An Improved Machine Learning Method for Efficient Fraud Detection from CC Database

Mr. Vipul Patil[1] Dr. Umesh Kumar Lilhore[2]
[1]M.Tech Scholar [2]Head of Department
[1,2]Department of Computer Science & Engineering
[1,2]NIIST Bhopal, India

Abstract— Credit card fraud detection methods are widely used for CC fraud detections. These techniques are based on data mining, artificial intelligence and machine learning methods. In this research work we are presenting, an improved machine learning hybrid method for efficient fraud detection from cc database. Proposed CC Fraud detection method uses the quality of improved support vector machine algorithm with k-NN and genetic algorithm. A SVM method is widely used for data classification due to its quality of results and Genetic method is used to select the best attributes set by applying best evolutionary method. Once data set is reduced by GA, SVM method is apply to classify the resulting patterns. Existing SVM method is improved by modification in Gaussian kernel parameters to improve the mapping between fraud class and normal class. Improvements are made to the hybrid with the aid of the use of a correlation degree between attributes as a health degree to update the weaker participants within the populace with newly lengthy-established chromosomes. This injects more range and will growth the overall fitness of the population. Existing k-NN method is used for best feature selection with GA. Existing SVM method and proposed ISVMGA are implemented over MATLAB Simulator and various performance measuring parameters are calculated such as accuracy, precision, sensitivity, specificity, MCC and BCR.

Key words: SVM, GA, Improved SVM, Credit Card, Fraud Detection, Machine Learning Method

I. INTRODUCTION

A credit card provides online as well offline both types of financial services to customers. In every day financial transactions credit score rating card transactions the procurement of products and services assist on-line transactions or card swiping procurements [5]. This results in growth in online transactions the usage of credit and debit playing cards evolving to a global of to be had expenditure. Frauds concerned in the credit score card (CSC) segment have brought on immoderate damage to the customers and the agency provider and is stated to be even worse in coming days [1]. This paper is presenting an efficient CC fraud detection method based in improved SVM, GA and k-NN. This paper is organised in following chapters include introduction, machine learning and cc fraud detection methods, proposed method and simulation results, comparisons and finally conclusions and future work.

II. MACHINE LEARNING & CC FRAUD DETECTION

Machine learning is an application of artificial intelligence (AI) that gives systems the functionality to robotically examine and decorate from revel in without being explicitly programmed. Machine learning specializes in the improvement of laptop packages which could get proper to get entry to information and use it examine for themselves [4].

Fraud will be delineated as "a dishonest action to create false statements so as to achieve cash or take pleasure in a private or from an organization". It's additionally referred to as "scams" or, a lot of elegantly, as "economic offenses". By any definition, it's a criminal offense, which within the e-commerce network atmosphere will value businesses billions of bucks a year. Many studies show that fraud prices billions of bucks worldwide per annum, and nearly everyone is suffering from fraud directly or indirectly.

III. EXISTING METHODS & PROBLEMS

Credit card fraud may be a growing downside within the MasterCard trade. Within America alone, losses from all kinds of MasterCard fraud area unit projected to exceed $950 million, representing the tenth increase in fraud losses in bulk amount. Although tiny in comparison to MasterCard losses because of charge-offs of seriously delinquent accounts (charge-offs accounted for the US $21.5 billion of losses in the year 2015).

Reference	Key Methods	Benefits
[1]	Machine learning methods, naïve Bayes; decision tree; logistic regression.	K-NN shows significant performance for all metrics evaluated except for accuracy in the 10:90 data distribution.
[2]	Ada-Boost and Majority Voting	The best MCC score is 0.823, achieved using majority voting. The majority voting method has yielded The best MCC score of 0.942 for 30% noise added to the dataset.
[3]	KNN and Outlier Detection	KNN method to determine the anomaly of the target instance.
[14]	Logistic Regression	Provides a probability formula for classification.
[9]	Hidden Markov Model.	Reduced the False Positive (FP) transactions
[16]	Decision Tree	D tree handles non-linear data as well as linear
[12]	k-Nearest Neighbour Algorithm	Not required any prediction model

[18]	SVM (Support vector machin)	It efficiently classified dataset with accuracy
[21]	Artificial Neural Network	Capable of detecting the fraudulent activity at the time of the transaction.
[8]	Genetic Algorithm	Select best chromosomes

Table 1:

Fraud represents Associate in fraud increasing proportion of total charge volume, indicating that it's growing quicker than the MasterCard business itself. From the year 2008 through 2015, the scale of the fraud downside grew from eight basis points to over twenty [11]. Though MasterCard fraud takes several forms, there is a unit many principal classes. Credit card fraud detection methods are widely used for CC fraud detections. These techniques are based on data mining, artificial intelligence, and machine learning methods.

Existing methods have several challenges which are needed to resolve such as:

- Poor accuracy %
- Higher detection time
- Poor TPR, FPR, MCC, FDR, and NPV

IV. PROPOSED METHODS

In this research work researchers are presenting, an improved machine learning hybrid method for efficient fraud detection from cc database. Proposed CC Fraud detection method uses the quality of improved support vector machine algorithm with k-NN and genetic algorithm.

An SVM method is widely used for data classification due to its quality of results and Genetic method is used to select the best attributes set by applying best evolutionary method. Once the data set is reduced by GA, SVM method is applied to classify the resulting patterns. Existing SVM with the k-NN method is improved by modification in Gaussian kernel parameters to improve the mapping between fraud class and normal class.

Improvements are made to the hybrid with the aid of the use of a correlation degree between attributes as a health degree to update the weaker participants within the populace with newly lengthy-established chromosomes. This injects more range and will grow the overall fitness of the population. Existing k-NN method is used for best feature selection with GA. Existing SVM with the k-NN method and proposed ISVMGA are implemented over MATLAB Simulator and various performance measuring parameters are calculated such as accuracy, precision, sensitivity, specificity, MCC, and BCR.

A. Algorithm for Proposed ISVMGA

Proposed CC Fraud detection method uses the quality of improved support vector machine algorithm with the kNN- and genetic algorithm.

1) Step-1 Load data set
2) Step-2 Apply data pre-processing (Cleaning and re-categorized)
   2.1) User_data = Search (Attributes_titles, User_Attributes)
   2.2) While (i !=n) where initially i=1
   {
   2.3) IF (User_data > 2) //Set the threshold
   {
   2.4) Add (User_data as a approved data set)
   } // End if
   } // End while
3) Step-3 // Training & Testing
   3.1) Set initial variable
   W= [2] and C=8
   3.2) Now calculate h
   Where h= C / W [i] ---------- 3.2.1
   3.3) X1 = Binary to Decimal (1 to h)
   X2 = Binary te Decimal (h+1 to C)
   K (X, Xi) = p (X) p (Xi) K (X,Xi)
   n p
   WCSS = ∑ ∑ (Xij - Xkj) pow 2 -------
   i=1 j=1
   Where (WCSS=Sum of square with in the cluster)
   Xij = is the object
   n = number of object for i (1 to n)
   p = number of object for j (1 to p)
   Xkj = it is the mean value j of all the elements in group k
   3.4) While (there are no change in mean)
   {
   3.5) For (j=1 to K)
   {
   3.6) Now Compute value (aj – bj)
   3.7) Cj = new mean (C1, C2.....Ck)
   3.8) Print the value of C1, C2
   } // End of While
   } // End of for loop
   3.9) Fraud (min (Class_1, Class_2));
4) Step 4 End

V. EXPERIMENTAL RESULTS

Proposed ISVMGA method and existing SVM with k-NN method both are implemented in MATLAB simulator.

For simulation results validation German credit card database is used. Description of the German credit dataset-

1) Title: German Credit data
2) Source Information-
   Dataset1-https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/german/
   Dataset2-https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/autralian
3) Number of Instances: 100000
4) Two datasets are provided. The original dataset, in the form provided by Prof. Hofmann, contains categorical/symbolic attributes and is in the file "german.data".

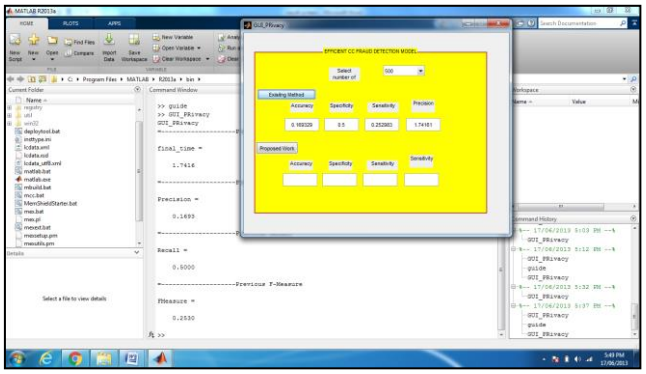
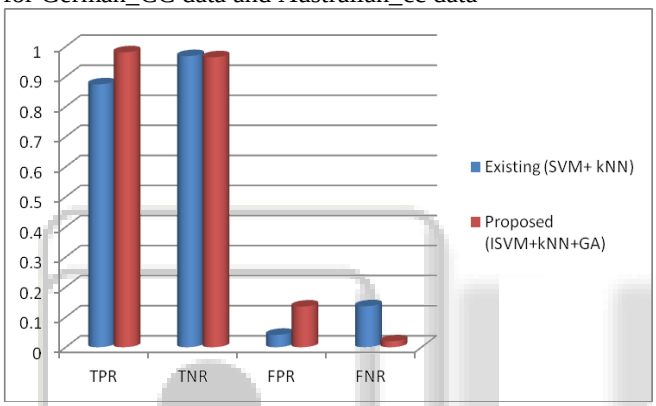


Fig. 5: Simulation Screen

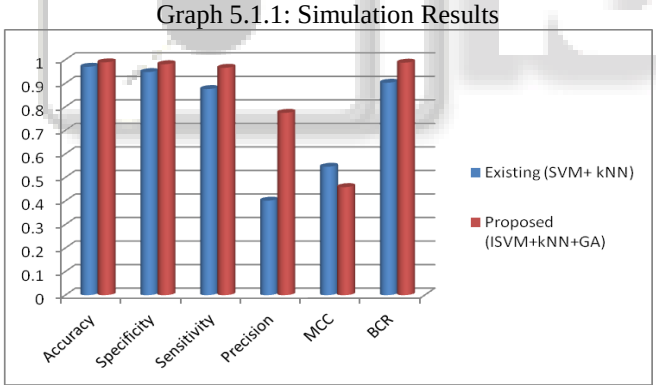
A. Results

Following resulting parameters are calculated for both of the methods

Dataset selection 60 % Training and 40 % for testing for German_CC data and Australian_cc data



Graph 5.1.1: Simulation Results



Graph 5.1.2: Simulation Results

The above results shows that proposed method (ISVM+kNN+GA) performs better over existing method (SVM+kNN).

VI. CONCLUSIONS & FUTURE WORKS

In this research work we are presenting, an improved machine learning hybrid method for efficient fraud detection from cc database. Existing SVM method and proposed ISVMGA are implemented over MATLAB Simulator and various performance measuring parameters are calculated such as accuracy, precision, sensitivity, specificity, MCC, and BCR. An experimental result clearly shows that proposed ISVMGA performs outstanding over existing SVM method.

In future work, we can apply proposed method for real-time data in the real-time environment. Explore continuous updating and adaption of subsystems and combiner. Also extend adaptive classifier to finite mixture model (more flexible), approximate logistic regression and RBF networks. Can apply more realistically handle the delayed fraud label.

REFERENCES

[1] John O. Awoyemi, Adebayo O. Adetunmbi , Samuel A. Oluwadare, "Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis", IEEE International Conference CCI 2017, pp 978-981

[2] Kuldeep Randhawa, Chu Kiong Loo, Manjeevan see," Credit Card Fraud Detection Using AdaBoost and Majority Voting", IEEE Access VOLUME 6, 2018. Pp14277-14283.

[3] N.Malini, Dr.M.Pushpa, "Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection", 3rd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEEICB17, IEEE, 978-984.

[4] B.Pushpalatha, C.Willson Joseph, "Credit Card Fraud Detection Based on the Transaction by Using Data mining Techniques", International Journal of Innovative Research in Computer and Communication Engineering , Vol. 5, Issue 2, February 2017, pp 1785-1794.

[5] You Dai∗, Jin Yan∗, Xiaoxin Tang∗, Han Zhao† and Minyi Guo∗, "Online Credit Card Fraud Detection: A Hybrid Framework with Big Data Technologies", 2016 IEEE TrustCom/BigDataSE/ISPA, pp 1644-1653.

[6] Nuno Carneiroa, Gonc¸alo Figueiraa,———————————, Miguel Costa, "A data mining based system for credit-card fraud detection in e-tail", Decision Suppoty System , Sep 2016, pp 918-925.

[7] D. Viji1, S. Kothbul Zeenath Banu, "An Improved Credit Card Fraud Detection Using K-Means Clustering Algorithm", International Journal of Engineering Science Invention (IJESI) , pp 59-65.

[8] Shivangi Sharma,Puneet Mittal, "An Approach to Detect Credit Card Frauds using Attribute Selection and Ensemble Techniques", International Journal of Computer Applications (0975 – 8887), Volume 180 – No.21, February 2018, pp 1-8.

[9] S. K. Saravanan1, Dr. G. N. K. Suresh Babu, "Literature Study –Data Mining Techniques on Detecting Fradulent Activities in Credit Card",International Journal of Emerging Research in Management &Technology ISSN: 2278-9359, (Volume-6, Issue-10), pp 60-70.

[10] Ishan Sohony,Rameshwar Pratap,"Ensemble Learning for Credit Card Fraud Detection",CoDS-COMAD '18, January 11–13, 2018, Goa, India, pp217-223.

[11] TAO GUO, GUI-YANG LI, "NEURAL DATA MINING FOR CREDIT CARD FRAUD DETECTION",Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 12-15 July 2008, pp3630-3641.

[12] Ogwueleka, F. N., (2011). Data Mining Application in Credit Card Fraud Detection System, Journal of Engineering Science and Technology, Vol. 6, No. 3, pp. 311-322

[13] Bahnsen, A. C., Stojanovic, A., Aouada, D., & Ottersten, B. (2013). Cost sensitive credit card fraud detection using Bayes minimum risk. In Machine Learning and Applications (ICMLA), 2013 12th International Conference, IEEE, Vol. 1, pp. 333-338.
[14] Y. Sahin, S. Bulkan, and E. Duman, ``A cost-sensitive decision tree approach for fraud detection," Expert Syst. Appl., vol. 40, no. 15, pp. 5916-5923, 2013.
[15] RamaKalyani, K. and UmaDevi, D., (2012). Fraud Detection of Credit Card Payment System by Genetic Algorithm, International Journal of Scientific & Engineering Research, Vol. 3, Issue 7, pp. 1 – 6.
[16] V. Bhusari, and S. Patil, "Study of Hidden V. Bhusari, and 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.
[17] E. Duman, A. Buyukkaya, and I. Elikucuk, ``A novel and successful credit card fraud detection system implemented in a Turkish bank," in Proc. IEEE 13th Int. Conf. Data Mining Workshops (ICDMW), Dec. 2013, pp. 162-171. Khyati Chaudhary, Jyoti Yadav, Bhawna Mallick, "A review of Fraud Detection Techniques: Credit Card", International Journal of Computer Applications (0975-8887) Volume 45– No.1, May 2012.
[18] M. Syeda, Y.-Q. Zhang, and Y. Pan, "Parallel granular neural networks for fast credit card fraud detection," in Fuzzy Systems, 2002. FUZZIEEE' 0-. Proceedings of the 2002 IEEE International Conference on, vol. 1, 2002, pp. 572–577.
[19] X. Wu, X. Zhu, G. Q. Wu, and W. Ding, "Data mining with big data," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1,pp. 97–107, Jan 2014.

