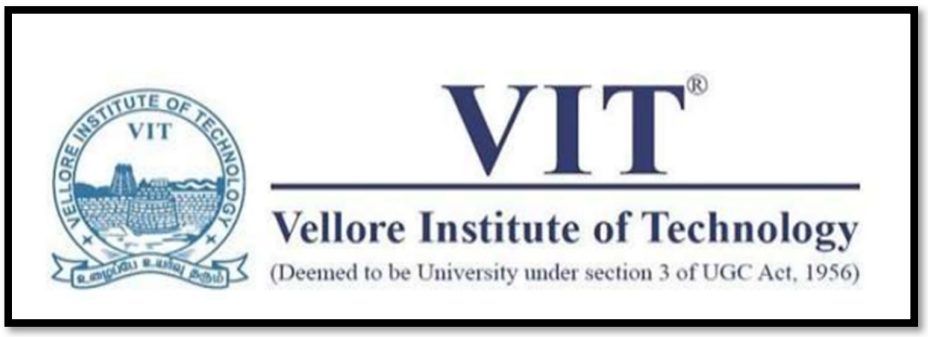

CREDIT CARD FRAUD DETECTION SYSTEM

J-Component Document

DATA MINING TECHNIQUES (ITE2006)

NILADRI MITRA (20BIT0381)

Submitted to

Prof. B. VALARMATHI, SITE

School of Information Technology and Engineering

OBJECTIVE

The main objective behind the proposed project is to detect fraud that happens when credit card is used by customer using algorithms such as random forest and ensemble method.

ABSTRACT

The most widely publicised problems in the credit card business today are those involving credit card fraud. To differentiate between the many types of credit card deception and to assess optional processes that have been used in fraud detection is the main idea. Numerous alternative steps may be put in place to lessen the frauds depending on the numerous forms of credit card frauds that financial institutions like banks or the credit card businesses deal with. Utilizing these methods and strategies has as its goal reducing credit card fraud. Some unresolved issues with the current methods cause some legitimate credit card users to be mistakenly labelled as fraudulent. Incorporating the top classification method from a group of algorithms, which is likely to indicate the level of financial fraud, is the major goal of this paper.

INTRODUCTION

A credit card is a large, practical plastic card that carries personal information, such as a signature or photo, card numbers, or magnetic stripe/chip data. It enables the person listed on it to make purchases or handle accounts in his name, for which he will occasionally be charged. The information on the card is now accessible through banks, store bar code readers, automated teller machines (TMs), and online web banking systems. They have a certain CCV number that is quite significant. Its security is reliant on the plastic card's physical security, same like how the credit card number is protected. The number of credit card swaps is increasing quickly, which has led to a significant rise in fraudulent activity. The phrase "credit card fraud" refers to a broad range of theft and misrepresentation that uses a credit card as a fictitious source of funds in a particular transaction. To address this fraud recognition issue, quantifiable procedures and many information mining computations are used in significant numbers. The bulk of artificial intelligence-based systems for credit card extortion rely on synthetic reasoning, meta-learning, and pattern matching .

Fraud detection is keeping an eye on how users spend their money in order to identify, spot, or prevent unfavourable behaviour. Fraud related to credit cards is increasing as it becomes the most popular method of payment for both normal and online purchases. Fraud detection focuses on swiftly identifying and recording fraudulent behaviours in addition to fraudulent occurrences. In today's world, credit cards are often used. Fraud is a multimillion dollar industry that keeps growing. Fraud has a tremendous financial impact on our economy globally.

For the purpose of identifying fraudulent credit card transactions, modern approaches based on data mining, machine learning, sequence alignment, fuzzy logic, genetic programming, etc. have been developed.

LITERATURE SURVEY

S.NO	TITLE OF PAPER AND YEAR	ALGORITHM USED	DATSET USED	PERFORMANCE MEASURE	SCOPE FOR FUTURE WORK
1.	Credit Card Fraud Detection using Local Outlier Factor and Isolation Forest (2019)	Local Outlier Factor and Isolation Forest	dataset is performed which is taken from Kaggle	Accuracy of 97% by Local Outlier Factor and 76% by Isolation Forest	To test using more number of dataset
2.	Cost-sensitive payment card fraud detection based on dynamic random forest and k-nearest neighbours(2018)	dynamic random forest and k-nearest neighbours	dataset is analysed and classified linearly	Accuracy= 81.69%	To test using more number of dataset
3.	Credit card fraud detection using neural network and geolocation & 2017	neural network and geolocation	location and the pattern	Accuracy = 80%	In future, proposed model has been more improve for the valid transaction.
4.	A Hybrid Approach for Credit Card Fraud Detection using Rough Set and Decision Tree Technique & 2016	Genetic Algorithm , Support Vector Machine , Decision Tree Classification	dataset is analysed and classified linearly , dataset using breadth-first approach	Accuracy = 80%	resources of the institutions can be focused on more suspicious transactions to decrease the fraud levels.
5.	The evolution of payment card fraud(2014)	Various different methods	Dataset of card holder	Accuracy-80%-90%	To test using a greater number of datasets
6.	Deep Convolution Neural Network Model for Credit-Card Fraud Detection and Alert (2021)	Deep Convolution Neural Network (DCNN), logistic regression algorithm	Dataset of A random sample of 5 million transactions over a duration of 24 hours is considered in which 6223 frauds exist.	Accuracy-90%	Fraud location, timing calculation and various other features may be incorporated with a single algorithm in future work
7.	Analysis on Credit Card Fraud Detection Methods(2011)	Various different methods	dataset is performed which is taken from Kaggle	Accuracy-80%-90%	resources of the institutions can be focused on more suspicious transactions to decrease the fraud levels.
8.	A Study on Credit Card Fraud Detection using Data Mining Techniques	Decision Tree, Neural Networks, K-Means Clusterin	Dataset of card holder	Compare various Algorithms	To test using a greater number of datasets
9.	CREDIT CARD	RANDOM	dataset is	Accuracy-90%	In future,

	FRAUD DETECTION USING RANDOM FOREST ALGORITHM	FOREST ALGORITHM	performed which is taken from Kaggle		proposed model has been more improve for the valid transaction.
10.	Comparative Analysis of Various Classification Algorithms in the Case of Fraud Detection & 2017	K Nearest Neighbor algorithm	UCI Machine Learning repository	accuracy of 91.8%	supervised learning feed forward back propagation algorithm which can be used in future analysis

S.NO	**TITLE OF PAPER AND YEAR**	**ALGORITHM USED**	**DATSET USED**	**PERFORMANCE MEASURE**	**SCOPE FOR FUTURE WORK**
1.	CREDIT CARD FRAUD DETECTION USING NEURAL NETWORKS & 2014	ARTIFICIAL NEURAL NETWORK	acquired from a data mining blog	Accuracy = 97%	increasing the number of records of the dataset for training the data in order to get accurate results and the network will be able to learn more efficiently with more number of records
2.	A Review On Credit Card Fraud Detection Using Machine Learning (2019)	Logistic Regression, Decision Tree	dataset is performed which is taken from Kaggle	Accuracy of 0.947 by Logistic Regression and 0.908 by Decision Tree	Future work will include applying semi-supervised learning methods for classification of alert in FDS
3.	A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms(2017)	Decision Tree and Random Forest Algorithms	a real-world credit card data set from a financial institution is analyzed	Accuracy-90%	This method in future proves accurate in deducting fraudulent transaction and minimizing the number of false alert.
4.	Credit Card Fraud Detection Using Hidden Markov Model & 2016	Hidden Markov Model	Microarray dataset.	Accuracy = 99.9%	Fast Clustering based feature selection algorithm can be compared with existing feature algorithm
5.	Genetic K-means Algorithm for	K-means Algorithm or	unsupervised learning	Accuracy = 99.9%	In future this model can be

	Credit Card Fraud Detection & 2015	Genetic Algorithm	approach for dataset		extended by adding various rules in rule engine to improve accuracy of the system.
7.	A Survey on Different Data Mining & Machine Learning Methods for Credit Card Fraud Detection(20 18)	Genetic algorithms Decision Tree	a real-world credit card data set from a financial institution is analyzed	majority voting method achieves good accuracy rates in detecting fraud cases in credit cards.	In future work we will develop an efficient CC fraud detection method based on existing data mining and machine learning methods.
8.	Distributed Data Mining in Credit Card Fraud Detection(2013)	Distributed Data Mining	unsupervised learning approach for dataset	accuracy of 91.8%	In future work we will develop an efficient CC fraud detection method based on existing data mining and machine learning methods.
9.	Detection of credit card fraud: State of art(2018)	Fuzzy Logic, Artificial neural network (ANN)	NSL-KDD dataset	Fraud Detection Model (AFDM), which increase the accuracy up to 25%, reduce the cost up to 85%, and decrease system response time up to 40% compared to	attempt to propose a hybrid model that is both able to handle imbalanced dataset and the realtime problem (to have a response during the financial transaction runtime) with an improved accuracy
10.	Novel Artificial Neural Networks and Logistic Approach for Detecting Credit Card Deceit & 2015	Fuzzy Darwinian, Hidden Markov Model , & GENETIC ALGORITHM	artificial intelligence, Meta learning and pattern matching.	Accuracy = 90%	This method in future proves accurate in deducting fraudulent transaction and minimizing the number of false alert.

S.NO	TITLE OF PAPER AND YEAR	ALGORITHM USED	DATSET USED	PERFORMANCE MEASURE	SCOPE FOR FUTURE WORK
1.	Credit Card Fraud Detection through Parenclitic Network Analysis(2018)	networkbased classifica tion algorithm	credit card transactions dataset.	Results confirm that features extracted from a network-based representation of data, leveraging on a recently proposed parenclitic approach	relevance is evaluated by means of a large dataset of real transactions, by comparing the yielded increase in the classification score when compared to the use of a standard ANN algorithm
2.	Machine Learning For Credit Card Fraud Detection System (2018)	Logistic Regressi on, Decision Trees, Random Forest	Dataset from kaggle and it contains a total of 2,84,808 credit card transactions	The result shows of accuracy for logistic regression, Decision tree and random forest classifier are 90.0, 94.3, 95.5 respectively	By comparing all the three methods, found that random forest classifier is better than the logistic regression and decision tree and to continue with random forest in future
3.	Credit Card Fraud Detection with Unsupervised Algorithms & 2016	PCA AND SKM ALGORITHM	it can be applied to very large data sets	Accuracy = 99.08%	In future, proposed model has been developed to satisfy the two conditions of calculation simplicity and operation transparency
4.	A novel feature engineering methodology for credit card fraud detection with a deep learning architecture(2019)	Deep Learning	Publicly available dataset were used	Accuracy = 90 to 99 %	In future work, we want to carry out further research from two aspects. The first focuses on researching the computational demand of a real-time fraud detection system. The second is to explore the application of more advanced machine learning methods and possible combinations of

					deep learning methods and traditional data mining methods in fraud detection
5.	Supervised Machine Learning Algorithms for Credit Card Fraud Detection(2017)	Supervised Machine Learning Algorithms	it can be applied to very large data sets	Accuracy = 80 to 95 %	Future researchers in this field may apply the resampling techniques to the respective datasets being used. This technique helps to reduce the imbalance ratio of datasets which in turn produces better classification results.
6.	Credit Card Fraud Detection Using Hmm And Image Click Point Authentication & 2015	Hidden Markov Model and Image Click Point Authentication technique	Image Click Point Authentication	Accuracy = 80 to 95 %	Proposed system solves drawbacks of existing system ,this type of system used even now and used in future also with better update.
7.	An Efficient Credit Card Fraud Detection Model Based on Machine Learning Methods (2020)	Logistic Regression, KNearest Neighbors	dataset includes transactions by European credit cardholders throughout the year 2013	Accuracy of 90.448 % by Logistic Regression and 94.999 % by KNearest Neighbors	can be implemented and tested on large size realtime data with different more machine learning methods
8.	Using Generative Adversarial Networks for Improving Classification Effectiveness in Credit Card Fraud Detection(2017)	Generative Adversarial Networks	Publicly available dataset were used	Accuracy = 99%	while our method can handle frauds which are similar, in essence, to malevolent transactions seen before, it can be expected to be largely ineffective in spotting frauds that are completely novel, where

					there is no information to generalize upon.
9.	Adaptive Model for Credit Card Fraud Detection (2020)	Fuzzy Associati on Rules (FAR),D eep Learning (DL)	NSL-KDD Dataset from International Journal of Advanced Research in Computer and Communication Engineering	Accuracy-85%	implementation and evaluation of the framework as a tool for credit card fraud detection
10.	An Approach to Identify Credit Card Frauds through Support Vector Machine using Kernel Trick & 2017	Support Vector Machine using Kernel Trick	regular pattern	Accuracy = 75%	aim of Kernel function is to take input vector in the original space and then to return the dot product of the two vectors in a feature space

[1] Today technology is increasing at very rapid pace, which can be used for good as well as for bad purposes. So with this growing technology e-commerce and online transactions also grown up which mostly contain transactions through credit cards. Credit cards help People to enjoy buy now and pay later for both online and offline purchases. It provides cashless shopping at every shop in all countries. As the usage of credit cards is increasing more, the chances of credit card frauds are also increasing dramatically. Credit card system is most vulnerable for frauds. These credit card frauds costs financial companies and consumers a very huge amount of money annually, fraudsters always try to find new methods and tricks to commit these illegal and outlaw actions. Online transaction fraud detection is most challenging issue for banks and financial companies. So it is much essential for banks and financial companies to have efficient fraud detection systems to reduce their losses due to these credit card fraud transactions. Various approaches have been found by many researchers till date to detect these frauds and to reduce them. Comparison of Local Outlier Factor and Isolation Factor algorithms using python and their detailed experimental results are proposed in this paper. After the analysis of the dataset we got the accuracy of 97% by Local Outlier Factor and 76% by Isolation Forest. The rate of credit card frauds is increasing at a very higher rate because attackers are becoming more and more sophisticated and are well equipped. This paper presents a paradigm to detect credit card frauds using artificial neural networks. We present a model using Neuroph IDE which provides an environment to work with neural networks. An artificial neural network is similar to a biological neuron which accepts many inputs, processes it and gives us a single output. The neural network needs to be trained continuously with a set of inputs. The network is made to learn so that when a new data is fed into it, it can properly classify it based on the learning it acquired.

The learning can be said to be accurate and efficient if it gives the expected output for the test data.

[2] Payment card fraud leads to heavy annual financial losses around the world, thus giving rise to the need for improvements to the fraud detection systems used by banks and financial institutions. In the academe, as well, payment card fraud detection has become an important research topic in recent years. With these considerations in mind, we developed a method that involves two stages of detecting fraudulent payment card transactions. The extraction of suitable transactional features is one of the key issues in constructing an effective fraud detection model. In this method, additional transaction features are derived from primary transactional data. A better understanding of cardholders' spending behaviors is created by these features. After which the first stage of detection is initiated. A cardholder's spending behaviors vary over time so that new behavior of a cardholder is closer to his/her recent behaviors. Accordingly, a new similarity measure is established on the basis of transaction time in this stage. This measure assigns greater weight to recent transactions. In the second stage, the dynamic random forest algorithm is employed for the first time in initial detection, and the minimum risk model is applied in cost-sensitive detection. We tested the proposed method on a real transactional dataset obtained from a private bank. The results showed that the recent behavior of cardholders exerts a considerable effect on decision-making regarding the evaluation of transactions as fraudulent or legitimate. The findings also indicated that using both primary and derived transactional features increases the F-measure. Finally, an average 23% increase in prevention of damage (PoD) is achieved with the proposed cost-sensitive approach.

[3] The most acknowledged payment mode is credit card for both disconnected and online mediums in today's day and age. It facilitates cashless shopping everywhere in the world. It is the most widespread and reasonable approach with regards to web based shopping, paying bills, what's more, performing other related errands. Thus danger of fraud exchanges utilizing credit card has likewise been expanding. In the Current Fraud Detection framework, false exchange is recognized after the transaction is completed. As opposed to the current system, the proposed system presents a methodology which facilitates the detection of fraudulent exchanges while they are being processed, this is achieved by means of Behaviour and Locational Analysis(Neural Logic) which considers a cardholder's way of managing money and spending pattern. A deviation from such a pattern will then lead to the system classifying it as suspicious transaction and will then be handled accordingly.

[4] To make the business accessible to a large number of customers worldwide, many companies small and big have put up their presence on the internet. Online businesses gave birth to e-commerce platforms which in turn use digital modes of transaction such as credit-card, debit card etc. This kind of digital transaction attracted millions of users to transact on the internet. Along came the risk of online credit card frauds. Hence the need to have secure payment transactions arose and many techniques based on Neural Network,

Decision Tree, Artificial Intelligence, Artificial Immune System, Fuzzy based systems, Nearest neighbor algorithm, Support Vector Machines, Genetic Algorithm were developed to detect the fraudulent online credit card transactions. This paper presents hybrid Approach for Credit Card Fraud detection using Rough Set and Decision Tree Technique which can be used in credit card fraud detection mechanisms. In this contribution we explore the possibility of using complex networks as a way of improving credit card fraud detection. Specifically, networks are used to synthesize complex features representing card transactions, relying on the recently proposed approach of parenclitic networks (Section 3). Afterwards, their relevance is evaluated by means of a large dataset of real transactions, by comparing the yielded increase in the classification score when compared to the use of a standard ANN algorithm (Section 4). We additionally show that the combined data mining/complex networks approach is able to outperform a commercial system in some specific situations.

[5] Steve Gold looks at the problem of credit and debit card fraud, the evolution of the technology, its problem with backwards compatibility, and analyses some of the more leading-edge fraud methodologies used by increasingly technology-aware cyber-criminal gangs. Credit and debit card fraud has changed significantly over the past few years. Gone are the days when fraudsters used simple skimming techniques – with pinhole cameras to record the associated PIN – at the ATM to harvest card credentials for card cloning. Today's fraudsters harvest card credentials – and allied information – on a commodity basis, using a variety of techniques.

[6] With the exponential increase in the usage of the internet, numerous organisations, including the financial industry, have operationalized online services. The massive financial losses occur as a result of the global growth in financial fraud. Henceforth, devising advanced financial fraud detection systems can actively detect the risks such as illegal transactions and irregular attacks. Over the recent years, these issues are tackled to a larger extent by means of data mining and machine learning techniques. However, in terms of unknown attack pattern identification, big data analytics and speed computation, several improvements must be performed in these techniques. The Deep Convolution Neural Network (DCNN) scheme based financial fraud detection scheme using deep learning algorithm is proposed in this paper. When large volume of data is involved, the detection accuracy can be enhanced by using this technique. The existing machine learning models, auto-encoder model and other deep learning models are compared with the proposed model to evaluate the performance by using a real-time credit card fraud dataset. Over a time duration of 45 seconds, a detection accuracy of 99% has been obtained by using the proposed model as observed in the experimental results.

[7] Due to the rise and rapid growth of E-Commerce, use of credit cards for online purchases has dramatically increased and it caused an explosion in the credit card fraud. 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. In real life, fraudulent

transactions are scattered with genuine transactions and simple pattern matching techniques are not often sufficient to detect those frauds accurately. Implementation of efficient fraud detection systems has thus become imperative for all credit card issuing banks to minimize their losses. Many modern techniques based on Artificial Intelligence, Data mining, Fuzzy logic, Machine learning, Sequence Alignment, Genetic Programming etc., has evolved in detecting various credit card fraudulent transactions. A clear understanding on all these approaches will certainly lead to an efficient credit card fraud detection system. This paper presents a survey of various techniques used in credit card fraud detection mechanisms and evaluates each methodology based on certain design criteria.

[8] The credit card system is the process which is initiates the global economy to develop significantly. Credit card providers issued millions of credit cards to their customers. They don't know whether the card holder is wrong customer, if card is issued to wrong customer that can be a very crucial factor of the financial crisis. This survey represents an organized analysis of data mining methods and its applications in credit card process. Our survey mainly focuses on data mining methods implemented especially in credit card process which helps to emphasis much larger parts. Hence, this survey must be very helpful for any credit card providers to select an appropriate solution for their problem as well as for researchers to have comprehensive of the review of literature in their area.

[9] This Project is focused on credit card fraud detection in realworld scenarios. Nowadays credit card frauds are drastically increasing in number as compared to earlier times. Criminals are using fake identity and various technologies to trap the users and get the money out of them. Therefore, it is very essential to find a solution to these types of frauds. In this proposed project we designed a model to detect the fraud activity in credit card transactions. This system can provide most of the important features required to detect illegal and illicit transactions. As technology changes constantly, it is becoming difficult to track the behavior and pattern of criminal transactions. To come up with the solution one can make use of technologies with the increase of machine learning, artificial intelligence and other relevant fields of information technology, it becomes feasible to automate this process and to save some of the intensive amounts of labour that is put into detecting credit card fraud. Initially, we will collect the credit card usage data-set by users and classify it as trained and testing dataset using a random forest algorithm and decision trees. Using this feasible algorithm, we can analyse the larger data-set and user provided current data-set. Then augment the accuracy of the result data. Proceeded with the application of processing of some of the attributes provided which can find affected fraud detection in viewing the graphical model of data visualization.

[10] Credit card payment has become very popular today. Credit card is an easiest way to pay directly through your bank account. But we all know that everything have some pros as well as some cons. In the case of credit card, fraudsters are the main intruder. These intruders can access some unauthorised transactions. It is very important to prevent your

account transaction from these intruders. In this paper we used three different classification algorithms (KNN, Neural network, C5. 0) for fraud detection.

[11] The rate of credit card frauds is increasing at a very higher rate because attackers are becoming more and more sophisticated and are well equipped. This paper presents a paradigm to detect credit card frauds using artificial neural networks. We present a model using Neuroph IDE which provides an environment to work with neural networks. An artificial neural network is similar to a biological neuron which accepts many inputs, processes it and gives us a single output. The neural network needs to be trained continuously with a set of inputs. The network is made to learn so that when a new data is fed into it, it can properly classify it based on the learning it acquired. The learning can be said to be accurate and efficient if it gives the expected output for the test data.

[12] The study showed that as size of training dataset increases the number of fraud detected by SVM are less than fraud identified by decision tree method. Here in author presented fraud detection system using a Naive Bayes K-Nearest Neighbors method. The main aim of proposed system was to improve accuracy. Naive Bayes Classifier predicts probabilities of fraud in transaction while KNN classifier predicts how near the undefined sample data is to kth training dataset. The author compared both this classifier and showed that both work differently for given dataset. Most of predictive model used for detecting fraud in credit card transaction faces the issue of concept drift. The author presented two FDS based on sliding window and ensemble learning and showed that classifier need to be trained separately using feedback and delayed samples. The outcome of the two was than aggregated to improve the alert precision in FDS. Thus the author showed that to solve the issue of concept drift, the feedback and delayed samples are to be handled separately.

[13] In this paper we mainly focus on credit card fraud detection in real world. Here the credit card fraud detection is based on fraudulent transactions. Generally credit card fraud activities can happen in both online and offline. But in today's world online fraud transaction activities are increasing day by day. So in order to find the online fraud transactions various methods have been used in existing system. In proposed system we use Random Forest Algorithm(RFA) for finding the fraudulent transactions and the accuracy of those transactions. This algorithm is based on supervised learning algorithm where it uses decision trees for classification of the dataset. After classification of dataset a confusion matrix is obtained. The performance of Random Forest Algorithm is evaluated based on the confusion matrix. The results obtained from processing the dataset gives accuracy of about 90%.

[14] Credit card is the most popular mode of payment for both online as well as offline. The use of credit cards has increased day by day and a fraudulent transaction has also increased day by day in today’s world. Credit card provides cashless shopping at every shop in the world. In Credit card fraud detection system, fraudulent transaction will be detected after transaction is done. Credit card fraud can be detected using Hidden Markov Model (HMM) during transactions. Hidden Markov Model is the tools for solve “hidden” problems. In this

paper, the sequence of operations in credit card transaction processing system using a Hidden Markov Model and show how it can be used for the detection of fraud. Using Hidden Markov Model, the fraud detection processing system is trained with the standard procedures and spending patterns of a card user.

[15] Rapid growth in electronic commerce technology has led to a tremendous increase in the use of online credit card payment mode. With the usage of credit cards, the number of frauds associated with it also increases. In order to avoid credit card frauds, proper security measures need to be taken. This work reflects an attempt to detect fraudulent credit card transactions by using k-means along with genetic algorithm. Genetic Algorithm is a powerful optimization technique. The k-means algorithm groups the credit card transactions based on the independent attribute values. But, with the increase in input size, it results in outliers. Hence to provide optimized detection of frauds, we used genetic algorithm. The significant results by proposed model are observed over simple K-means and Simple Genetic Algorithm.

[17] The rapid participation in online based transactional activities raises the fraudulent cases all over the world and causes tremendous losses to the individuals and financial industry. Although there are many criminal activities occurring in financial industry, credit card fraudulent activities are among the most prevalent and worried about by online customers. Thus, countering the fraud activities through data mining and machine learning is one of the prominent approaches introduced by scholars intending to prevent the losses caused by these illegal acts. Primarily, data mining techniques were employed to study the patterns and characteristics of suspicious and non-suspicious transactions based on normalized and anomalies data. On the other hand, machine learning (ML) techniques were employed to predict the suspicious and non-suspicious transactions automatically by using classifiers. Therefore, the combination of machine learning and data mining techniques were able to identify the genuine and non-genuine transactions by learning the patterns of the data. This paper discusses the supervised based classification using Bayesian network classifiers namely K2, Tree Augmented Naïve Bayes (TAN), and Naïve Bayes, logistics and J48 classifiers. After pre-processing the dataset using normalization and Principal Component Analysis, all the classifiers achieved more than 95.0% accuracy compared to results attained before pre-processing the dataset.

[18] Companies and institutions move parts of their business, or the entire business, towards online services providing e-commerce, information and communication services for the purpose of allowing their customers better efficiency and accessibility. Payment card fraud has become a serious problem throughout the world. Companies and institutions loose huge amounts annually due to fraud and fraudsters continuously seek new ways to commit illegal actions. In this we will try to detect fraudulent transaction through the with the genetic algorithm. Genetic algorithm are used for making the decision about the network topology, number of hidden layers, number of nodes that will be used in the design of neural network for our problem of credit card fraud detection.

[19] Credit card fraud is costing the payment card industry, literally billions of dollars annually. Financial institutions try to improve continually their fraud detection systems, but fraudsters are in same time inventing new techniques to hack systems. That said; the prevention and detection of credit card fraud become an emergency. Data mining techniques are providing great help in financial fraud detection, since dealing with the large and complex among of financial data are big challenges for financial institutions. In recent years, several studies have used machine learning and data mining techniques to face this problem. In this paper, we propose a state of the art on various techniques of credit card fraud detection. The purpose of this study is to give a review of implemented techniques for credit card fraud detection, analyse their incomes and limitless, and synthesise the finding in order to identify the techniques and methods that give the best results so far.

[20] By the rise and rapid growth of E-Commerce, use of credit cards for online purchases has more increased and it caused an explosion in the credit card fraud. The most accepted payment mode is credit card for both online as well as regular purchase, pay bills etc. So frauds associated with it are also rising. In real life, fraudulent transactions are scattered with genuine transactions and simple pattern matching techniques are not often sufficient to detect those frauds accurately. Implementation of efficient fraud detection systems has thus become imperative for all credit card issuing banks to minimize their losses. Many modern mechanisms are developed such as CHIP & PIN the mechanism do not prevent the most common fraud type such as fraudulent credit card usages over virtual POS terminals through internet or mail orders. Finally fraud detection is the essential for stop such type of frauds. In this study, classification model based on Artificial Neural Networks (ANN) and Logistic Regression (LR) are developed and applied on credit card fraud detection problem.

[21] In this paper we explore the possibility of using complex networks as a way of improving credit card fraud detection. Specifically, networks are used to synthesize complex features representing card transactions, relying on the recently proposed approach of parenclitic networks. Afterwards, their relevance is evaluated by means of a large dataset of real transactions, by comparing the yielded increase in the classification score when compared to the use of a standard ANN algorithm (Section 4). We additionally show that the combined data mining/complex networks approach is able to outperform a commercial system in some specific situations.

[22] The aim of this paper is to propose a hybrid model for credit card fraud detection, with three level of security and it is based on profile of customer and transaction. The proposition is adapted to real time transaction and to reduce the rate of false alarm. The rest of this paper is organized as follows. The section 2 contains the state of art analysis. In section 3, the framework background is described. The section 4 detailed the proposed framework. Finally, we conclude and propose our future work in Section 5.

[23] Throughout this paper, we introduce an effective credit card fraud identification system with a feedback system, centered on machine learning techniques. That feedback approach contributes to boosting the classifier's detection rate and performance. Also

analysis the performance of different classification methods including random forest, tree classifiers, artificial neural networks, supporting vector machine, Naïve Baiyes, logistic regression including gradient boosting classifier approaches, on even a highly skewed credit card fraud database. This complete research paper is divided into different sections including; introduction portion, related activities, credit card fraud obfuscation techniques for machine learning, and the obstacles. Subsequently, the implementation for machine learning techniques as well as the estimation and evaluation of different performance measurement parameters are covered and then the findings of the entire research are covered and also suggested further enhancements.

[24] Multiple financial frauds cause massive losses to companies and global financial institutions. Without the knowledge of the authenticated user, unauthorized credit and debit card transactions, credit card theft and several other such fraudulent activities are alarming the world governments, clients and banking sector [8]. The financial fraud detection systems have the ability to identify unusual attacks and unauthorized access. These fraud detection mechanisms are constantly updated by financial institutions. These issues are addressed by data mining and machine learning tools that are commonly used over the last few years [9]. Various research literatures have proposed optimal solutions by using these tools and techniques. However, the integration of big data, memory cost and computational cost may still be improved by using these techniques to meet the requirements of the growing financial sector [10]. The major challenges addressed in financial fraud detection schemes are the constantly varying fraudulent behavior, lack of fraud transaction information tracking mechanism, limitations of machine learning algorithms and other existing models and algorithms that are hard to train with the highly skewed financial fraud datasets.

[25] The study showed that as size of training dataset increases the number of fraud detected by SVM are less than fraud identified by decision tree method. Here in [6] author presented fraud detection system using a Naive Bayes K-Nearest Neighbors method. The main aim of proposed system was to improve accuracy. Naive Bayes Classifier predicts probabilities of fraud in transaction while KNN classifier predicts how near the undefined sample data is to kth training dataset. The author compared both this classifier and showed that both work differently for given dataset. Most of predictive model used for detecting fraud in credit card transaction faces the issue of concept drift. The author [7] presented two FDS based on sliding window and ensemble learning and showed that classifier need to be trained separately using feedback and delayed samples. The outcome of the two was than aggregated to improve the alert precision in FDS. Thus the author showed that to solve the issue of concept drift, the feedback and delayed samples are to be handled separately.

[26] The most acknowledged payment mode is credit card for both disconnected and online mediums in today's day and age. It facilitates cashless shopping everywhere in the world. It is the most widespread and reasonable approach with regards to web based shopping,

paying bills, what's more, performing other related errands. Thus danger of fraud exchanges utilizing credit card has likewise been expanding. In the Current Fraud Detection framework, false exchange is recognized after the transaction is completed. As opposed to the current system, the proposed system presents a methodology which facilitates the detection of fraudulent exchanges while they are being processed, this is achieved by means of Behaviour and Locational Analysis(Neural Logic) which considers a cardholder's way of managing money and spending pattern. A deviation from such a pattern will then lead to the system classifying it as suspicious transaction and will then be handled accordingly.

[27] The paper focuses on the way to detect this kind of fraud is to analyze the spending patterns on every card and to figure out any inconsistency with respect to the “usual” spending patterns. Fraud detection based on the analysis of existing purchase data of cardholder is a promising way to reduce the rate of successful credit card frauds. Since humans tend to exhibit specific behavioristic profiles, every cardholder can be represented by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent, etc. Deviation from such patterns is a potential threat to the system.

[28] To make the business accessible to a large number of customers worldwide, many companies small and big have put up their presence on the internet. Online businesses gave birth to e-commerce platforms which in turn use digital modes of transaction such as credit-card, debit card etc. This kind of digital transaction attracted millions of users to transact on the internet. Along came the risk of online credit card frauds. Hence the need to have secure payment transactions arose and many techniques based on Neural Network, Decision Tree, Artificial Intelligence, Artificial Immune System, Fuzzy based systems, Nearest neighbor algorithm, Support Vector Machines, Genetic Algorithm were developed to detect the fraudulent online credit card transactions. This paper presents hybrid Approach for Credit Card Fraud detection using Rough Set and Decision Tree Technique which can be used in credit card fraud detection mechanisms.

[29] Credit card frauds come in a wide variety of shapes and systems, like misrepresentation by using any depiction's installment card, and this is only the beginning. Likewise the reasons behind the misrepresentation of the card update. Others are meant to gain money from accounts, while others wish for nothing to get merchandise. Over time the credit card payment market has grown exponentially with the rise and universal Internet. Most businesses and industries have transformed their business into online services to provide ecommerce, ease of access, and connectivity to allow better productivity and accessibility for their customers [3]. This development is of immense value because it attributes to improved productivity and sustainability, but it still has its own limitations. A larger scope for risks emerges along with this development. One of the most significant problems in conducting business on the internet is that when making the transaction, neither the card nor the cardholder needs to be present [4]. This makes it difficult for the

retailer to test whether or not the client who is making a transaction is the genuine cardholder. This makes it conveniently possible for a fraudster to conduct a fraudulent transaction.

[30] This paper investigates the performance of logistic regression, decision tree and random forest for credit card fraud detection. Dataset of credit card transactions is collected from kaggle and it contains a total of 2,84,808 credit card transactions of a European bank data set. It considers fraud transactions as the “positive class” and genuine ones as the “negative class” .The data set is highly imbalanced, it has about 0.172% of fraud transactions and the rest are genuine transactions. The author has been done oversampling to balance the data set, which resulted in 60% of fraud transactions and 40% genuine ones. The three techniques are applied for the dataset and work is implemented in R language. The performance of the techniques is evaluated for different variables based on sensitivity, specificity, accuracy and error rate. The result shows of accuracy for logistic regression, Decision tree and random forest classifier are 90.0, 94.3, 95.5 respectively. The comparative results show that the Random forest performs better than the logistic regression and decision tree techniques.

EXISTING SYSTEMS

1) Various supervised and unsupervised learning techniques are utilised for credit card fraud detection.

2) Using Hidden Markov Models, the researchers tested the detection of credit card fraud and reported 80% accuracy.

GAP IDENTIFIED

A better algorithm should be used to increase accuracy and success rates so that users are safeguarded against fraud.

PROPOSED METHOD

KNN, decision trees, AdaBoost Classifier, Tree Random Forest, and SVM models are the methods we're utilising.

DATASET DESCRIPTION AND SIMPLE DATA

8a) Data Set Information:

This is a simulated credit card transaction dataset containing legitimate and fraud transactions. It covers credit cards of customers doing transactions with a pool for buying essentials merchants. The provided dataset comprises of data from. The dataset consists of 3075 rows and 12 columns. They were used, one for training, one for validation and one for testing.

The Dataset available on UCI Machine learning repository.

Link: UCI Machine Learning Repository: default of credit card clients Data Set

8b) Attribute Information:

This research employed a binary variable, default payment (Yes = 1, No = 0), as the response variable. This study reviewed the literature and used the following 23 variables as explanatory variables:
X1: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.
X2: Gender (1 = male; 2 = female).
X3: Education (1 = graduate school; 2 = university; 3 = high school; 4 = others).
X4: Marital status (1 = married; 2 = single; 3 = others).
X5: Age (year).
X6 - X11: History of past payment. We tracked the past monthly payment records (from April to September, 2005) as follows: X6 = the repayment status in September, 2005; X7 = the repayment status in August, 2005; . . .;X11 = the repayment status in April, 2005. The measurement scale for the repayment status is: -1 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; . . .; 8 = payment delay for eight months; 9 = payment delay for nine months and above.
X12-X17: Amount of bill statement (NT dollar). X12 = amount of bill statement in September, 2005; X13 = amount of bill statement in August, 2005; . . .; X17 = amount of bill statement in April, 2005.
X18-X23: Amount of previous payment (NT dollar). X18 = amount paid in September, 2005; X19 = amount paid in August, 2005; . . .;X23 = amount paid in April, 2005.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	Y
1	ID	LIMIT_BA	SEX	EDUCATIC	MARRIAG	AGE	PAY_(	PAY_2	PAY_3	PAY_4	PAY_5	PAY_6	BILL_AMT	BILL_AMT2	BILL_AMT	BILL_AMT	BILL_AMT	BILL_AMT	PAY_AMT	PAY_AMT	PAY_AMT	PAY_AMT	PAY_AMT	PAY_AMT	default payment
2	1	20000	2	2	1	24	2	2	-1	-1	-2	-2	3913	3102	689	0	0	0	0	689	0	0	0	0	1
3	2	120000	2	2	2	26	-1	2	0	0	0	2	2682	1725	2682	3272	3455	3261	0	1000	1000	1000	0	2000	1
4	3	90000	2	2	2	34	0	0	0	0	0	0	29239	14027	13559	14331	14948	15549	1518	1500	1000	1000	1000	5000	0
5	4	50000	2	2	1	37	0	0	0	0	0	0	46990	48233	49291	28314	28959	29547	2000	2019	1200	1100	1069	1000	0
6	5	50000	1	2	1	57	-1	0	-1	0	0	0	8617	5670	35835	20940	19146	19131	2000	36681	10000	9000	689	679	0
7	6	50000	1	1	2	37	0	0	0	0	0	0	64400	57069	57608	19394	19619	20024	2500	1815	657	1000	1000	800	0
8	7	500000	1	1	2	29	0	0	0	0	0	0	367965	412023	445007	542653	483003	473944	55000	40000	38000	20239	13750	13770	0
9	8	100000	2	2	2	23	0	-1	-1	0	0	-1	11876	380	601	221	-159	567	380	601	0	581	1687	1542	0
10	9	140000	2	3	1	28	0	0	2	0	0	0	11285	14096	12108	12211	11793	3719	3329	0	432	1000	1000	1000	0
11	10	20000	1	3	2	35	-2	-2	-2	-2	-1	-1	0	0	0	0	13007	13912	0	0	0	13007	1122	0	0
12	11	200000	2	3	2	34	0	0	2	0	0	-1	11073	9787	5535	2513	1828	3731	2306	12	50	300	3738	66	0
13	12	260000	2	1	2	51	-1	-1	-1	-1	-1	2	12261	21670	9966	8517	22287	13668	21818	9966	8583	22301	0	3640	0
14	13	630000	2	2	2	41	-1	0	-1	-1	-1	-1	12137	6500	6500	6500	6500	2870	1000	6500	6500	6500	2870	0	0
15	14	70000	1	2	2	30	1	2	2	0	0	2	65802	67369	65701	66782	36137	36894	3200	0	3000	3000	1500	0	1
16	15	250000	1	1	2	29	0	0	0	0	0	0	70887	67060	63561	59696	56875	55512	3000	3000	3000	3000	3000	3000	0
17	16	50000	2	3	3	23	1	2	0	0	0	0	50614	29173	28116	28771	29531	30211	0	1500	1100	1200	1300	1100	0
18	17	20000	1	1	2	24	0	0	2	2	2	2	15376	18010	17428	18338	17905	19104	3200	0	1500	0	1650	0	1
19	18	320000	1	1	1	49	0	0	0	-1	-1	-1	253286	246536	194663	70074	5856	195599	10358	10000	75940	20000	195599	50000	0
20	19	360000	2	1	1	49	1	-2	-2	-2	-2	-2	0	0	0	0	0	0	0	0	0	0	0	0	0
21	20	180000	2	1	2	29	1	-2	-2	-2	-2	-2	0	0	0	0	0	0	0	0	0	0	0	0	0
22	21	130000	2	3	2	39	0	0	0	0	0	-1	38358	27688	24489	20616	11802	930	3000	1537	1000	2000	930	33764	0
23	22	120000	2	2	1	39	-1	-1	-1	-1	-1	-1	316	316	316	0	632	316	316	316	0	632	316	0	1
24	23	70000	2	2	2	26	2	0	0	2	2	2	41087	42445	45020	44006	46905	46012	2007	3582	0	3601	0	1820	1
25	24	450000	2	1	1	40	-2	-2	-2	-2	-2	-2	5512	19420	1473	560	0	0	19428	1473	560	0	0	1128	1
26	25	90000	1	1	2	23	0	0	0	-1	0	0	4744	7070	0	5398	6360	8292	5757	0	5398	1200	2045	2000	0
27	26	50000	1	3	2	23	0	0	0	0	0	0	47620	41810	36023	28967	29829	30046	1973	1426	1001	1432	1062	997	0