time-consuming and parameters like the number of hidden neurons and learning rate of delta rules require extensive tuning and trainingto achieve the best performance.[10] In the domain of fraud detection, supervised neural networks like back-propagation are known as efficient tool that have numerous applications.[11,12,13]

RaghavendraPatidar, *et al*.[14] used a dataset to train a three layers backpropagation neural network in combination with genetic algorithms (GA)[15] forcredit card fraud detection. In this work, genetic algorithms was responsible for making decision about the network architecture, dealing with the network topology, number of hidden layers and number of nodes in each layer.

Also, Aleskerov *et al*.[16] developed a neural network based data mining system for credit card fraud detection. The proposed system (CARDWATCH) had three layers autoassociativearchitectures. They used a set of synthetized data for training and testing the system. The reportedresultsshow very successfulfraud detection rates.

In,[17] a P-RCE neural network was applied for credit card fraud detection. P-RCE is a type of radial-basis function networks,[18,19] that usually applied for pattern recognition tasks. Krenker*et al*. proposed a model for real time fraud detection based on bidirectional neural networks.[20] They used a large data set of cell phone transactions provided by a credit card company. It was claimed that the system outperforms the rule based algorithms in terms of false positive rate.

Again in.[21] a parallel granular neural network (GNN) is proposed to speed up data mining and knowledge discoveryprocess for credit card fraud detection. GNN is a kind of fuzzy neural network based on knowledge discovery (FNNKD).The underlying dataset was extracted from SQL server database containing sample Visa Card transactions and then preprocessed for applying in fraud detection. They obtained less average training errors in the

presence of larger training dataset.

2.4 Unsupervised techniques

The unsupervised techniques do not need the previous knowledge of fraudulent and normalrecords. These methodsraise alarm for those transactions that are most dissimilar from the normalones. These techniques are often used in user behavior approach. ANNs can produce acceptable result for enough large transaction dataset. They need a long training dataset. Self- organizing map (SOM) is one of the most popular unsupervised neural networks learning which was introduced by.[22] SOM provides a clustering method, which is appropriate for constructing and analyzing customer profiles, in credit card fraud detection, as suggested in.[23] SOM operates in two phase: training and mapping. In the former phase, the map is built and weights of the neurons are updated iteratively, based on input samples,[24] in latter, test data is classified automatically into normal and fraudulent classes through the procedure of mapping. As stated in,[25] after training the SOM, new unseen transactions are compared to normal and fraud clusters, if it is similar to all normal records, it is classified as normal. New fraud transactions are also detected similarly.

One of the advantages of using unsupervised neural networks over similar techniques is that these methods can learn from data stream. The more data passed to a SOM model, the more adaptation and improvement on result is obtained. More specifically, the SOM adapts its model as time passes. Therefore it can be used and updated online in banks or other financial corporations. As a result, the fraudulent use of a card can be detected fast and effectively. However, neural networks has some drawbacks and difficulties which are mainly related to specifying suitable architecture in one hand and excessive training required for reaching to best performance in other hand.

2.5 Hybrid supervised and unsupervised techniques

In addition to supervised and unsupervised learning models of neural networks, some researchers have applied hybrid models. John Zhong Leiet. Al.[26] proposed hybrid supervised (SICLN) and unsupervised (ICLN)learning network for credit card fraud detection. They improved the reward only rule of SICLNmodel to ICLN in order to update weights according to both reward and penalty. This improvement appeared in terms of increasing stability and reducing the training time. Moreover, the number of final clusters of the ICLN is independent from the number of initial network neurons. As a result the inoperable neurons can be omitted from the clusters by applying the penalty rule. The results indicated that both the ICLN and

the SICLN havehigh performance, but the SICLN outperforms well-known unsupervised clustering algorithms.

2.6 Artificial Immune System (AIS)

The natural immune system is a highly complex system, comprised of an intricate network of specialized tissues, organs, cells and chemical molecules. These elements are interrelated and act in a highly co- ordinate and specific manner when they recognize, remember disease causing foreign cells and eliminate them. Any element that could be recognized by the immune system is named an antigen. The immune system‟s detectors are the antibodies that are capable to recognition and destruction harmful and risky antigens.[27]

The immune system consists of the two main response of immune and defense: innate immune response and acquired immune response. The body‟s first response for defense is made of the outer, unbroken skin and the „mucus membranes‟ lining internal channels, such as the respiratory and digestive tracts. If the harmful cells could pass through innate immune defense the acquired immunity will defense. In fact, adaptive immune response performs based on antigen-specific recognition of almost unlimited types of infectious substances, even if previously unseen or mutated. It is worth mentioning that the acquired immune response is capable of "remembering" every infection, so that a second exposure to the same pathogen is dealt with more efficiently.

There are two organs responsible for the generation and development of immune cells: the bone marrow and the thymus. The bone marrow is the site where all blood cells are generated and where some of them are developed. The thymus is the organ to which a class of immune cells named T- cells migrates and maturates.[28] There exist a great number of different immune cells, but lymphocytes (white blood cells), are the prevailing ones. Their main function is distinguishing self- cells, which are the human body cells, from non-self cells, the dangerous foreign cells (the pathogens). Lymphocytes are classified into two main types: B-cells and T-cells, both originated in the bone marrow. Those lymphocytes that develop within the bone marrow are named B-cells, and those that migrate to and develop within the thymus (the organ which is located behind the breastbone) are named T-cells.

Artificial Immune System (AIS) is a recent sub field based on the biological metaphor of the immune system.[29] The immune system can distinguish between self and non-self-cells, or more specific, between harmful cells (called as pathogens) and other cells. The ability to

recognize differences in patterns and being all to detect and eliminate infections precisely has attracted the engineer‟s intention in all fields.

Researchers have used the concepts of immunology in order to develop a set of algorithms, such as negative selection algorithm,[30] immune networks algorithm,[31] clonal selection algorithm,[32] and the dendritic cells algorithm,[33]

2.7 Negative Selection

Negative Selection Algorithm or NSA proposed by,[34] is a change detection algorithm based on the T-Cells generation process of biological immune system. It is one of the earliest AIS algorithms applied in various real-world applications. Since it was first conceived, it has attracted many researchers and practitioners in AIS and has gone through some phenomenal evolution. NSA has two stages: generation and detection. In generation stage, the detectors are generated by some random process and censored by trying to match self samples. Those candidates that match (by affinity of higher than affinity threshold) are eliminated and the rest are kept as detectors. In detection stage, the collection of detectors (or detector set) is used in checking whether an incoming data instance is self or non-self. If it matches (by affinity of higher than affinity threshold) any detector, it is claimed as non-self or anomaly.

Brabazon*et.al,*[35] proposed an AIS based model for online credit card fraud detection. Three AIS algorithms were implemented and their performance was standardized against a logistic regression model. Their three chosenalgorithms were the unmodified negative selection Algorithm, the modified negative selection algorithm and the Clonal selection algorithm. They proposed the Distance Value Metric for calculating distance between records. This metric is based on the probability of data occurrence in the training set. Where the detection rate increased, but the number of false alarms and missed frauds remained.

2.8 Clonal selection

Clonal selection theory is used by the immune system to explain the basic features of an immune response to an antigenic stimulus. The selection mechanism guarantees that only those clones (antibodies) with higher affinity for the encountered antigen will survive. On the basis of clonal selection principle, clonal selection algorithm was initially proposed in.[36] and formally explained in.[37] The general algorithm was called CLONALG.

Gadi *et.al* in[36] applied the AIRS in fraud detection on credit card transactions. AIRS is a

classification algorithm that is based on AIS which applies clonal selection to create detectors. AIRS generatesdetectors for all of the classes in the database and in detectionstage uses k Nearest Neighbor algorithm (also called K-NN)in order to classify eachrecord. They compared their method with other methods like the neural networks, Bayesian networks, and decision trees and claimed that, after improving the input parameters for all the methods, AIRS has show the best results of all, partly perhaps since the number of input parameters for AIRS is comparatively high. If we consider a particular training dataset, and set the parameters depending on the same database, the results indicate a tendency to improve. The experiment was carried out on Weka package.

Soltani *et.al* in,[8] proposed AIRS on credit card fraud detection. Since AIRS has a long training time, authors have implemented the model in Cloud Computing environment to shorten this time. They had used MapReduce API which works based on Hadoop distributed file system, and runs the algorithm in parallel.

2.9 Immune Network

The nature immune system is applied through the interactions between a huge numbers of different types of cells. Instead of using a central coordinator, the nature immune systems sustain the appropriate level of immune responses by maintaining the equilibrium status between antibody suppression and stimulation using idiotypes and paratopes antibodies,[38,39] The first Artificial Immune Network (AIN) proposed by.[40] Neal M.*et.al*,[41] introduced the AISFD, which adopted thetechniques developed by CBR (case based reasoning) communityand applied various methods borrowed from genetic algorithm andother techniques to clone the B cells (network nodes) for mortgage fraud detection.

2.10 Danger Theory

The novel immune theory, named Danger Theory was proposed in 1994.[42] It embarked from the concept that defined “self-non-self” in the traditional theories and emphasizes that the immune system does not respond to “non-self” but to danger. According to the theory a useful evolutionarily immune system should focus on those things that are foreign and dangerous, rather than on those that are simply foreign.[43] Danger is measured by damage inflicted to cells indicated by distress signals emitted when cells go through an unnatural death (necrosis).

Dendritic cells (DCs), part of the innate immune system, interact with antigens derived from

the host tissue; therefore, the algorithm inspired by Danger Theory is named Dendritic cell algorithm. Dendritic cells control the state of adaptive immune system cells by emitting the following signals:

- PAMP (pathogen associated molecular pattern)
- Danger
- Safe
- Inflammation

PAMP is released from tissue cells following sudden necrotic cell death; actually, the presence of PAMP usually indicates an anomalous situation.

The presence of Danger signals may or may not indicate an anomalous situation; however the probability of an anomaly is higher than the same, under normal circumstances.

Safe signal act as an indicator of healthy tissue.

Inflammation signal is classed as the molecules of an inflammatory response to tissue injury. In fact, the presence of this signal amplifies the above three signals.

DCs exist in a number of different states of maturity, depending on the type of environmental signal present in the surrounding fluid. They can exist in immature, semi-mature or mature forms. Initially, when a DC enters the tissue, it exists in an immature state. DCs which have the ability to present both the antigen and active T-cells are mature. For an immature DC to become mature it should be exposed to PAMP and danger signals predominantly. The immature DCs exposed to safe signals predominantly are termed “semi-mature”; they produce semi-mature DCs output signaling molecule, which has the ability to de-activate the T-cells. Exposure to PAMP, danger and safe signals lead to an increase in co-stimulatory molecules production, which in turn ends up in removal from the tissue and its migration to local lymph nodes.

2.11 Hybrid AIS or methods

Some researchers applied different algorithms (i.e. vaccination algorithm, CART and so on) by AIS algorithm which are presented below:

Wong,[44] presents the AISCCFD prototype proposed to measure and manage the memory population and mutate detectors in real time. In their work both the two algorithms the

vaccination and negative selection were combined. The results were tested for different fraud types. The proposed method demonstrated higher detection rates when vaccination algorithm was applied, but it failed to detect some types of fraud precisely.

Huang *et.al*[45] presented a novel hybrid Artificial Immune inspired model for fraud detection by combining triplealgorithms: CSPRA, the dendritic cell algorithm (DCA), and CART. Though their proposed method had high detection rate and low false alarm, their approach was focused on logging data and limited to VoD (video on demand) systems and not credit card transactions.

Ayara*et.al*,[46] applied AIS to predict failures of ATM[1]. Their approach is enriched by adding a generation of new antibodies from the antigens that correspond to the unpredicted failures.

2.12 Genetic Algorithm (GA)

Inspired from natural evolution, Genetic algorithms (GA), were originally introduced by John Holland.[15] GA searches for optimum solution with a population of candidate solutions that are traditionally represented in the form of binary strings called chromosomes.

The basic idea is that the stronger members of the population have more chance to survive and reproduce. The strength of a solution is its capability to solve the underlying problem which is indicated by fitness. New generation is selected in proportion to fitness among previous population and newly created offspring. Normally, new offspring will be produced by applying genetic operators such as mutation and crossing over on some fitter members of current generation (parents). As generations progress, the solution are evolved and the average fitness of population increases. This process is repeated until some stopping criteria, (i.e. often passing a pre-specified number of generations) is satisfied.

Genetic Programming (GP).[47] is an extension of genetic algorithms that represent each individual by a tree rather than a bit string. Due to hierarchy nature of the tree, GP can produce various types of model such as mathematical functions, logical and arithmetic expressions, computer programs, networks structures, etc.

Genetic algorithms have been used in data mining tasks mainly for feature selection. It is also widely used in combination with other algorithms for parameter tuning and optimization. Due to availability of genetic algorithm code in different programming languages, it is a popular and strong algorithm in credit card fraud detection. However, GA is very expensive in

consuming time and memory. Genetic programming has also various applications in data mining as classification tool.

Ekrem Duman *et al*. developed a method for credit card fraud detection.[48] They defined a cost- sensitive objective function that assigned different cost to different misclassification errors (e.g. false positive, false negative). In this case, the goal of a classifier will be the minimization of overall cost instead of the number of misclassified transactions. This is due the fact that the correct classification of some transactions was more important than others. The utilized classifier in this work was a novel combination of the genetic algorithms and the scatter search. For evaluating the proposed method, it was applied to real data and showed promising result in comparison to literature. Analyzing the influence of the features in detecting fraud indicated that statistics of the popular and unpopular regions for a credit card holder is the most important feature. Authors excluded some type of features such as the MCC and country statistics from their study that resulted in less generality for typical fraud detection problem.

K.Rama Kalyani *et al*.[49] presented a model of credit card fraud detection based on the principles of genetic algorithm. The goal of the approach was first developing a synthetizing algorithm for generating test data and then to detect fraudulent transaction with the proposed algorithm.

Bentley et al.[50] developed a genetic programming based fuzzy system to extract rules for classifying data tested on real home insurance claims and credit card transactions.

In,[51] authors applied Genetic Programming to the prediction of the price in the stock market of Japan. The objective of the work was to make decision in stock market about the best stocks as well as the time and amount of stocks to sell or buy. The experimental results showed the superior performance of GP over neural networks.

2.13 Hidden Markov Model (HMM)

A Hidden Markov Model is a double embedded stochastic process which is applied to model much more complicated stochastic processes as compared to a traditional Markov model. The underlying system is assumed to be a Markov process with unobserved states. In simpler Markov models like Markov chains, states are definite transition probabilities are only unknown parameters. In contrast, the states of a HMM are hidden, but state dependent

outputs are visible.

In credit card fraud detection a HMM is trained for modeling the normal behavior encoded in user profiles.[52] According to this model, a new incoming transaction will be classified to fraud if it is not accepted by model with sufficiently high probability. Each user profile contains a set of information about last 10 transactions of that user liketime; category and amount of for each transaction.[52,53,54] HMM produces high false positive rate.[55] V. Bhusari et al.[56] utialized HMM for detecting credit card frauds with low false alarm. The proposed system was also scalable for processing huge number of transactions.

HMM can also be embedded in online fraud detection systems which receive transaction details and verify whether it is normal or fraudulent. If the system confirms the transaction to be malicious, an alarm is raised and related bank rejects that transaction. The responding cardholder may then be informed about possible card misuse.

2.14 Support Vector Machine (SVM)

Support vector machine (SVM).[57] is a supervised learning model with associated learning algorithms that can analyze and recognize patterns for classification and regression tasks.[48] SVM is a binary classifier. The basic idea of SVM was to find an optimal hyper-plane which can separate instances of two given classes, linearly. This hyper-plane was assumed to be located in the gap between some marginal instances called support vectors. Introducing the kernel functions, the idea was extended for linearly inseparable data. A kernel function represents the dot product of projections of two data points in a high dimensional space. It is a transform that disperses data by mapping from the input space to a new space (feature space) in which the instances are more likely to be linearly separable. Kernels, such as radial basis function (RBF), can be used to learn complex input spaces. In classification tasks, given a set of training instances, marked with the label of the associated class, the SVM training algorithm find a hyper-plane that can assign new incoming instances into one of two classes. The class prediction of each new data point is based on which side of the hyper-plane it falls on feature space.

SVM has been successfully applied to a broad range of applications such as.[58,59,60] In credit card fraud detection, Ghosh and Reilly,[61] developed a model using SVMs and admired neural networks. In this research a three layer feed-forward RBF neural network applied for detecting fraudulent credit card transactions through only two passes required to churn out a

fraud score in every two hours.

Tung-shou Chen *et al*.[62] proposed a binary support vector system (BSVS), in which support vectors were selected by means of the genetic algorithms (GA). In proposed model self-organizing map (SOM) was first applied to obtain a high true negative rate and BSVS was then used to better train the data according their distribution.

In,[63] a classification model based on decision trees and support vector machines (SVM) was constructed respectively for detecting credit card fraud. The first comparative study among SVM and decision tree methods in credit card fraud detection with a real data set was performed in this paper. The results revealed that the decision tree classifiers such as CART outperform SVM in solving the problem under investigation.

Rongchang Chen *et al*.[64] suggested a novel questionnaire-responder transaction (QRT) approach with SVM for credit card fraud detection. The objective of this research was the usage of SVM as well as other approaches such as Over-sampling and majority voting for investigating the prediction accuracy of their method in fraud detection. The experimental results indicated that the QRT approach has high degree of efficiency in terms of prediction accuracy.

Qibei Lu *et al*.[65] established a credit card fraud detection model based on Class Weighted SVM. Employing Principal Component Analysis (PCA), they initially reduced data dimension to less synthetic composite features due to the high dimensionality of data. Then according to imbalance characteristics of data, an improved Imbalance Class Weighted SVM (ICW-SVM) was proposed.

2.15 Bayesian Network

A Bayesian network is a graphical model that represents conditional dependencies among random variables. The underlying graphical model is in the form of directed acyclic graph. Bayesian networks are usefulfor finding unknown probabilities given known probabilitiesin the presence of uncertainty.[66] Bayesian networks can play an important and effective role in modeling situations where some basic information is already known but incoming data is uncertain or partially unavailable.[67,68,69] The goal of using Bayes rules is often the prediction of the class label associated to a given vector of features or attributes.[70] Bayesian networks have been successfully applied to various fields of interest for instance churn prevention.[71]

in business, pattern recognition in vision,[72] generation of diagnostic in medicine.[73] and fault diagnosis,[74] as well as forecasting.[75] in power systems. Besides, these networks have also been used to detect anomaly and frauds in credit card transactions or telecommunication networks.[76,77,5]

In,[70] two approaches are suggested for credit card fraud detection using Bayesian network. In the first, the fraudulent user behavior and in the second the legitimate (normal) user behavior are modeled by Bayesian network. The fraudulent behavior net is constructed fromexpert knowledge, while the legitimate net is set up in respect to available data from non fraudulent users. During operation, legitimatenet is adapted to a specific user based on emerging data.Classification of new transactions were simplyconducted by inserting it to both networksand then specify the type of behavior (legitimate/fraud) according to correspondingprobabilities. Applying Bayes rule, gives the probability of fraud for new transactions.[78] Again, Ezawa and Nortondeveloped a four-stage Bayesian network.[79] They claimed that lots of popular methods such as regression, K-nearest neighbor and neural networks takes too long time to be applicable in their data.

2.16 Fuzzy Logic Based System

Fuzzy logic based system is the system based on fuzzy rules. Fuzzy logic systems address the uncertainty of the input and output variables by defining fuzzy sets and numbers in order to express values in the form of linguistic variables (e.g. small, medium and large). Two important types of these systems are fuzzy neural network and fuzzy Darwinian system.

2.17 Fuzzy Neural Network (FNN)

The aim of applying Fuzzy Neural Network (FNN) is to learn fromgreatnumber of uncertain and imprecise records of information, which is very common in real world applications.[80] Fuzzy neural networks proposedin,[81] to accelerate rule induction for fraud detection in customer specific credit cards. In this research authors applied GNN (Granular Neural Network) method which implements fuzzy neural network based on knowledge discovery (FNNKD), for accelerating thetraining network and detecting fraudster in parallel.

2.18 Fuzzy Darwinian System

Fuzzy Darwinian Detection,[82] is a kind of Evolutionary-Fuzzy system that uses genetic programming in order to evolve fuzzy rules. Extracting the rules, the systemcan classify the transactions into fraudulent and normal. This system was composed of genetic programming

(GP) unit in combination with fuzzy expert system. Results indicated that the proposed system has very high accuracy and low false positive rate in comparison with other techniques, but it is extremely expensive.[83]

2.19 Expert Systems

Rules can be generated from information which are obtained from a human expert and stored in a rule-based system as IF-THEN rules. Knowledge base system or an expert system is the information which is stored in Knowledge base. The rules in the expert system appliedin order to perform operations on a data to inference to reach appropriate conclusion. Powerful and flexible solutions for many application problemsprovides by expert system. Financial analysis and fraud detection are one of the general areas which it can be apply. By applying expert system suspicious activity or transaction can be detected from deviations from "normal' spending patterns.[84]

In,[85] authors presented a model to detect credit card frauds in various payment channels. In their model fuzzy expert system gives the abnormal degree which determines how the new transaction is fraudulent in comparison with user behavioral. The fraud tendency weight is achieved by user behavioral analysis. So, this system is named FUZZGY. Also, another research,[86] proposed expert system model to detect fraud for alert financial institutions.

2.20 Inductive logic programming (ILP)

ILP by using a set of positive and negative examples uses first order predicate logic to define a concept. This logic program is then used to classify new instances. Complex relationship among components or attributes can be easily expressed, in this approach of classification. The effectiveness of the system improves by domain knowledge which can be easily represented in an ILP system.[87] Muggleton et al.[88] proposed the model applying labeled data in fraud detection which using relational learning approaches such as Inductive Logic Programming (ILP) and simple homophily- based classifiers on relational databases. Perlich, et al.[89] also propose novel target-dependent detection techniques for converting the relational learning problem into a conventional one.

2.21 Case-based reasoning (CBR)

Adapting solutions in order to solve previous problems and use them to solve new problems is the basic idea of CBR. In CBR, cases introduce as descriptions of past experience of human specialists and stored in a database which uses for later retrieval when the user

encounters a new case with similar parameters. These cases can apply for classification purposes. A CBR system attempts to find a matching case when face with a new problem. In this method the model defined as the training data, and in test phase when a new case or instance is given to the model it looks in all the data to discover a subset of cases that are most similar to new case and uses them to predict the result.

Nearest neighbor matching algorithm usually applied with CBR, although there are several another algorithms which used with this approach such as.[90]

Case-based reasoning is well documented both as the framework for hybrid frauddetection systems and as an inference engine in.[91]

Also, E.b. Reategui applied hybrid approaches of CBR and NN which divides the task of fraud detection into two separate components and found that this multiple approach was more effective than either approach on its own.[92] In this model, CBR looks for best matches in the case base while an artificial neural net (ANN) learns patterns of use and misuse of credit cards. The case base included information such as transaction amounts, dates, place and type, theft date, and MCC (merchant category code). The hybrid CBR and ANN system reported a classification accuracy of 89% on a case base of 1606 cases.

Table 1: Advantages and disadvantages of fraud detection methods.

Techniques	Advantages	Disadvantages
Artificial Neural Network (ANN)	Ability to learn from the past/lack of need to be reprogrammed/ Ability to extract rules and predict future activities based on the current situation/ High accuracy/ Portability/ high speed in detection/ the ability to generate code to be used in real-time systems/ the easiness to be built and operated/ Effectiveness in dealing with noisy data, in predicting patterns, in solving complex problems, and in processing new instances/Adaptability /Maintainability /knowledge discovery and data miming	Difficulty to confirm the structure/high processing time for large neural networks and excessive training/ poor explanation capability/ difficult to setup and operate/high expense/ non numerical data need to be converted and normalized/Sensitivity to data format.
Artificial Immune System (AIS)	High capability in pattern recognition/powerful in Learning and memory/Self-organization/ easy in integration with other systems/dynamically changing	Need high training time in NSA/ poor in handle missing data in ClonalG and NSA

		coverage/ self Identity/ multilayered/ has diversity/ noise tolerance/ fault tolerance/ predator-prey dynamics/ Inexpensive / no need to training phase in DCA.	
Genetic Algorithm		Works well with noisy data/easy to integrate with other systems/ usually combined into other techniques to increase the performance of those techniques and optimize their parameters/ easy in build and operate/In expensive/fast in detection/ Adaptability/Maintainability/knowledge discovery and data miming	Requires extensive tool knowledge to set up and operate and difficult to understand.
Hidden Markov Model (HMM)		Fast in detection	Highly expensive/ low accuracy/not scalable to large size data sets
Support Vector Machines (SVM)		SVMs deliver a unique solution, since the optimality problem is convex/by choosing an appropriate generalization grade,SVMs can be robust, even when the training sample has some bias.	Poor in process largedataset/expensive/has low speed of detection/ medium accuracy/lack of transparency of results
Bayesian Network		High processing and detection speed/high accuracy	Excessive training need/ expensive
Fuzzy Logic Based System	**Fuzzy Neural Network**	Very fast in detection/good accuracy	Expensive
	Fuzzy Darwinian System	Very high accuracy/ Maintainability	Has very low speed in detection/ High expensive
Expert System		Easy to modify the KB/ easy to develop and build the system/ easy to manage complexity or missing information/high degree of accuracy/ explanation facilities/good performance/Rules from other techniques such as NN and DT can be extracted, modified, and stored in the KB.	Poor in handling missing information or unexpected data values/poor in process different data types /knowledge representation languages do not approach human flexibility/ poor in build and operate/ poor in integration
Inductive logic programming (ILP)		Powerful in process different data types/ powerful modeling language that can model complex relationships/powerful in handle missing data	Has low predictive accuracy/extremely sensitive to noise/ their performance deteriorates rapidly in the presence of spurious data.
Case based reasoning (CBR)		Useful in domain that has a large number of examples/ has the ability to work with incomplete or noisy data/effective/ flexible/ easy to update and maintain/can be used in a hybrid approach.	May suffer from the problem of incomplete or noisy data.

Decision tree (DT)	High flexibility/good haleness/ explainable/easy to implement/easy to display and to understand	Requirements to check each condition one by one. In fraud detection condition is transaction.

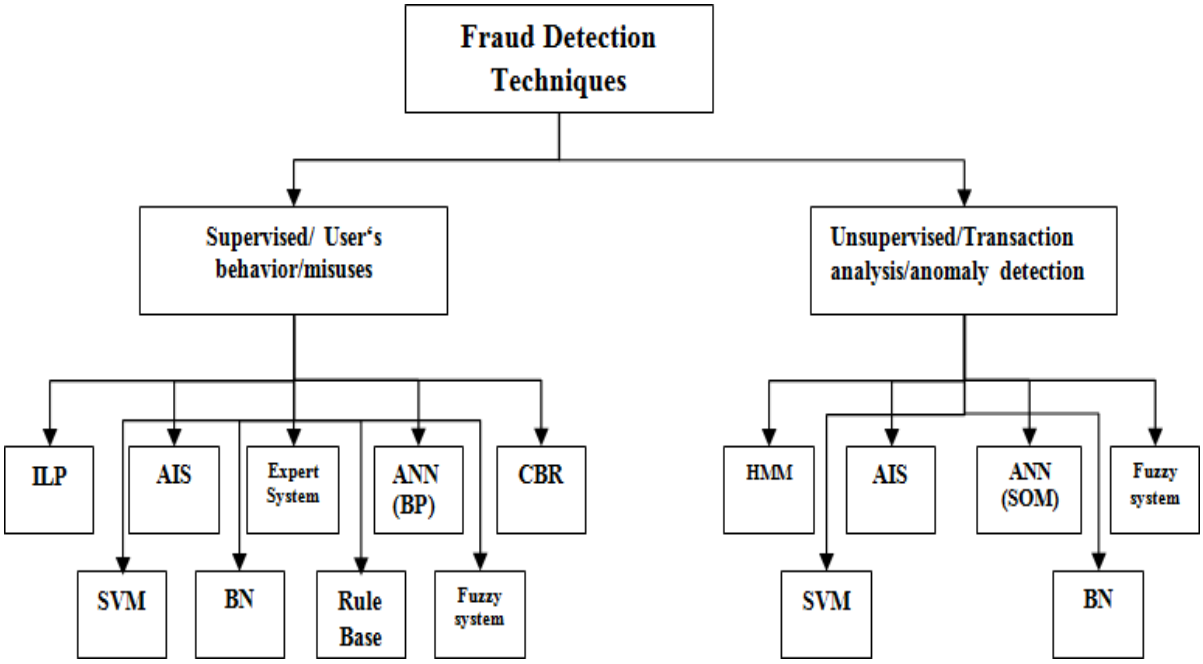

Fig. 2 A complete classification of credit card fraud detection techniques

3 METHODOLOGY

3.1 Data set and evaluation

The volume of fraud in every dataset is different. This might be because of the different security protocols used by different organizations and banks and so on. Whatever the reason is, this fact causes different fraud characteristics on each dataset, which affects the performance of the fraud detection system. Therefore considering the dataset's characteristics will help the system having more precise results.

A proper data set is a data set which covers various fraud and several attributes of customer profile or behavior. The contribution of attributes is a critical factor that should be considered. Also, a proper data set should be able to reflect the real world of credit card.

Credit card transaction datasets usually divided in to two types:

(i) Numerical and

(ii) Categorical attributes.

In statistics, categorical data is a statistical data type consisting of categorical variables, used for observed data whose value is one of a fixed number of nominal categories, or for data that

has been converted into that form, for example as grouped data. However numeric data are numbers like age, cost, etc.

In fraud detection applications customer's gender and name are the typical numerical attribute, and categorical attributes are those like merchant category code, date of transaction, amount of transaction and etc. Some of these categorical variables can, depending on the dataset, have hundreds and thousands of categories.

Finally, Fig. 3 shows a complete classification in two groups: numerical and categorical attributes which is suitable for each algorithm.

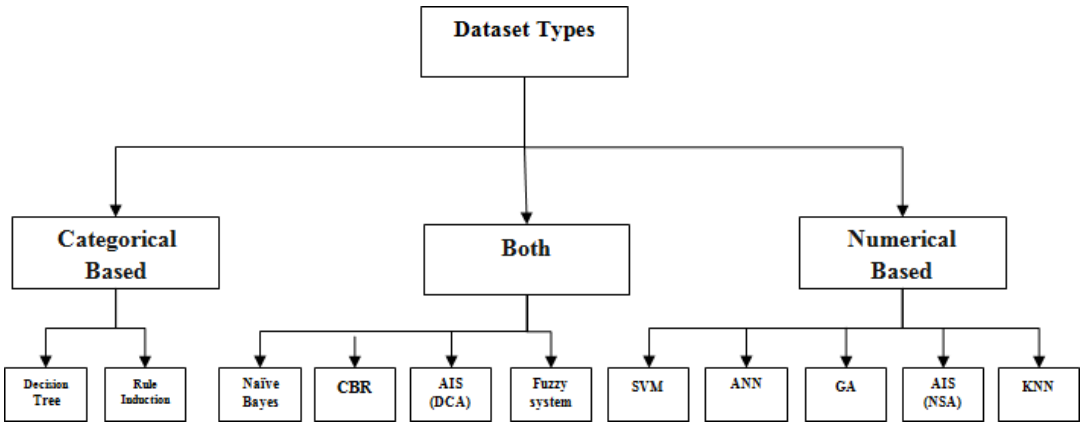

Fig. 3: A complete classification of the dataset's attribute.

- **Evaluation**

There are a variety of measures for various algorithms and these measures have been developed to evaluate very different things. So it should be criteria for evaluation of various proposed method. False Positive (FP), False Negative (FN), True Positive (TP), and True Negative (TN) and the relation between them are quantities which usually adopted by credit card fraud detection researcher to compare the accuracy of different approaches. The definitions of mentioned parameters are presented below:

- FP: the false positive rate indicates the portion of the non-fraudulent transactions wrongly being classified as fraudulent transactions.
- FN: the false negative rate indicates the portion of the fraudulent transactions wrongly being classified as normal transactions.
- TP: the true positive rate represents the portion of the fraudulent transactions correctly being classified as fraudulent transactions.
- TN: the true negative rate represents the portion of the normal transactions correctly

being classified as normal transactions.

Table 2 shows the details of the most common formulas which are used by researchers for evaluation of their proposed methods.

Table 2: Evaluation criteria for credit card fraud detection.

Measure	Formula	Description	Used in
Accuracy (ACC)/Detection rate	TN + TP/TP + FP + FN + TN	Accuracy is the percentage of correctly classified instances. It is one the most widely used classification performance metrics	Nicholas Wong et al. (2012)[97] Manoel Fernando, et al. (2008)[36] soltani et al.[8],A. Brabazon et al. (2011)[35], Siddhartha et al. (2008)[59], P. Ravisankar et al. (2011)[99], AbhinavSrivastava et al. (2008)[52], John Zhong et al. (2012)[26], Qibei Lu et al. (2011)[65], AmlanKundu (2006)[98]
Precision/Hit rate	TP/TP + FP	Precision is the number of classified positive or fraudulent instances that actually are positive instances.	Manoel Fernando, et al. (2008)[36], Siddhartha et al. (2008)[50], John Zhong et al. (2012)[26], Qibei Lu et al. (2011)[65], AmlanKundu (2006)[98]
True positive rate/Sensitivity	TP/TP + FN	TP (true positive) is the number of correctly classified positive or abnormal instances. TP rate measures how well a classifier can recognize abnormal records. It is also called sensitivity measure. In the case of credit card fraud detection, abnormal instances are fraudulent transactions.	Maes S. et al. (2002)[5], Siddhartha et al. (2008)[59], Tao guo et al. (2008)[93], P. Ravisankar et al. (2011)[99], AbhinavSrivastava et al. (2008)[52], John Zhong et al. (2012)[26], Qibei Lu et al. (2011)[65], AmlanKundu (2006)[98]
True negative rate /Specificity	TN/TN + FP	TN (true negative) is the number of correctly classified negative or normal instances. TN rate measures how well a classifier can recognize normal records. It is also	Siddhartha et al. (2008)[59] Philip K. Chan (1999)[95], Tao guo et al. (2008)[93], P. Ravisankar et al. (2011)[99], John Zhong et al. (2012)[26], Maes S. et al. (2002)[5], Qibei Lu et al. (2011)[65], AmlanKundu (2006)[98]

		called specificity measure.	
False positive rate (FPR)	FP/FP+TN	Ratio of credit card fraud detected incorrectly	Nicholas Wong et al. (2012)[97], soltani et al.[8], Maes S. et al. (2002)[5], Philip K. Chan (1999)[95], AbhinavSrivastava et al. (2008)[52], John Zhong et al. (2012)[26], Qibei Lu et al. (2011)[65], AmlanKundu (2006)[98]
ROC	True positive rate plotted against false positive rate	Relative Operating Characteristic curve, a comparison of TPR and FPR as the criterion changes	Manoel Fernando, et al. (2008)[36], Maes S. et al. (2002)[5], Tao guo et al. (2008)[93],John Zhong et al. (2012)[26], Qibei Lu et al. (2011)[65],AmlanKundu (2006)[98]
Cost	Cost = 100 * FN + 10 * (FP +TP)		Manoel Fernando, et al. ,(2008)[36], soltani et al.[8], Philip K. Chan (1999)[95], Qibei Lu et al. (2011)[65]
F1-measure	2 × (Precision ×Recall)/(Precision +Recall)	Weighted average of the precision and recall	Siddhartha et al. (2008)[59]

The goal of all algorithms and techniques is to minimize FP and FN rate and maximize TP and TN rate and with a good detection rate at the same time.

3.2 Pattern Database Construction Using Frequent Item set Mining (Training)

Frequent itemsets are sets of items that occur simultaneously in as many transactions as the user defined minimum support. The metric support(X) is defined as the fraction of records of database D that contains the itemset X as a subset:

Support (X) = Count (X)/ │ D │ ---------- (1)

For example, if the database contains 1000 records and the itemset X appears in 800 records, then the support(X) = 800/1000 = 0.8 = 80%; that is, 80% of transactions support the itemset X.

In credit card transaction data, the legal pattern of a customer is the set of attribute values specific to a customer when he does a legal transaction which shows the customer behavior. It is found that the fraudsters are also behaving almost in the same manner as that of a

customer.[1] This means that fraudsters are intruding into customer accounts after learning their genuine behavior only. Therefore, instead of finding a common pattern for fraudster behavior it is more valid to identify fraud patterns for each customer. Thus, in this research, we have constructed two patterns for each customer—legal pattern and fraud pattern. When frequent pattern mining is applied to credit card transaction data of a particular customer, it returns set of attributes showing same values in a group of transactions specified by the support. Generally the frequent pattern mining algorithms like that of Apriori[31] return many such groups and the longest group containing maximum number of attributes is selected as that particular customer's legal pattern. The training (pattern recognition) algorithm is given below.

Step 1. Separate each customer's transactions from the whole transaction database D
Step 2. From each customer's transactions separate his/her legal and fraud transactions.
Step 3. Apply Apriori algorithm to the set of legal transactions of each customer. The Apriori algorithm returns a set of frequent item sets. Take the largest frequent item set as the legal pattern corresponding to that customer. Store these legal patterns in legal pattern database.
Step 4. Apply Apriori algorithm to the set of fraud transactions of each customer. The Apriori algorithm returns a set of frequent item sets. Take the largest frequent item set as the fraud pattern corresponding to that customer. Store these fraud patterns in fraud pattern database.

The pseudo code of training algorithm is given in Algorithm 1.
Input: Customer Transactions Database D, Support S
Output: Legal Pattern Database LPD, Fraud Pattern Database FPD

Begin
Group the transactions of each customer together.
Let there are "n" groups corresponds to "n" customers f **or i** = 1 to n **do**
Separate each group Gi into two different groups LGi and FGi of legal and fraud transactions.
Let there are "m" legal and "k" fraud transactions
FIS = Apriori (LGi, S, m); //Set of frequent itemset LP = max (FIS); //Large Frequent Itemset
LPD(i) = LP;
FIS = Apriori (FGi, S, k); //Set of frequent itemset FP = max (FIS); //Large Frequent Itemset
FPD(i) = FP;
End for
Return LPD & FPD;

End

Fraud Detection Using Matching Algorithm (Testing)

After finding the legal and fraud patterns for each customer, the fraud detection system traverses these fraud and legal pattern databases in order to detect frauds. These pattern databases are much smaller in size than original customer transaction databases as they contain only one record corresponding to a customer. This research proposes a matching algorithm which traverses the pattern databases for a match with the incoming transaction to detect fraud. If a closer match is found with legal pattern of the corresponding customer, then the matching algorithm returns "0" giving a green signal to the bank for allowing the transaction. If a closer match is found with fraud pattern of the corresponding customer, then the matching algorithm returns "1" giving an alarm to the bank for stopping the transaction. The size of pattern databases is $n \times k$ where n is the number of customers and k is the number of attributes. The matching (testing) algorithm is explained below.

Step 1. Count the number of attributes in the incoming transaction matching with that of the legal pattern of the corresponding customer. Let it be lc.

Step 2. Count the number of attributes in the incoming transaction matching with that of the fraud pattern of the corresponding customer. Let it be fc.

Step 3. If fc = 0 and lc is more than the user defined matching percentage, then the incoming transaction is legal.

Step 4. If lc = 0 and fc is more than the user defined matching percentage, then the incoming transaction is fraud.

Step 5. If both fc and lc are greater than zero and $fc \geq lc$, then the incoming transaction is fraud or else it is legal.

The pseudocode of the testing algorithm is given in Algorithm 2.

Input: Legal Pattern Database LPD, Fraud Pattern Database FPD, Incoming Transaction T, Number of costumers "n", Number of attributes "k", matching percentage "mp"

Output: 0 (if legal) or 1 (if fraud)

Assumption

1. First attribute of each record in pattern databases and incoming transaction is Customer ID
2. If an attribute is missing in the frequent itemset (ie, this attribute has different values in each transaction and thus it is not contributing to the pattern) then we considered it as invalid.

Begin

lc = 0; //legal attribute match count fc = 0; //fraud attribute match count

for i =1 to n **do**

if (LPD(i, 1) =T(1)) **then** //First attribute

for j =2 to k **do**

if (LPD(i, j) is valid and LPD(i, j) = T(j)) **then**

lc = lc + 1;

endif endfor

endif endfor

for i =1 to n **do**

if (FPD(i, 1)= T(1)) **then for j** =2 to k **do**

if (FPD(i,j) is valid and FPD(i,j) = T(j)) **then**

fc = fc + 1;

endif endfor

endif endfor

if (fc = 0) **then** //no fraud pattern

if ((lc/no. of valid attributes in legal pattern) ≥ mp) **then return** (0); //legal transaction

else return (1); //fraud transaction

endif

elseif (lc = 0) **then** //no legal pattern

if ((fc/no. of valid attributes in fraud pattern) ≥ mp) **then return** (1); //fraud transaction

else return (0); //legal transaction

endif

elseif (lc > 0 && fc > 0) **then** //both legal and fraud patterns are available

if (fc ≥ lc) **then return** (1); //fraud transaction

else return (0); //legal Transaction

endif endif

End

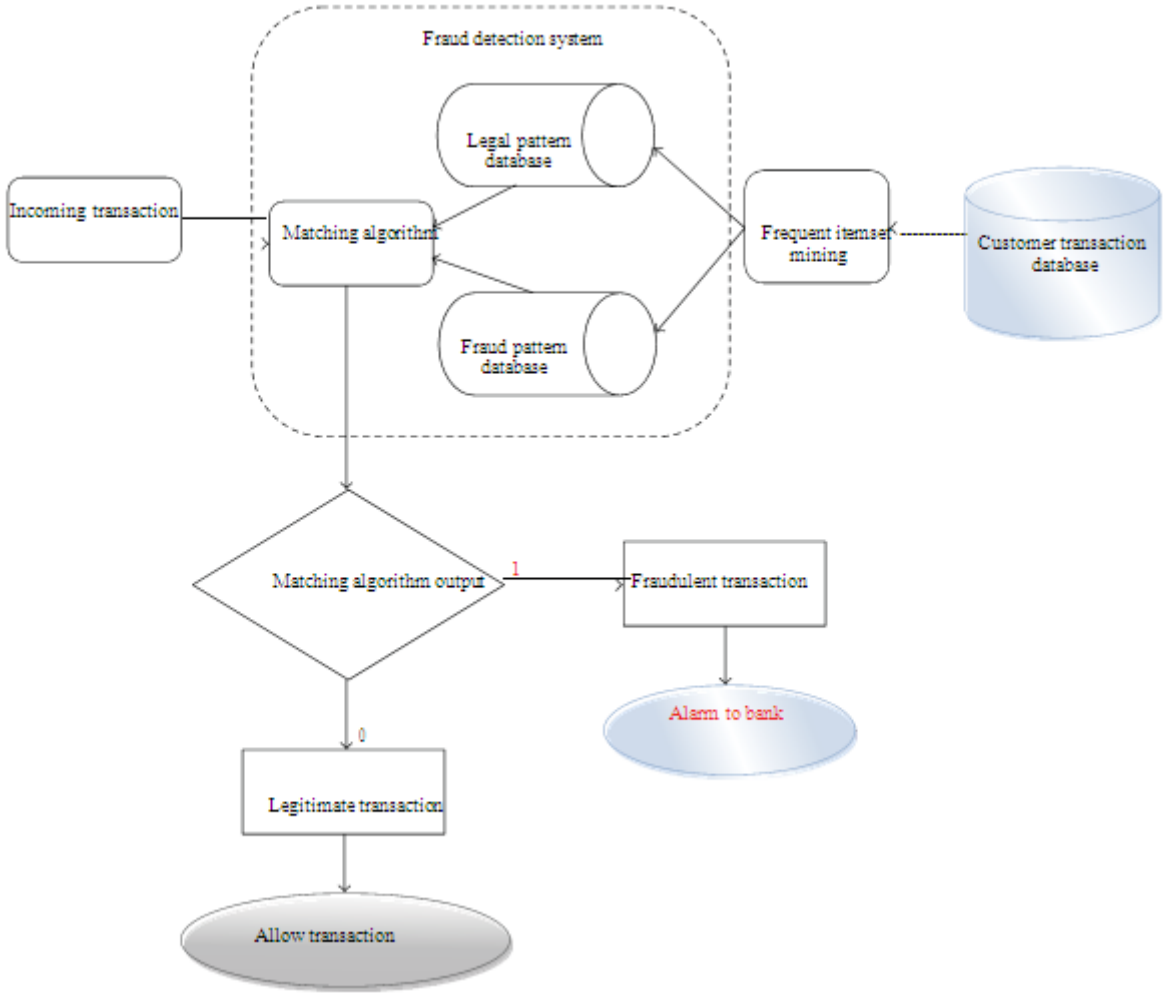

Figure 4: **Proposed credit card fraud detection model.**

4. CONCLUSIONS

Credit card is one alternative of cash payment. Some card holders may abuse their responsibility in credit card usage and repayment. Apart from that, credit card transaction is also prone to fraudulent where unauthorized parties perform illegal transactions using credit cards. Therefore, it is the responsibility of card issuers or the banks to find an effective way to reduce the cost that may incur when the issues above happen. One way to address these issues is via data mining. Due to the characteristics such as overlapping class samples and unbalanced class distribution that exist in credit card data sets, it gives challenges to data mining researchers. On top of that, the weakness of general learning algorithms also contributes to the difficulties of classifying the minority class, which is usually the important class, of the data sets.

The experimental results showed that the proposed MCS outperformed their work. In general,

the proposed MCS demonstrates its superiority in handling the credit data sets that inherit the characteris- tics of overlapping classes and unbalanced class distribution. However, there are rooms to improve the TPR for the minority classes. Other MCS combination strategies have been planned for the current research work, particularly the hybrid combination. Currently, researchers had attempted deep learning algorithms such as long short-term memory (lstm) and deep Belief Networks for detecting anomalies in credit card transactions. We are also considering combining the deep learning algorithms, as in the study of, for promising detection results.

REFERENCES

1. A. Engelbrecht and H. Viktor, Rule Improvement Through Deci- sion Boundary Detection Using Sensitivity Analysis (Lecture Notes in Computer Science). Berlin, Germany: Springer, 1999; 78–84. doi: 10.1007/bfb0100474.
2. A. Fernández, S. Del Río, N. V. Chawla, and F. Herrera, ‘‘An insight into imbalanced Big Data classification: Outcomes and challenges,’’ Complex Intell. Syst., 2017; 3(2): 105–120, doi: 10.1007/s40747- 017-0037-9.
3. A. G. De Sá, A. C. Pereira, and G. L. Pappa, ‘‘A customized classification algorithm for credit card fraud detection,’’ Eng. Appl. Artif. Intell., 2018; 72: 21–29, doi: 10.1016/j.engappai.2018.03.011.
4. A. O. Adewumi and A. A. Akinyelu, ‘‘A survey of machine-learning and nature-inspired based credit card fraud detection techniques,’’ Int.
5. A. Srivastava, A. Kundu, S. Sural, and A. K. Majumdar, “Credit card fraud detection using hidden Markov model,” IEEE Trans- actions on Dependable and Secure Computing, 2008; 5(1): 37–48.
6. Ceicdata. (2019). Malaysia Credit Card Statistics. Accessed: Dec., 2019. [Online]. Available: https://www.ceicdata.com/en/malaysia/credit-card- statistics.
7. D. Excell, “Bayesian inference—the future of online fraud protection,” Computer Fraud and Security, 2012; 2: 8–11.
8. D. J. Weston, “Off-the-peg and bespoke classifiers for fraud detection,” Computational Statistics and Data Analysis, 2008; 52(9): 4521–4532.
9. D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, “Association rules applied to credit card fraud detection,” Expert Systems with Applications, 2009; 36(2): 3630–3640.
10. De Carvalho, ‘‘Data complexity measures for imbalanced classification tasks,’’ in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2018; 1–8, doi: 10.1109/ijcnn.2018.8489661.

11. E. Aleskerov and B. Freisleben, "CARD WATCH: a neural network based database mining system for credit card fraud detection," in Proceedings of the Computational Intelligence for Financial Engineering, 1997; 220–226.
12. E. Duman and M. H. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search," Expert Systems with Applications, 2011; 38(10): 13057–13063.
13. E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature," Decision Support Systems, 2011; 50(3): 559– 569.
14. G. Bing, ''Learning from class-imbalanced data: Review of methods and applications,'' Expert Syst. Appl., May 2017; 73: 220–239, doi: 10.1016/j.eswa.2016.12.035.
15. G. Blunt and D. J. Hand, "The UK credit card market," Tech. Rep., Department of Mathematics, Imperial College, London, UK, 2000.
16. G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Holmes, A. Donkin, and I. H. Witten, "Weka: a machine learning workbench," in Proceedings of the 2nd Australia and New Zealand Conference on Intelligent Information Systems, 1994.
17. G. Lemaitre, F. Nogueira, and C. Aridas, ''Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning,''.
18. G. M. Weiss, ''Mining with rarity,'' SIGKDD Explor. Newsl., 2004; 6(1): 7, doi: 10.1145/1007730.1007734.
19. I. H. Witten, E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, San Francisco, Calif, USA, 3rd edition, 2011.
20. J. Błaszczyński and J. Stefanowski, ''Neighbourhood sampling in bagging for imbalanced data,'' Neurocomputing, Feb. 2015; 150: 529–542, doi: 10.1016/j.neucom.2014.07.064.
21. J. Gao, L. Gong, J. Y. Wang, and Z. C. Mo, ''Study on unbalanced binary classification with unknown misclassification costs,'' in Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage. (IEEM), Dec. 2018; 1538–1542, doi: 10.1109/ieem.2018.8607671.
22. J. Mach. Learn. Res., 2017; 18(1): 1–5.
23. J. Mathew, C. K. Pang, M. Luo, and W. H. Leong, ''Classification of imbal- anced data by oversampling in kernel space of support vector machines,'' IEEE Trans. Neural Netw. Learn. Syst., Sep. 2018; 29(9): 4065–4076, doi: 10.1109/tnnls.2017.2751612.
24. J. Syst. Assur. Eng. Manag., Nov. 2017; 8(S2): 937–953, doi: 10.1007/s13198-016-0551-y.
25. J. T. S. Quah and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence," Expert Systems with Applications, 2008; 35(4): 1721–1732.

26. K. Napierała, J. Stefanowski, and S. Wilk, ‘‘Learning from imbalanced data in presence of noisy and borderline examples,’’ in Rough Sets and Current Trends in Computing. Brookline, MA, USA: Microtome Publishing, 2010; 158–167, doi: 10.1007/978-3-642-13529-3_18.
27. L. Seyedhossein and M. R. Hashemi, “Mining information from credit card time series for timelier fraud detection,” in Proceed- ing of the 5th International Symposium on Telecommunications (IST '10), pp. 619–624, Tehran, Iran, December 2010.
28. M. Syeda, Y.-Q. Zhang, and Y. Pan, “Parallel granular neural networks for fast credit card fraud detection,” in Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ- IEEE ’02), vol. 1, pp. 572–577, Honolulu, Hawaii, USA, May 2002.
29. M. Zareapoor, K. R. Seeja, and A. M. Alam, “Analyzing credit card: fraud detection techniques based on certain design criteria,” International Journal of Computer Application, 2012; 52(3): 35–42.
30. N. Chawla, N. Japkowicz, and A. Kotcz, ‘‘Editorial,’’ ACM SIGKDD Explor. Newslett., 2004; 6(1): doi: 10.1145/1007730.1007733.
31. N. Japkowicz and S. Stephen, ‘‘The class imbalance problem: A systematic study1,’’ IDA, Nov. 2002; 6(5): 429–449, doi: 10.3233/ida- 2002-6504.
32. N. O. Francisca, “Data mining application in credit card fraud detection system,” Journal of Engineering Science and Technol- ogy, 2011; 6(3): 311–322.
33. N. Wong, P. Ray, G. Stephens, and L. Lewis, “Artificial immune systems for the detection of credit card fraud,” Information Systems, 2012; 22(1): 53–76.
34. Nilsonreport. Accessed: 2019. [Online]. Available: https://nilsonreport. com/upload/content_promo/The_Nilson_Report_10-17-2016.pdf.
35. P. Juszczak, N. M. Adams, D. J. Hand, C. Whitrow, and P. K. Chan, W. Fan, A. L. Prodromidis, and S. J. Stolfo, “Distributed data mining in credit card fraud detection,” IEEE Intelligent Systems and Their Applications, 1999; 14(6): 67– 74.
36. P. Vuttipittayamongkol, E. Elyan, A. Petrovski, and C. Jayne, ‘‘Overlap- based undersampling for improving imbalanced data classification,’’ in Proc. Intell. Data Eng. Automated Learn. (IDEAL), 2018; 689–697. doi: 10.1007/978-3-030-03493-1_72.
37. Q. Dong, S. Gong, and X. Zhu, ‘‘Class rectification hard min- ing for imbalanced deep learning,’’ in Proc. IEEE Int. Conf. Com- put. Vis. (ICCV), Oct. 2017, pp. 1851–1860. Accessed, Aug. 2018, doi: 10.1109/iccv.2017.205.
38. Q. Lu and C. Ju, “Research on credit card fraud detection model based on class weighted support vector machine,” Journal of Convergence Information Technology, 2011; 6(1):

62–68.

39. R. J. Bolton and D. J. Hand, "Statistical fraud detection: a review,"
40. R. J. Bolton and D. J. Hand, "Unsupervised profiling methods for fraud detection," in Proceedings of the Conference on Credit Scoring and Credit Control, Edinburgh, UK, 2001.
41. RM51.3mil in payment card fraud losses reported in 2016. Accessed, 2019. [Online]. Available: https://www.thestar.com.my/news/nation/2017/08/03/ rm422mil-in-credit-card-fraud-losses-reported-in-2016.
42. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: a comparative study," Decision Support Systems, 2011; 50(3): 602–613.
43. S. Ghosh and D. L. Reilly, "Credit card fraud detection with a neural-network," in Proceedings of the 27th Hawaii International Conference on System Sciences, vol. 3, pp. 621–630, Wailea, Hawaii, USA, January 1994.
44. S. Jha, M. Guillen, and J. C. Westland, "Employing transaction aggregation strategy to detect credit card fraud," Expert Systems with Applications, 2012; 39(16): 12650–12657.
45. S. Maes, K. Tuyls, B. Vanschoenwinkel, and B. Manderick, "Credit card fraud detection using Bayesian and neural net- works," in Proceedings of the 1st International NAISO Congress on Neuro Fuzzy Technologies, 1993; 261–270.
46. S. Maheshwari, J. Agrawal, and S. Sharma, ''A new approach for classifi- cation of highly imbalanced datasets using evolutionary algorithms,'' Int. J. Sci. Eng. Res., 2011; 2(7): 1–5.
47. S. Makki, ''Fraud analysis approaches in the age of big data—A review of state of the art,'' in Proc. IEEE 2nd Int. Workshops Found. Appl. Self Syst. (FASW), 2017; 243–250, doi: 10.1109/fas-w.2017.154.
48. S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, "Credit card fraud detection: a fusion approach using Dempster-Shafer theory and Bayesian learning," Information Fusion, 2009; 10(4): 354–363.
49. Statistical Science, 2002; 17(3): 235–255.
50. T. Lokman. 3.6 Million Credit Card Holders Have RM36.9 Billion Out- standing Balance. NST Online. Accessed: Dec., 2019. [Online]. Available: https://www.nst.com.my/news/nation/2017/08/270620/36-million-credit- card-holders-have-rm369-billion-outstanding-balance.
51. V. H. Barella, L. P. F. Garcia, M. P. De Souto, A. C. Lorena, and V. Zaslavsky and A. Strizhak, "Credit card fraud detection using self organizing maps," Information & Security,

2006; 18: 48–63.

52. Y. Yong, ‘‘The research of imbalanced data set of sample sampling method based on K-means cluster and genetic algorithm,’’ Energy Procedia, 2012; 17: 164–170. doi: 10.1016/j.egypro.2012.02.078.