Fraud Detection in banking using Novel strategy

Ritesh Patil[1], Suraj Bhoyar[2], Abhijeet Salunke[3], Sumeet Bhingardive[4], Geeta Atkar[5]
[1,2,3,4]UG Student, Dept. of Computer Engineering, GHRCEM, Pune, Maharashtra, India
[5]Professor, Dept. of Computer Engineering, GHRCEM, Pune, Maharashtra, India

Abstract— Now a day's online payment gaining popularity because of easy and convenience use of ecommerce. It became very easy mode of payment. People choose online payment and e-shopping; because of time convenience, transport convenience, etc. As the result of huge amount of e-commerce use, there is a vast increment in credit card fraud also. Machine Learning has been successfully applied to finance databases to automate analysis of huge volumes of complex data. Machine Learning has also played a salient role in the detection of credit card fraud in online transactions. Fraud detection in credit card is a big problem, it becomes challenging due to two major reasons–first, the profiles of normal and fraudulent behaviors change frequently and secondly due to reason that credit card fraud data sets are highly skewed.

Keywords- Data Analysis, Fraud in credit card, naïve bayes, Machine Learning, Security.

I. INTRODUCTION

Credit card fraud is a growing concern with far reaching consequences in the government, corporate organizations, finance industry, In Today's world high dependency on internet technology has enjoyed increased credit card transactions but credit card fraud had also accelerated as online and offline transaction. As credit card transactions become a widespread mode of payment, focus has been given to recent computational methodologies to handle the credit card fraud problem. There are many fraud detection solutions and software which prevent frauds in businesses such as credit card, retail, e-commerce, insurance, and industries. Machine Learning is one notable and popular methods used in solving credit fraud detection problem. It is impossible to be sheer certain about the true intention and rightfulness behind an application or transaction. In reality, to seek out possible evidences of fraud from the available data using mathematical algorithms is the best effective option. Fraud detection in credit card is the truly the process of identifying those transactions that are fraudulent into two classes of legit class and fraud class transactions, several techniques are designed and implemented to solve to credit card fraud detection such as genetic algorithm, artificial neural network frequent item set mining, migrating birds optimization algorithm, comparative analysis of decision tree and random forest is carried out. Credit card fraud detection is a very popular but also a difficult problem to solve. Firstly, due to issue of having only a limited amount of data, credit card makes it challenging to match a pattern for dataset. Secondly, there can be many entries in dataset with truncations of fraudsters which also will fit a pattern of legitimate behavior. Also the problem has many constraints.

Firstly, data sets are not easily accessible for public and the results of researches are often hidden and censored, making the results inaccessible and due to this it is challenging to benchmarking for the models built. Datasets in previous researches with real data in the literature is nowhere mentioned. Secondly, the improvement of methods is more difficult by the fact that the security concern imposes a limitation to exchange of ideas and methods in fraud detection, and especially in credit card fraud detection. Lastly, the data sets are continuously evolving and changing making the profiles of normal and fraudulent behaviors always different that is the legit transaction in the past may be a fraud in present or vice versa. This paper evaluates two advanced machine learning, Decision tree and random forests and then a collative comparison is made to evaluate that which model performed best. Credit card transaction datasets are rarely available, highly imbalanced and skewed. Optimal feature (variables) selection for the models, suitable metric is most important part of mining to evaluate performance of techniques on skewed credit card fraud data. A number of challenges are associated with credit card detection, namely fraudulent behavior profile is dynamic, that is fraudulent transactions tend to look like legitimate ones, Credit card fraud detection performance is greatly affected by type of sampling approach used, selection of variables and detection technique used.

II. LITERATURE REVIEW

In this section, we briefly review the related work on credit card fraud system and their different techniques.
In [1] this document proposes a new comparative measure of the comparison rules that reasonably represents the profits and losses due to fraud detection. A cost-sensitive method based on the minimum Bayes risk is presented using the proposed cost measure. Improvements of up to 23% are obtained by comparing this method and other latest-generation algorithms. The data set for this document is based on the real-life transactional data of a large European company and personal data in the data is kept confidential. The accuracy of an algorithm is about 50%. The importance of this work was to find an algorithm and reduce the

cost measurement. The result was 23% and the algorithm they found was the minimal risk of Bayes.

In [2] Several modern techniques based on sequence alignment, machine learning, artificial intelligence, genetic programming, data mining, etc. They have been developed and are still being developed to detect fraudulent credit card transactions. A solid and clear understanding of all these approaches is needed, which will undoubtedly lead to an efficient credit card fraud detection system. This document shows a survey of different techniques used in credit card fraud detection mechanisms and the evaluation of each methodology based on certain design criteria. An analysis of credit card fraud detection methods was performed. The survey in this document was based solely on detecting the efficiency and transparency of each method. The importance of this document was to conduct a survey to compare different credit card fraud detection algorithms to find the most appropriate algorithm to solve the problem.

In [3] A comparison was made between models based on artificial intelligence together with a general description of the fraud detection system developed in this document, such as the naive Bayesian classifier and the Bayesian network model, the clustering model. And finally, conclusions are drawn on the results of the model evaluation tests. The number of legal truncation was determined to be greater than or equal to 0.65, ie its accuracy was 65% using the Bayesian network. The importance of this document is to compare the models based on artificial intelligence together with a general description of the developed system and to establish the accuracy of each model together with the recommendation to create the best model.

In [4] Nutan and Suman on review on credit card fraud detection they have supported the theory of what is credit card fraud, types of fraud like telecommunication, bankruptcy fraud etc. and how to detect it, in addition to it they have explained numerous algorithms and methods on how to detect fraud using Glass Algorithm, Bayesian, networks, Hidden Markova model, Decision Tree and 4 more. They have explained in detail about each algorithm and how this algorithm works along with mathematical explanation. Types of machine learning along with classifications has been studied. Pros and cons of each method is listed.

III. PROPOSED APPROACH

In this system evidences from current as well as past Behaviour are combined. A fraud detection system is proposed that includes rule based filter, Dempster Shafer adder, transaction history database and Bayesian learner. In rule base the suspicion level of each incoming transaction is determined. Dumpster Shafer is used to combine multiple such evidences and an initial belief is computed. Based on this belief the transactions are classified as normal, abnormal or suspicious. The incoming transactions are initially handled by the rule base using probability values. After this the values are combined using Dumpster Shafer Adder. If the transaction is declared as fraudulent then it is handled by the card holder. If suppose the transaction is suspicious then it is fed in the suspicious table. The score of transaction is updated in the database with the help of machine learning classification. This architecture is flexible such that new kinds of fraud can be handled easily. With the help of learner the system can dynamically adapt to the changing needs.

System Diagram:

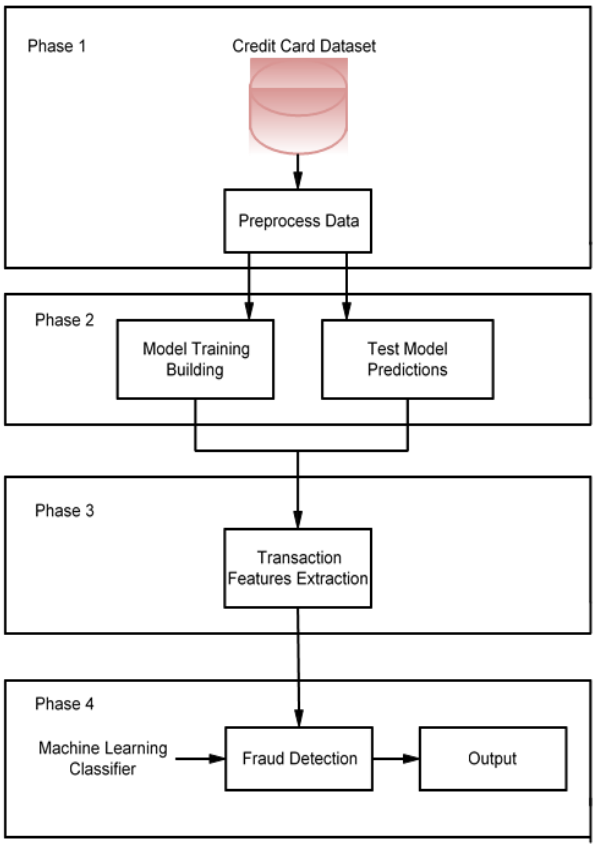

Fig 1. System Architecture

A. Algorithm

1. Naive Bayes

Steps:

1. Given training dataset D which consists of documents belonging to different class say Class A and Class B
2. Calculate the prior probability of class A=number of objects of class A/total number of objects
   Calculate the prior probability of class B=number of objects of class B/total number of objects
3. Find NI, the total no of frequency of each class
   Na=the total no of frequency of class A
   Nb=the total no of frequency of class B

4. Find conditional probability of keyword occurrence given a class:
   P (value 1/Class A) =count/ni (A)
   P (value 1/Class B) =count/ni (B)
   P (value 2/Class A) =count/ni (A)
   P (value 2/Class B) =count/ni (B)
   ……..
   ……..
   ……..
   P (value n/Class B) =count/ni (B)
5. Avoid zero frequency problems by applying uniform distribution
6. Classify Document C based on the probability p(C/W)
   a. Find P (A/W) =P (A)*P (value 1/Class A)* P (value 2/Class A)……. P(value n /Class A)
   b. Find P (B/W) =P (B)*P (value 1/Class B)* P (value 2/Class B)……. P(value n /Class B)
7. Assign document to class that has higher probability.

IV. Result and Discussion

a. Comparison Graph:

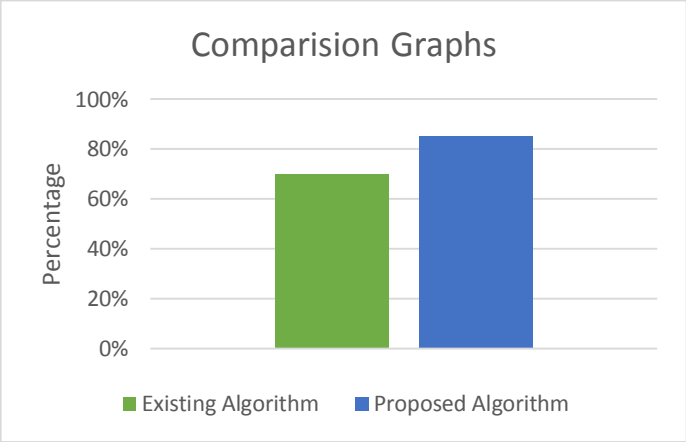

Fig2. Graph

b. Comparison Table:

Sr.No	Existing Algorithm	Proposed Algorithm
1	65%	86%

Table 1.comparative result

V. Conclusion

Credit card fraud detection is a fascinating domain. From this survey, we analyze that machine learning is the best compared to forecasting and classification. Machine learning techniques are mainly preferred in fraud detection, due to their high accuracy and detection rate. Even so, researchers find it difficult to achieve greater accuracy and detection speed. In addition, organizations are interested in finding ways to reduce costs and increase profits; you can find and select the method of previous studies.

VI. References

[1] Jain R., Gour B., Dubey S., A hybrid approach for credit card fraud detection using rough set and decision tree technique, International Journal of Computer Applications 139(10) (2016).
[2] Dermala N., Agrawal A.N., Credit card fraud detection using SVM and Reduction of false alarms, International Journal of Innovations in Engineering and Technology (IJIET) 7(2) (2016).
[3] Phua C., Lee V., Smith, Gayler K.R., A comprehensive survey of data mining-based fraud detection research. arXiv preprint arXiv:1009.6119 (2010).
[4] Bahnsen A.C., Stojanovic A., Aouada D., Ottersten B., Cost sensitive credit card fraud detection using Bayes minimum risk. 12th International Conference on Machine Learning and Applications (ICMLA) (2013), 333-338.
[5] Carneiro E.M., Dias L.A.V., Da Cunha A.M., Mialaret L.F.S., Cluster analysis and artificial neural networks: A case study in credit card fraud detection, 12th International Conference on Information Technology-New Generations (2015), 122-126.
[6] Hafiz K.T., Aghili S., Zavarsky P., The use of predictive analytics technology to detect credit card fraud in Canada, 11th Iberian Conference on Information Systems and Technologies (CISTI) (2016), 1-6.
[7] Sonepat H.C.E., Bansal M., Survey Paper on Credit Card Fraud Detection, International Journal of Advanced Research in Computer Engineering & Technology 3(3) (2014).VarrePerantalu K., BhargavKiran, Credit card Fraud Detection using Predictive Modeling (2014).
[8] Stolfo S., Fan D.W., Lee W., Prodromidis A., Chan P., Credit card fraud detection using meta-learning: Issues and initial results, AAAI-97 Workshop on Fraud Detection and Risk Management (1997).
[9] Maes S., Tuyls K., Vanschoenwinkel B., Manderick, B., Credit card fraud detection using Bayesian and neural networks, International Journal of Pure and Applied Mathematics Special Issue 836Proceedings of the 1st international naiso congress on neuro fuzzy technologies (2002), 261-270.
[10] Chan P.K., Stolfo S.J., Toward Scalable Learning with Non-Uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection, In KDD (1998), 164-168.