IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DETECTION OF CREDIT CARD FRAUD USING QUERY BASED ALGORITHM

Dr.Sudheer S Marar , Mr. Pramod K, Anshidha K K
HOD ,Associate Professor, MCA student scholar,
Department of MCA, Nehru College of engineering and research centre, Thrissur, India

Abstract: - Data mining techniques have been used extensively to detect insurance fraud and financial fraud, which is a growing area of tremendous importance. Financial deception will pay special attention to spotting fraudulent credit card transactions. Credit utilisation Due to significant advancements in internet commerce, the use of credit cards has expanded dramatically. Technology. Credit cards have become the most popular method of payment for both online and offline purchases. As with any ordinary purchase, the number of incidences of fraud related with it is increasing by the day. In this Using a Hidden Markov Model to sequence operations in credit card transaction processing (HMM) and demonstrate how it might be used to detect fraud. Initially, an HMM is trained with a cardholder's regular behaviour. An incoming credit card transaction is considered fraudulent if the trained model does not accept it with a high enough probability. It also ensures that legitimate transactions are not denied. Credit cards are one of the most convenient methods of payment when shopping online. Payment is made by providing information such as the card number, security code, and expiration date of the credit card while shopping online. Every cardholder's spending method is modelled using HMM to correct the risk elements of using the credit card. Data encryption is commonly used to protect sensitive data from security threats such as "attacks on confidentiality. "In the present, Large text messages need a long time to encrypt before they can be transferred, causing a delay in subsequent information transmission. Security dangers include those that occur during the transmission of secret information across insecure communication networks.

Keyword - Computer Science Cybernetics, Credit Card Fraud, False Positive, Fraud Detection System, Thresh hold value QBA encryption, True Positive.

I. INTRODUCTION

In credit card transactions, 'fraud' refers to the unlawful and unwelcome use of an account by someone other than the account owner. To stop this misuse, necessary preventative steps should be adopted, and the behaviour of such fraudulent acts can be analysed to decrease it and defend against future occurrences. In other words, credit card fraud occurs when a person uses another person's credit card for personal gain while the card's owner and issuing authorities are ignorant.

1.Fraud Detection System

Fraud is defined as improper or criminal deception intended to achieve monetary or personal advantage or to harm another person without necessarily resulting in immediate legal consequences. The fraught hindrance and fraud detection systems are the two basic strategies for preventing frauds and losses due to fraudulent activity. Fraud prevention is a proactive strategy aimed at reducing the prevalence of fraud. Once the fraudsters have overcome the fraud deterrent measures and have begun deceptive dealings, fraud detection systems come into play. MasterCard fraud can take several forms, including direct thievery, application fraud, and counterfeit cards (where the cardboard holder absence). In on-line fraud, communications are formed remotely and solely where the card's details are required. A manual signature, a PIN or a card imprint don't seem to be needed at the time of purchase, though hindrance mechanisms like CHIP&PIN decrease the fallacious activities through straightforward thieving, counterfeit cards and NRI. Online frauds (Internet and order frauds) are still increasing in each quantity and range of transactions. There has been a growing quantity of monetary losses to MasterCard frauds because the usages of the credit cards become a lot and lot of common things.

2.Introduction of Data mining

Extraction of hidden predictive information from massive datasets is a strong new technology that has a lot of promise for helping firms

focus on the most critical data in their data warehouses. By allowing enterprises and knowledge-driven decisions, data mining technologies will be able to forecast future trends and behaviours with ease. Data mining's automated and prospective analysis go beyond the retrospective analysis provided by retrospective methods. As a result, data mining technologies can answer all of the business issues that were before difficult to answer. They explore databases for hidden patterns and uncover predicted data that experts may overlook because it falls outside of their assumptions. The majority of businesses now collect and refine vast amounts of data.

Most data mining techniques may be integrated with new products and applied on existing software and hardware platforms to increase the value of existing information resources. These data mining techniques can examine databases and provide correct responses when they are deployed on a high-performance client/server or parallel processing computer.

3.Data Mining Task

In general, the datasets utilised for knowledge discovery are sourced from existing warehouses, databases, and data markers. By combining Apriori knowledge in the form of application-specific scaling and encoding, the pre-processing stage provides an ideal representation for a data-mining technique. It outlines the many processing processes carried out on raw data in order to prepare it for further processing. Regardless of the powerful data mining method used, the resulting model will not be valid if the data is not pre-processed correctly and efficiently. As a result, it implies that any processing of raw data can prepare it for further processing. It converts data into a simple and useful manner.Because of the poor quality of the data collected, pre-processing has become critical in analysis. The majority of the time, data is collected by third parties and may not be in the format required by the application. Data errors or outliers, data with gaps or missing values, and duplicates, all of which affect the performance of the data mining process, are all possible. Pre-processing techniques aim to smooth out noise, identify outliers, fill in missing values, and fix data discrepancies.

4.Research Back Ground

Data mining is useful in detecting credit card fraud since it is frequently used to extract and expose hidden truths from massive amounts of data. Data mining is the process of finding intriguing patterns in datasets that can then be utilised to make decisions. Another definition of data mining is the process of extracting and identifying usable information from a huge database using statistical, mathematical, artificial intelligence, and machine learning approaches.

5.Problem Statement

The study's major goal is to identify how credit fraud will be used in the next years. Clustering includes the disciplines of data analysis and machine learning. The goal of the study is to pre-process data so that it may be processed further. HMM can detect fraudulent transactions in progress. The key characteristic of the HMM-based model is that it reduces False Positive (FP) transactions, which are transactions that a fraud detection system incorrectly classifies as fraud despite the fact that they are actual customers. It only detects fraud after a few transactions, thus it is not secure for initial transactions. For added security, we apply the QBA algorithm procedure.

6.Objectives

The goal of this thesis is to figure out how to spot credit card fraud. To achieve the goal, a comparative analysis of the existing algorithm and a detailed investigation were conducted to detect the fraudulent transaction.
The following will be pursued in order to attain the goal:

- It prevents card fraud during financial transactions.
- It identifies detection methods for securing transactions involving OTP through SMS and Secret Queries.
- It encrypts the information provided by the user in order to detect Master card fraud.
- It tracks the unusual card holder who has misplaced or stolen the card.
- Check for any inconsistencies in the comparison of the payment pattern to the standard pattern.

7.Motivation

The primary goal of this study is to identify financial fraud, such as credit and debit card transactions. Since (general financial frauds technology-based frauds) identifies fraudulent transactions in anomaly detection, it is necessary to assess the performance of an intrusion detection system that identifies any inconsistencies in the comparison of spending patterns with the usual pattern 12 and tests the ability of each profile to distinguish between legitimate and fraudulent usage. It protects the master card from illegal abnormal user maltreatment.

8.Implementation

With the available data set, data-oriented research has become the fundamental area of research nowadays. As a result, a variety of analyses are carried out in order to construct various computational systems (mathematical models), which have formed the foundation for various corporate operations in the public domain. As a result, data-driven decision-making has become a critical component of successful financial transactions. Many academics have used data sets from diverse sectors to develop detection systems for financial scams involving credit or debit cards.

In order to detect overlay fraud situations in major enterprises' telecommunications networks, data mining techniques are used to analyse user profiles. Fraudulent actions in mobile technologies drive research in telecommunications fraud detection. Methods for detecting fraud might be supervised or unsupervised. The user's future behaviour can then be compared to his profile to see if there is any variation from his profile that could indicate fraudulent conduct.

II. LITERATURE SURVEY

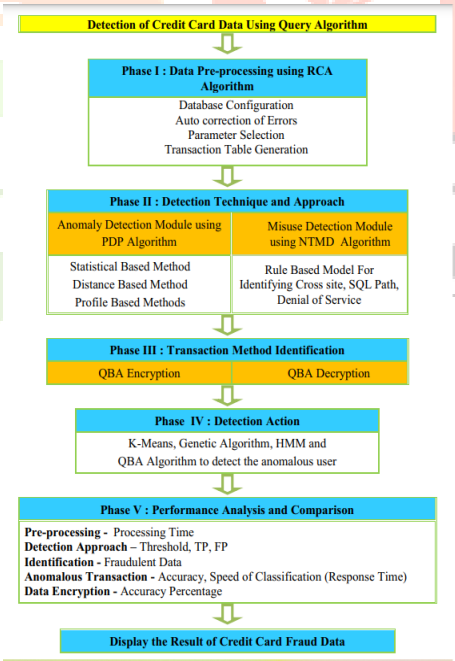
Fraud is defined as an illegal or criminal deception intended to gain financial or personal gain. It is a purposeful act committed in violation of a law, rule, or policy with the intent of obtaining unlawful financial gain.

Several literatures on anomaly or fraud detection in this domain have previously been published and are open to the public. Data mining applications, automated fraud detection, and adversarial detection are among the strategies used in this domain, according to a comprehensive survey undertaken by Clifton Phua and his colleagues. Suman, Research Scholar, GJUS&T at Hisar HCE, proposed strategies for credit card fraud detection such as Supervised and Unsupervised Learning in another work. Despite their unexpected success in some areas, these methods and algorithms failed to provide a long-term and consistent answer to fraud detection.Wen-Fang YU and Na Wang presented a comparable research domain in which they employed outlier mining, outlier detection mining, and distance sum algorithms to accurately forecast fraudulent transactions in financial transactions. Outlier mining is a type of data mining that is commonly utilised in the financial and internet industries. It is responsible for recognising objects that have become disconnected from the main system. Those transactions that aren't real. They've taken characteristics of client behaviour and calculated the value of such features.

They estimated the distance between the observed value of a characteristic and its planned value for those attributes. Unconventional methods, such as hybrid data mining and a complicated network classification algorithm, can detect anomalies. Based on a network reconstruction approach that permits producing illegal occurrences in a real card transaction data set On medium-sized datasets, representations of one instance's divergence from a reference group have proven effective. transaction made through the internet.

There have also been attempts to move forward from an entirely other perspective. In the event of a fraudulent transaction, efforts have been made to improve the alarm feedback interaction. If a fraudulent transaction is detected, the authorised system is notified, and a feedback is provided to refuse the current transaction. One of the ways that provided new light on this topic was the Artificial Genetic Algorithm, which tackled fraud from a different angle. It was successful in detecting fraudulent transactions and reducing the amount of false alarms. Despite this, there was a categorization issue with fluctuating misclassification costs.

III. METHODOLOGY



Methodology Frame Work

1. Credit Card Fraud Detection Methodology

The FDS system is utilised as a domain to anticipate credit card usage, and the mechanism for accomplishing the goal is framed. Pre-processing, detection approach, transaction method identification, and action against detection are all part of the methodology's structure.

The pre-processing will organise the data so that it may be used for future processing. The approach is identified using the Profile Detection and Network Monitoring algorithms. The real positive and false positive values are predicted using the Hidden Markov model. Finally, the illegal user is detected using a query-based technique.

a. Data Pre-Processing using RCA Algorithm:

Data pre-processing refers to any process that is conducted on data in order to prepare it for several procedures. Data pre-processing is a type of data mining that puts data into a format that cardholders can easily and effectively process. The raw data from the credit card transaction is used in the data pre-processing phase.

The Pre-processing module's main goal is to preprocess the data sets. Its goal is to obtain data sets that contain only relevant transactions and no duplicates. The data Preprocessing uses the training and test sets of the data set as inputs. This processing is done independently for both data sets in order to reduce the number of transactions required by deleting the unnecessary transactions, which are mostly single transactions.

Many attribute values are inconsistent, partial, or missing in the data set. Data cleaning and data transformation must be applied to the data to overcome this. This is accomplished using the following algorithm:

- System statistics and audit logs are used to configure the database.
- Singular value decomposition and non-negative matrix factorization are used to automatically fix errors.
- The RCA Algorithm is used to choose parameters or qualities.
- The transaction table is created following the preceding steps.

b. Detection Technique and Approach using PDP and NTMD Algorithm

The input for the Detection process is the Transaction table generated during the pre-processing step, which is done using two modules: Anomaly and Misuse Detection. Anomaly-based detection aims to recreate the data flow seen in everyday situations without the use of investigative tactics. In contrast, techniques in Misuse Detection aim to predetermine information about the pattern in the data flow to send to detective algorithms in the form of explicit signatures.

a. Transaction Method Identification using QBA Encryption

Data mining techniques are used to assess user profiles with the goal of detecting superimposed fraud situations in large-scale communications networks. The fraudulent actions in mobile technologies are driving research in telecommunications fraud detection. Unsupervised or supervised fraud detection approaches are available. Thus, the user's future behaviour can be compared to their profile in order to check for consistency with his profile's regular behaviour, which could indicate fraudulent activity.

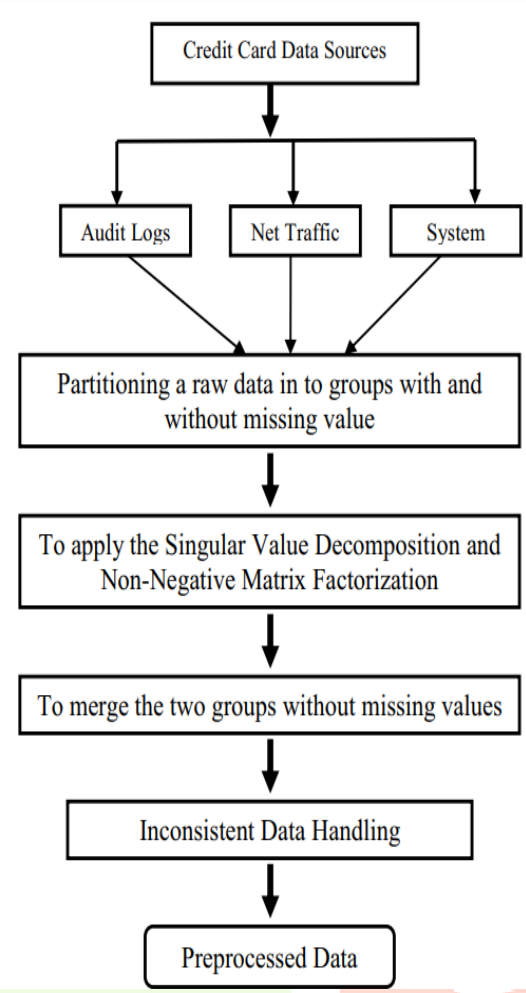
b. Detection Action

For the synthetic data of financial fraud based on credit card domain, pre-processing, detection methodologies, and procedures were used. Our goal is to evaluate real-time transaction data from a variety of institutions. The algorithm in use is intended to detect Credit Card Fraud transaction data. A novel query-based approach is introduced in light of the above factor. The query-based algorithm should be provided training.

c. Performance Analysis

For comparison, algorithms such as Classification, Clustering k-means, and Genetic algorithm were used. QBA employed one of these algorithms to train the data set and detect the erroneous transaction, and its accuracy percentage outperformed the other methods.

2. Pre-Processing of Credit Card Data



Credit Card Data Pre-processing Methodology

3. Inconsistent Data Handling

Dataset quality issues can be found in a financial transaction dataset for bank transactions, such as spelling errors, unrelated numbers entered during data entry, missing information, or other erroneous data. When several data sources must be connected, such as in web mining information systems, merged or combined data bases, and data warehouses, the demand for data cleaning becomes significantly.This is because the sources frequently provide redundant data in various forms. Consolidation of disparate data formats and the deletion of redundant information are required to offer access to correct and consistent data.

It is clear that combining the dataset of these reasons will result in inconsistencies. Inconsistencies arise, for example, from the use of different formats to represent dates (Date attributes – Date of Lone sanctioned month & year and Lone Issued date), as well as spelling errors (i.e., in the names of village and beneficiary, family type, married status, religion, business name, and so on). The usage of variance, such as for the name "Aruna Devi," and the use of a closed name or nick name, such as "Aruna" or "Devi" with the family name or spouse name "Devi Barathi," are two further grounds for contradictions. For example, statistics based on duplicated or missing data will be inaccurate or misleading ("garbage in, trash out"). Because of the large range of data that can be collected.

In order to regulate the data reduction and avoid too much inconsistent changing the properties of data, inconsistent data checking is employed after the replacement of missing values. The criteria define the amount of dimensional reduction data that can be accepted. The following methods are offered for calculating the dataset's inconsistency rate:

- When matched, two instances can be considered inconsistent if they are identical except for their class labels.
- The inconsistency count is calculated by subtracting the number of matching instances from the number of most frequently seen class labels; for example, if there are m matching examples, S1 tuples belong to class I1, S2 to class I2, and S3 to class I3, where S1 + S2 + S3 = m (if S3 is the largest among the three, the inconsistency count is n – S3);

- The inconsistency rates are obtained by adding all the counts of inconsistencies and then dividing by the total number of instances.

4. Data Transformation

The inconsistency rates are calculated by multiplying the total number of incidents by the total number of counts of discrepancies.

Step 1: Gather information Experiments are carried out on a real-world data set gathered from bank data for any city. The following is an example of the data obtained in Excel format:
Two-thirds of the data set is used for training, while the remaining data is used for testing.

a. Set of Instructions A training set is a collection of data used in information science to uncover potentially predictive associations. Artificial intelligence, machine learning, genetic programming, intelligent systems, and statistics all employ training sets. A training set serves the same purpose in all of these domains and is sometimes used in conjunction with a test set.

b. Practice Set A test set is a collection of data used in information science to determine the strength and utility of a prediction connection. Artificial intelligence, machine learning, genetic programming, and statistics all require test sets. A test set plays a similar role in all of these domains.

Step 2: Pick a training scenario. The profile size can be adjusted as needed. When the profile size is huge,raises the accuracy while also increasing the training time.

5. Data Reduction

The dataset was created using compressed data, which is substantially smaller than the original data yet retains its integrity. As a result, data mining will have a greater impact on the condensed dataset, yielding the same (or nearly identical) analysis result. To eliminate and avoid redundancy and inconsistency in the output dataset, as well as to increase the accuracy and speed of data mining, the mutually linked different data sources should be brought together.

The following are the specific procedures:
Select three fields from the card transaction log table: card number, transaction amount, and tran date to create a dataset, then summarise the data by card id and usage date, calculating monthly usage amounts for each account, and finally integrating the data records into an annual usage amount table for clustered mining. The table below contains a portion of the combined data.

6. Parameter Selection

The following strategies are used to pick the parameters that are chosen for each individual instance in particular attributes such as customer id, customer name, dob, secret query, tran id, tran amount, tran date, and time for fraudulent transactions.

Principal Component Analysis of Kernels

Kernel principal component analysis (kernel PCA) is a multivariate statistics extension of principal component analysis (PCA) that employs kernel methods. The originally linear PCA processes were performed in a reproducing kernel Hilbert space with a non-linear mapping using a kernel.

Analysis of Independent Components

By maximising the statistical independence of the estimated components, ICA finds the independent components (also known as factors, latent variables, or sources). With the two biggest definitions of independence for ICA as 61 and 62, one of the various ways to define independence and the form of the ICA algorithm can be picked. • Mutual information minimization • Non-Gaussianity maximisation The MMI family of ICA algorithms uses measurements like Kullback-Leibler Divergence and maximum entropy to minimise mutual information.

Analysis of Canonical Correlation

CCA (Canonical-Correlation Analysis) is a method for understanding cross-covariance matrices. When two vectors of random variables, X = (X1,Xn) and Y = (Y1,Ym), have correlations between them, canonical-correlation analysis will find a linear combination of the Xi and Yj that has the highest correlation. Almost all regularly used parametric significance tests can be thought of as specific examples of canonical correlation analysis, which is a broad process for examining the correlations between two sets of data.

Algorithm RCA

The train/test data is pre-processed by listing out the unique client ID. To search for the attribute "custattr1" in both data sets, out of a total of ten attributes, this corresponds to the Customer ID. Scanning the full data set for the respective transactions of each customer ID. Only if a customer ID has more than five transactions is it maintained in the preprocessed transaction lists; otherwise, it is eliminated. This approach is repeated for both the train and test sets, yielding preprocessed data sets as outputs. The train set has been reduced to 22,545 transactions after pre-processing, while the test set has been reduced to 9,425 numbers of transactions.

RCA Algorithm

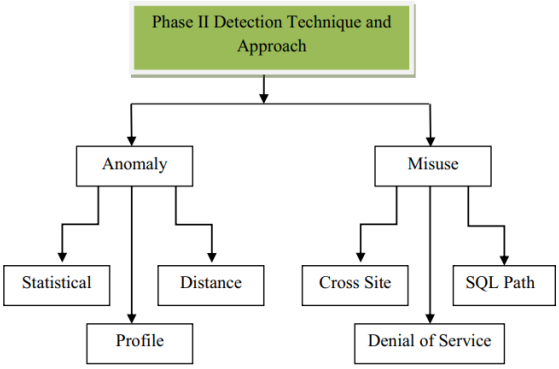
Step 1. Data Pre-processing Input: Load the train / Test data
Step 2.Output: Preprocessed list of data
Step 3. Initialize: Attribute matrix Customer Id = find (field Titles, custattr1)
Step 4. For I =1 to n If (Customer Id>5)
Step 5.Add (Customer Id Accepted Set of Users)Step
6. End for Step
7.End if

IV. DETECTION TECHNIQUE FOR TRANSACTION METHOD IDENTIFICATION

This chapter describes the Anomaly and Misuse detection strategy and approach for identifying transaction methods based on online or offline transactions, as demonstrated by the algorithms Profile Detection Processing and Network Traffic Monitor Detecting. QBA encryption and decryption employing SKC and PKC transformations identify the Transaction Method, which performs the process for continuing with valid transaction. Once the transaction has been verified as real, a false alarm for a fraudulent transaction is received.

1.Anomaly Detection Approach Using PDP

Finding patterns in data that do not conform to expectations is known as anomaly detection. In different covering areas, these non-conforming patterns are called to as Anomalies, Outliers, Discordant Observation, Exceptions, Surprises, Peculiarities, or Contaminants. Anomalies and Outlier are the two terms most typically used in the context of anomaly technique, and they are occasionally used interchangeably. Anomaly spotting is used in a range of applications, including credit card fraud detection, insurance fraud detection, and health fraud detection, usurpation detection for cyber-security, malfunction detection in safety-critical systems, and military monitoring for enemy operations. Figure explains anomaly that is created in data for a variety of causes, including malevolent conduct, such as Credit Card Fraud, cyberintrusion, terrorist action, or system failure, but all of the explanations have one thing in common: they are all fascinating to the analyst. A major component of anomalous signal detection is the "interestingness" or real-life tale relevance of odd people.



Detection Approach and Technique

HMM is used to identify the erroneous positive degree fire in the diagram above. The phone number of typical transactions is recognised as abnormal in a false positive attack. The hidden states are the different forms of purchases. Based on the transaction history, a normal new transaction is characterised as anomalous or normal. The user is categorised using the HMM depending on his disbursement profile. The number of typical transactions labelled as unusual is the false positive rate. The following pattern is used to determine the False Positive Degree Rate (FPR):

FPR= (Number of anomalous transactions/Number of normal transactions) * 100%.

2.Statistical Based Method

It measures the quantity of the variable and statistics over time to track user/meshwork behaviour. The limits of applied mathematics can be circumvented in this way. When the information second power measure is difficult to imagine inside the three-dimensional distributions, outlier identification procedures are used. Profile-based Method proficiency is comparable to rule-based technique, but visibility of traditional behaviour is built for various types of meshing traffics, users, and everyone throughout this type of method. Synonyms / hyponyms of noun device and divergence from this visibility agency to the intrusion (ordered by estimated frequency).
There are numerous scenarios in which determining the relationship between variables in a Worldwide Web environment is necessary. As a result, these dependencies are given crucial information.

3.Distance Based Technique

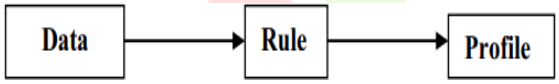
Dependency Signal Detection, Category Recognition, Category Description, and Exception/Outlier Detection are four common categories for data discovery activities. The core three task classifications correspond to patterns that apply to a large number of items or a large proportion of the data set's objects. Most data processing analysis falls into these three categories (for example, affiliation normal compartmentalization, noises clump, and conception generalisation). The fourth class in the note focuses on the share of knowledge objects, which is normally rejected as noise in machine learning and data processing. Outliers have been mentioned in machine learning and data processing, but only to the extent of accepting them regardless of the techniques.

4.Profile Based Technique

The conjunctive rule varies the client's individual behaviour, and the association and categorization rule selects a certain consumer. When it comes to customer behaviour, there are some advantages to employing rules.

A conjunctive rule is a well-studied notion utilised extensively in data processing, skilled systems, and logic programming, in addition to being intuitive and descriptive to model behaviour. Furthermore, the researchers have created an architectural plan for numerous rule discovery methods that have been published in the literature, particularly for association and classification rules, as well as a diagram to identify the profile process based on the phases.

Profile Process Based on the Phases



V. RESULT ANALISYS

The detection system predicts the data mining approaches that may be effectively employed to solve the socioeconomic problem of distributing the products of each customer purchase to the public through credit/debit transaction. The data mining algorithms created are used to forecast future transaction usage for the public in a certain location. The study is based on real data from the Fraud Detection System in several cities.

Certain conclusions have been reached regarding the current functionality of the Fraud increase the performance of the system and to optimise customer happiness as a result of the many surveys, studies, and in-depth analysis conducted in this research. The need to keep the Fraud Detection System running in the current Indian setting is a key problem, and transactions are efficiently processed using the expanding volumes of data. Data mining is a powerful method that allows data analysts to quickly and efficiently investigate big databases. The following two facts motivate the use of data mining in FDS analysis:

- When data is properly analysed, it becomes information.
- When information is effectively interpreted, it becomes knowledge.

Classifier Result

The accuracy of the classification is used to assess the performance of the classification algorithm. It's a hazy problem, and the correct answer could be dependent on the user. Traditional algorithm evaluation methods, such as calculating the space and time overhead, can be employed, but they are just secondary. The percentage of tuples placed in the proper class is commonly used to calculate classification accuracy. This ignores the possibility of a cost associated with submitting an improper assignment to the inappropriate class.

VI. CONCLUSION

The FDS System assists the government in detecting and reporting certain customer behaviours to the public. Various towns have established a model for adopting the FDS as a universal system with the purpose of eliminating financial loss caused by various fraudulent activities. The FDS's performance depends on the timely delivery of critical transaction attributes to the client. True Positive, False Positive, and Threshold Value are required for proper use of FDS. The Detective System's involvement is significant in the FDS's success in India. The clustering time and response time were considered in the methods utilised in this thesis. The technique aids in the effective clustering of data for predicting future transaction usage offline and online using the optimal method.

The research study aids the Credit Card FDS Cyber Crime section in better comprehending the rising volume of sales and predicting transactions in a timely and efficient manner. The proposed algorithm is known as a 'Query Based Algorithm,' and its major purpose is to supply decision-makers with timely and relevant transaction information.

The current research project aims to create a Query Based Algorithm that officers in the FDS department can use to quickly detect patterns and trends, establish linkages, and generate forecasts. Detection Technique and Approach were used to identify the employment of data mining techniques during the design and development process, which consists of five key phases, namely data preparation, Profile Detection Process, and Network Traffic Monitoring.

Detection of Fraud The classification of a certain client transaction is determined by calculating the mistake rate. In each phase, a performance evaluation of the techniques is proposed. An upgraded RCA algorithm is utilised to choose the parameters for processing the data, and the algorithm's efficiency is compared to KPCA, CCA, and ICA.

The algorithm is effective in terms of analysis speed, recognising common patterns, and future prediction, according to experimental data. In the current state of affairs, the proposed algorithm has potential value and can be employed as an effective algorithm by the FDS department for detecting client transactions and preventing the illicit use of Fallacious Transactions.

VII. REFERENCES

1. Network Based Database Mining System for Credit Card Fraud Detection," Department of Electrical Engineering and Computer Science, University of Siegen, pp. 220-226, 1997.
2. A. Abdelhalim and I. Traore, "Identity Application Fraud Detection using Web Mining and Rule-based Decision Tree," International Journal of Computer and Network Security (IJCNS), vol. 1, no. 1, pp. 31-44, 2009.
3. Abhinav Srivastava, Amlan Kundu, Shamik Sural and Arun K. Majumdar, "CreditCard Fraud Detection Using Hidden Markov Model" IEEE, Transactions On Dependable And Secure Computing, Vol. 5, No 1. , JanuaryMarch 2008.
4. A. M. Hormozi and S. Giles, "Data mining: A competitive weapon for banking and retail industries," Information Systems Management, pp. 62-71, 2004.
5. A. Brabazon, J. Cahill, P. Keenan and D. Walsh, "Identifying Online Credit Card Fraud using Artificial Immune," in IEEE Congress on Evolutionary Computation (CEC), Dublin, 2010.
6. W.-F. YU and N. Wang, "Research on Credit Card Fraud Detection Model Based on Distance Sum," International Joint Conference on Artificial Intelligence, pp. 353-356, 2009.
7. X. Zhu, "Semi-Supervised Learning Literature Survey," University of Wisconsin, Madison, 2008.
8. Y. Sahin and E. Duman, "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines," Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. I, pp. 1-6, 2011.
9. Z. Ghahramani, "An Introduction to Hidden Markov Models and Bayesian Networks," International Journal of Pattern Recognition and Artificial Intelligence, vol. 15, no. 1, 2001.
10. Zhou Xiaoyun, QIN Xiongpai. "Database lock table optimization based on light weight data mining" [J]. journal6, 2012, 48(8): 16-20.
11. Tsivtsivadze, Evgeni, BotondCseke, and Tom Heskes. "Kernel principal component ranking: Robust ranking on noisy data."