AN IDENTIFICATION AND DETECTION OF FRAUDULENCE IN CREDIT CARD FRAUD TRANSACTION SYSTEM USING DATA MINING TECHNIQUES

T. Kavipriya[1], N.Geetha[2]

[1]T.Kavipriya, Research Scholar, M. Phil., Computer Science, Vellalar College for Women, Erode-12.
[2]N.Geetha, Assistant Professor, Department of Computer Application, Vellalar College for Women, Tamil Nadu, India

---***---

Abstract- *Data mining concerns the extraction of implicit knowledge, data relationship or other patterns not explicitly stored in the large amount of data. Fraud mining in large amount of data is one of the powerful sources of high-level semantics. If these fraudulent transactions could be identified, detected and recognized automatically, they would be a valuable source of high-level semantics for indexing and retrieval. This thesis developed to analyze, detect and recognize the fraudulent transactions and the system is based on efficient clustering and classification methods such as apriori and support vector machine respectively. The result shows that the proposed method gives better results which helps to obtain high fraud coverage combined with a low false alarm rate than the existing Hidden Markov Model.*

Keywords: Data Mining, Credit card, Apriori, SVM

I. INTRODUCTION

Fraud detection is generally viewed as a data mining classification problem, where the objective is to correctly classify the credit card transactions as legitimate or fraudulent. The reason is the unavailability of real world data on which researchers can perform results since banks are not ready to reveal their sensitive user transaction data due to privacy reasons. Card fraud begins either with the theft of the physical card or with the compromise of data associated with the account, including the card account number or other information that would routinely and necessarily be available to a merchant during a legitimate transaction. Stolen cards can be reported quickly by cardholders, but a compromised account can be hoarded by a thief for weeks or months before any fraudulent use, making it difficult to identify the source of the compromise. This problem is challenging due to the cardholder may not discover fraudulent use until receiving a billing statement, which may be delivered infrequently. Credit card fraud has been divided into two types: On-line fraud and off line fraud. Online transaction processing (OLTP) a group of data that relieve and manage the transaction -oriented domain, typically for data entry and retrieval deal on an online database management system. It is used for financial transactions, customer relationship management (CRM) and retail sales.

OLTP payment mode is credit card amount transfer for both online and offline in today's world, it offer cashless shopping at every shop in all countries. It will be the most convenient way to do online shopping, paying bill etc. Hence, risks of fraud transaction using credit card has also been increasing. It is hard to identify fraudulent and regarding loses will be barred by issuing authorities. It is with this motivation, of the proposed system employs on fraud detection model to evaluate the performance with an anonymized dataset. The proposed model is its ability to handle class imbalance using frequent item set mining algorithm.

The objective of the Credit Card Fraud Detection Systems is to extract transactions from dataset which contains all the transactions of each user and group the legal transaction pattern and fraudulent transaction pattern of each user. Analyse whether their coming transaction is matching more with legal transaction pattern or fraudulent transaction pattern. Finally apply classification process to detect fraudulent transaction during transaction process which is made by the user. Obtain high fraud coverage combined with a low false alarm rate.

II. LITERATURE SURVEY

MaliniN. et.al [2017] described to detect the bank credit card fraud. In this paper they compare several techniques like Machine learning, Genetic Programming, fuzzy logic, sequence alignment, etc. for detecting credit card fraudulent transactions. Along with these techniques, KNN algorithm and outlier detection methods are implemented to optimize the best solution for the fraud detection problem. These approaches are proved to minimize the false alarm rates and increase the fraud detection rate.

Maria R. Lepoivre et.al [2016] discussed to develop an anti-fraud project by using a combination of two unsupervised algorithms. They using classification for generate a package. Then each package will be grouped by using clustering concept. The model has been applied to manually implemented data containing on many bank accounts. PCA, SIMPLEKMEANS unsupervised classification scheme has been applied to classify the transactions. It directly classifies the transactions with a good precision and it can detect new fraudulent behaviors.

Sumanet.al [2013] describes a survey of new techniques used in credit card fraud detection and telecommunication fraud process. Fraud detection methods based on neural network are the most popular ones. An artificial neural network consists of an interconnected group of artificial neurons. It is widely applied in classification and clustering. A improve neural networks, successfully and banks can detect fraudulent use of a card, and faster. Among the reported credit card fraud studies most have focused on using neural networks. In more practical terms neural networks are non-linear statistical data modelling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data. [5]

Bhusari V. et.al [2011] presented a credit card fraud can be detected using Hidden Markov Model during transactions. HMM model helps to obtain a high fraud reporting combined with a low false alarm rate. A HMM is a finite set and each state is linked with a probability distribution values. Transitions dataset state among these states are governed by a set of probabilities called transition probabilities. In a particular state a possible outcome or observation can be generated which is associated symbol of observation of probability distribution. Hence, HMM is a great value solution for addressing detection of fraud transaction through credit card. [1]

Benson Edwin Raj S. et.al [2011] discussed of fraud detection, two Bayesian networks to describe the behaviour of user is constructed. First, a DBN is make to model behaviour under the assumption that the user is fraudulent (F) and another model under the assumption the user is a legitimate (NF). The 'fraud net' is set up by using expert knowledge. The 'user net' is set up by using data from non-fraudulent users. In proposed method current operation user net is adapted to a specific user based on emerging data. By inserting evidence in these networks and propagating it through the network, the probability of the measurement x less than two above mentioned hypotheses is obtained. This probability value means, it gives judgments to what degree experiential user behaviour meets typical fraudulent or non-fraudulent behaviour.

Salvatore J. Stolfoet.al [2000] describes the results achieved using the JAM distributed data mining system for the real world problem of fraud detection in financial information systems. For this online process domain provide clear evidence that state-of-the-art commercial fraud detection systems can be substantially improved in stopping losses due to fraud by combining multiple models of fraudulent transaction shared among banks. Demonstrate that the traditional statistical metrics used to train and evaluate the performance of learning systems,(i.e. statistical accuracy or ROC analysis) are misleading and perhaps inappropriate for this application. [3]

III. SYSTEM DESIGN

The main aim of this thesis is to design and the software that extracts the textual information present in images.

A. Workflow of the Proposed System

The processing steps of the proposed system as shown in Fig.3.1 and the following steps are summarized as,

Step 1: Initially a series of n transactions is taken from the dataset. The input is given as a data with each row corresponds to the transaction and each column represents the attributes.

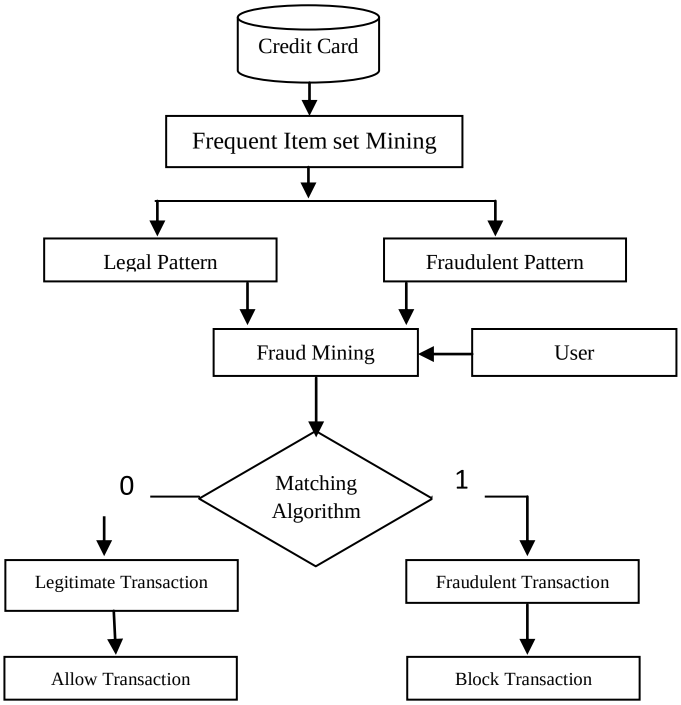

Fig 3.1 Steps of the Proposed System

Step 2: Process the transactions for finding the frequent item set in credit card transaction dataset.

Step 3: Split the amount of each transaction into different range and count the number of transactions falls into each range.

Step 4: The count is used to generate the total number of transactions done by each user.

Step 5: Form two group namely as legal and fraudulent transaction patterns for each user based on their previous transaction.

Step 6: It groups the user's transactions by considering the transactions made by each user based on the account number from which the transaction has taken place.

Step 7: The new transactions enters into the system. That new incoming transaction states similarity will be checked using the above grouping data which refers to the class information i.e., legal or fraud pattern.

Step 8: In the prediction state, the matching process is accomplished and gives the result.

Step 9: The result shows, the incoming transaction is more matched with either legal or fraudulent transaction pattern.

Step 10: If the output is legitimate transaction then that transaction will be allowed for further processing. Otherwise, the transaction should be suspected by the analyst / administrator.

B. Classification Algorithms

The following classification algorithms used in the proposed system,

- Apriori algorithm
- SVM classification algorithm

Apriori algorithm is one of the classification algorithms. This is applied to get a refined dataset. When applied along with genetic algorithm, it provides an optimal solution to the defined problem. It uses bottom-up approach and works on the basis of hash tree and BFS (breadth first search). It is a process where in frequent items in the dataset are mined using association rule mining. Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceed by identify the repeated individual items in the database and extending them to larger item sets as long as those item sets appear adequately often in the database. The recurrent item sets determined by apriori and it can be used to establish association rules which emphasize universal trends in the database: this has application in domain such as market basket analysis. [4]

An SVM method classifies data by decision the best hyper plane that separates all credit card fraud transaction data points of one class from those of the legal and illegal class. The best hyper plane for an SVM means the one with the largest margin between the two classes. Margin means the maximal width of the slab parallel to the hyper plane that has no interior credit transaction data points. The support vectors are the transaction data points that are closest to the separating hyper plane; these transaction points are on the boundary of the slab. The supervised learning model, first train transaction dataset a support vector machine, and then cross validate the classifier. Use the trained transaction machine to classify (predict) new credit data.

IV. RESULTS AND DISCUSSIONS

In this research experiments are performed using the UCI Machine Learning Repository. The credit card transaction data set is used in this research. It is downloaded from UCI Machine Learning Repository. It has 600 instances and 23 attributes. The attributes are 1 - amount of the given credit card, 2 - gender, 3 - Education, 4 – Marital status, 5 – Age, 6 – 11 – History of past payment (from April to September 2005), 12-17 – Amount of bill statement, 18-23 – Amount of previous payment. In order to evaluate the performance of the proposed method, these 600 instances and 23 attributes have been tested on the system.

The results for the experiments of different user transactions are presented in Table 4.1, and the corresponding values for Precision and Recall rates, F-Score and Accuracy are listed.

Algorithms	Credit Card Transaction Dataset		Measures (%)			
	No of Instances	No of Attributes	Precision	Recall	F-measure	Accuracy
HMM	600	23-Including class Label	61.5	75.4	67.7	68.2
SVM	600	23 - Including class Label	65.5	79.8	71.8	72.3

Table 4.1Comparisons of Average Performance measures

The proposed method, compared with the existing method and the average performance measures of comparisons are shown in graphically Fig 4.1

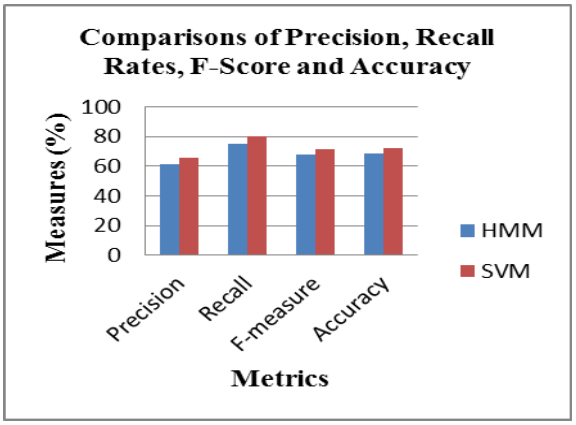

Fig4.1Column Chart showing the Comparisons of Precision and Recall rates, F-Score and Accuracy with two Algorithms

The proposed method is compared with existing method and the results show the ability to detect fraudulent transaction clearly than the existing method. The accuracy for the proposed method is shown in the Fig 4.2.

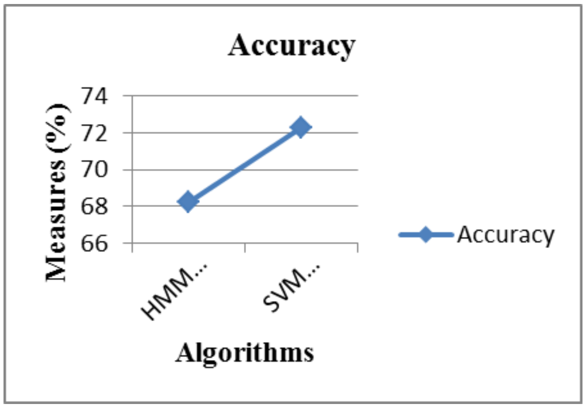

Fig 4.2 Comparison of Accuracy in line graph

Hence, from the results in all cases, it is shown that proposed method is found to be better than the proposed method for various types of transactions by providing better precision, recall, F-score and accuracy rates. The proposed method have been tested on various types of transactions with similar formatting and obtained encouraging results and experimental results show that proposed method outperforms the existing method. The results indicate that the method using efficient classification algorithms has the efficiency to discriminate between legal and fraudulent transactions of each user and detect the fraudulent transactions.

V. CONCLUSION

This thesis has discussed the credit card fraud detection system. The proposed method has been extensively tested on different types of transactions. The results were promising, almost all the fraudulent transactions could be detected successfully and the proposed method have been compared with the existing method and the result shows that the proposed method performs better than existing methods. In this research fraudulent transactions have been detected and recognized which illustrates the robustness of the proposed system. This proposed method enables the transaction at various types and improves the classification process, which can significantly improve the detection performance. The proposed system gives the following advantages are,

- Frequent item set mining is used apriori and association rule learning technique for grouping legal and fraud transaction pattern efficiently from the database which contains large item set.
- The matching process is used SVM classification prediction model technique clearly and accurately detects and identifies the fraudulent transactions and also improves the recognition performance.
- Obtain a high fraud coverage combined with a low false alarm rate. That is, it gives less number of false positives compared with existing method.

On the whole, the system is successful in achieving its objectives and can be utilized for automatic extraction, detection and recognition of credit card fraudulent transaction.

REFERENCES

[1] V. Bhusari S. Patil, “Study of Hidden Markov Model in Credit Card Fraudulent Detection” ,International Journal of Computer Applications (0975 – 8887) Volume 20– No.5, April 2011

[2] Priya Ravindra Shimpi, Prof. Vijayalaxmi Kadroli Angrish, “Survey on Credit Card Fraud Detection Techniques”, International Journal Of Engineering And Computer Science ISSN: 2319-7242

[3] Salvatore J. Stolfo, Wei Fan, WenkeLee, “Cost-based Modeling for Fraud and Intrusion Detection Results from the JAM Project”, In Proceedings of the ACM SIGMOD Conference on Management of Data, pages 207–216, 2014.

[4] Delamaire. L. Abdou, HAH and Pointon. J,”Credit card f raud and detection techniques”, Banks and Bank Systems, Volume 4, Issue 2, 2009

[5] Suman, Nutan, “Review Paper on Credit Card Fraud Detection”, International Journal of Computer Trends and Technology (IJCTT) – volume 4 Issue 7–July 2013.

[6] Renu, Suman, “Analysis on Credit Card Fraud Detection Methods”, International Journal of Computer Trends and Technology (IJCTT) – volume 8 number 1 – Feb 2014.

[7] Sushmito Ghosh and Douglas L. Reilly, “Credit Card Fraud Detection with a Neural-Network” Proc. IEEE First Int. Conf. on Neural Networks, 2014.

[8] Deepak Pawar, SwapnilRabse, Sameer Paradkar, NainaKaushi, “Detection of Fraud in Online Credit card Transactions” ,International Journal of Technical Research and Applications e-ISSN: 2320-8163.

[9] Mohamed Hegazy1, Ahmed Madian2, 3, Mohamed Ragaie, “Enhanced Fraud Miner: Credit Card Fraud Detection using Clustering Data Mining Techniques”, Egyptian Computer Science Journal (ISSN: 1110 – 2586) Volume 40 – Issue 03, September 2016