A NOVEL APPROACH PREDICTIVE MODELLING FOR CREDIT CARD FRAUD ANALYSIS

D.SEKHAR BABU[1], M.BHANU PRADEEP[2], M.JYOTHI[3], N.SHANKAR[4], P.HARI PRIYANKA[5].

[1] Assistant Professor, CSE,Chalapathi Institute of Technology,Guntur, India

[2] UG Student,CSE,Chalapathi Institute of Technology,Guntur, India

[3] UG Student,CSE,Chalapathi Institute of Technology,Guntur, India

[4] UG Student,CSE,Chalapathi Institute of Technology,Guntur, India

[5] UG Student,CSE,Chalapathi Institute of Technology,Guntur, India

ABSTRACT: In this project author proposed that fraud detection is a critical problem affecting large financial companies that have increased due to the growth in credit card transactions. This project presents detection of frauds in credit card transactions, Classification process of random forest algorithm to analyze data set and user current dataset. Finally optimize the accuracy of the result data. The performance of the techniques is evaluated based on accuracy, sensitivity, and specificity, and precision. Then processing of some of the attributes provided identifies the fraud detection and provides the graphical model visualization. The performance of the techniques is evaluated based on accuracy, sensitivity, and specificity, and precision This data set present transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The data set is highly unbalanced, the positive class(frauds) Account for 0.172% of all transactions.

1. INTRODUCTION

Credit card fraud is a major problem that involves payment card like credit card as illegal source of funds in transactions. Fraud is an illegal way to obtain goods and funds. The goal of such illegal transaction might be to get products without paying or gain an unauthorized fund from an account. Identifying such fraud is a troublesome and may risk the business and business organizations. In the real world FDS, investigator are not able to check all transactions.Here the Fraud Detection System monitors all the approved transactions and alerts the most suspicious one. Investigator verifies these alerts and provides FDS with feedback if the transaction was authorized or fraudulent. Verifying all the alerts everyday is a time consuming and costly process. Hence investigator is able to verify only few alerts each day.

The rest of the transactions remain unchecked until customer identifies them and report them as a fraud. Also the techniques used for fraud and the cardholder spending behavior changes over time.

This change in credit card transaction is called as concept drift. Hence most of the time it is difficult to identify the credit card fraud. Machine Learning is considered as one of the most successful technique for fraud identification. It uses classification and regression approach for recognizing fraud in credit card. The machine learning algorithms are divided into two types, supervised and unsupervised learning algorithm. Supervised learning algorithm uses labeled transactions for training the classifier whereas unsupervised learning algorithm uses peer group analysis that groups customers

according to their profile and identifies fraud based on customers spending behavior.

Many learning algorithm have been presented for fraud detection in credit card which includes , Logistic Regression , decision tree, Naive Bayes , Support Vector Machines , K-Nearest Neighbors and Random Forest. This project examines the performance of above algorithms based on their ability to classify whether the transaction was authorized or fraudulent and then compares them. The comparison is made using performance measure accuracy, specificity and precision. The result showed that Random Forest algorithm showed better accuracy and precision than other techniques.

2. LITERATURE REVIEW

The author has proposed a paper where they have first explained the proper performance measures which is used for fraud identification. The authors have structured a novel learning technique that can solve concept drift, verification latency, and class imbalance issues. The paper also showed effect of above issues in true credit card transactions.

A realistic Modeling and a Novel Learning Strategy presented two types of classifier using random forests which are used to train the behavior features of transactions. The authors have compared the two random forests and have analyzed their performance on fraud identification in credit card.

presented a FDS for credit card using Artificial Neural Network and Logistic Regression. The system used to monitor each transaction separately using classifier and then classifier would generate score for each transaction and label this transaction as legal or illegal transaction. A decision tree method was proposed

The method decreased overall misclassification costs and selected splitting property at each node. The author also compared the decision tree method for fraud identification with other models and proved that this approach performs well using performance measure like accuracy and genuine positive rate. developed a FDS for credit card transaction using support vector machines and decision tree. This study built seven alternative models that were created using support vector machines and decision tree. The author also compared this classifiers performance using performance measure accuracy. The study also showed that as size of training dataset increases the number of fraud detected by SVM are less than fraud identified by decision tree method.

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. 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

3. EXISTING SYSTEM

Credit card frauds are easy targets. Without any risks, a significant amount can be withdrawn without the owner's knowledge, in a short period. Fraudsters always try to make every fraudulent transaction legitimate, which makes fraud detection very challenging and difficult task to detect.

4. PROPOSED SYSTEM

Users can unfamiliarity is a very difficult problem in real-world when are called concept drift problems. Concept drift can be said as a variable which changes over time and in unforeseen ways. These variables cause a high imbalance in data. The main aim of our research is to overcome the problem of Concept drift to implement on real-world scenario.

1. The results obtained by the logistic regression algorithm is best compared to any other algorithms.

2.The accuracy obtained was almost equal to cent present which proves using of logistic algorithm gives best results.

3. The plots that were plotted according to the proper data that is processed during the implementation.

5.IMPLEMENTATION

Preprocessing of Data

Following are the Preprocessing steps that have been carried out

Importing Data

Importing Data set in CSV format file.

Checking the Missing Values in Data set
Balanced Data Set

Here it can be observed that the dataset is highly in balanced, and thus for accurate ML predictions and training, a balanced dataset has to be created.

Feature Scaling

Train data is fitted to a suitable classifier upon feature extraction, then once the classifier is trained enough then we predict the results of the test data using the classifier, then compare the original value to the value returned by the classifier.

Modeling

Here We are applied Various Machine learning algorithms applied. Such as

KNN
Logistic Regression
Decision Tree
Random Forest
Navie Bayes, SVM

Here the compare of different classifiers are shown among which the best classifier with highest accuracy percent is the chosen. Some factors such as f1-score, recall, and precision.etc. also accounts for consideration of the classifiers.

Visual Representation

Our final results are plotted as charts which contain different fields such as Genuine, Fraud in analysis. Thus it is chosen ml models

6. SCREEN SHORT

7. CONCLUSION

A new method for identifying fraud in credit card transactions has been proposed. The strategy was tested using a real-world dataset. The results demonstrate the possibility of fusing. many simple classification algorithms to generate a reliable solution. The need for a very low false positive rate as well as the This is a fairly busy site with a lot of valid transactions. a challenging issue The outcomes demonstrate the capabilities of the approach proposed to save at least 30% of

the money fraudulent credit card activities have put you at risk.

REFERENCES

[1] Jalinus, N., Nabawi, R. A., & Mardin, A. (2017). The Seven Steps of Project based Learning Model to Enhance Productive Competences of Vocational Students. In 1st International Conference on Technology and Vocational Teacher (ICTVT 2017). Atlantis Press. Advances in Social Science, Education and Humanities research (Vol.102, pp. 251-256).

[2] Andrea Dal Pozzolo, Giacomo Boracchi, Olivier Caelen, Cesare Alippi and Gianluca Botempi, ‖Credit card Fraud Detection : A realistic Modeling and a Novel Learning Strategy‖, IEEE Trans. on Neural Network and Learning system,vol.29,No.8, August 2018.

[3] Shiyang Xuan,Guanjun Liu,Zhenchuan Li,Lutao Zheng,Shuo Wang, Jiang,‖Random Forest for credit card fraud detection‖,Int.conf.on Networking, Sensing and control,2018.

[4] Y. Sahin , and Duman,E.,(2011) ―Detecting credit card fraud by ANN and logistic regression.‖ In Innovations in Intelligent Systems and Applications(INISTA),2011 international Symposium on (pp.315-319).IEEE

[5] 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

[6] Sahin Y. and Duman E. (2011),‖Detecting Credit Crad Fraud by Decision Trees and Support Vector Machines‖, International Multi-Conference Of Engineers and Computer Scientists(IMECS 2011),Mar 16-18,Hong Kong,Vol.1,pp.1-6

[7] Sai Kiran, Jypti Guru, Rishabh Kumar, Naveen Kumar, Deepak Katariya, Credit card fraud detection using Naïve Bayes model based and KNN classifier‖, Int. Journal of Adv. Research , Ideas and Innovations in Technology,vol.4,2018.

[8] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, ―Credit card fraud detection and concept-drift adaptation with delayed supervised information, in Proc. Int. Joint Conf. Neural Netw., 2015,pp. 1–8.

[9] A. C. Bahnsen, D. Aouada, and B. Ottersten, ―Example-dependent cost-sensitive decision trees,‖ Expert Syst. Appl., vol. 42, no. 19,pp. 6609–6619, 2015

[10] A. Dal Pozzolo, O. Caelen, and G. Bontempi, ―When is undersampling effective in unbalanced classification tasks?‖ in Machine Learning and Knowledge Discovery in Databases. Cambridge, U.K.: Springer, 2015

[11] N. Mahmoudi and E. Duman, ―Detecting credit card fraud by modified fisher discriminant analysis,‖ Expert Syst. Appl., vol. 42, no. 5,pp. 2510–2516, 2015