



A COMPARATIVE STUDY ON CREDIT CARD FRAUD DETECTION

Arpit Jain
Department of CS
IMS Engineering
College, Dr. APJ
Abdul Kalam
Technical
University,
Ghaziabad, Uttar
Pradesh, 201009,
India

Narendra Kr Sharma
Department of CS
IMS Engineering
College, Dr. APJ
Abdul Kalam
Technical
University,
Ghaziabad, Uttar
Pradesh, 201009,
India

Anshul Aggarwal
Department of CS
IMS Engineering
College, Dr. APJ
Abdul Kalam
Technical
University,
Ghaziabad, Uttar
Pradesh, 201009,
India

Love Aggarwal
Department of CS
IMS Engineering
College, Dr. APJ
Abdul Kalam
Technical
University,
Ghaziabad, Uttar
Pradesh, 201009,
India

Abstract - Credit card fraud is the term used to specify the transactions which are done without the knowledge of the account holder in order to cause monetary loss. Banks used various techniques to catch these fraud transactions but every year it happens and causes huge loss to the people and the bank. So there need to be a techniques which could use all the available data of the transaction and uses it to classify whether a transaction is fraud or not based on previous activity of the user. In this paper we've introduced some ideas of frauds associated with the credit cards or the other on-line card. We have tried to find a solution so that fraud get caught even before transaction takes place. In this paper we have applied various supervised algorithms on the dataset and the result founded are shown.

Keywords- svm, knn, ann, random forest, regression, decision tree

I. INTRODUCTION

Credit card fraud detection is becoming a topic of research because advanced techniques such as ml, big data are coming into play. Also this plays a big role in banking sector and other organizations where transactions are a major part of their process because it would help them to minimize unnecessary losses.

Generally people in bank generally detects anomaly when an unusual amount is debited or whenever signature or other credential does not match but if we replace it with a model which can measure the variations of the various features associated with the

transactions then the efficiency of catching the fraud would be increased.

The models studied in this paper are as follows:

1. KNN

Knn algorithms works on simple technique of classifying an input's label based on the neighbors which are closest and what their label is. KNN method is popular as it has simple implementation and works incredible well in practice based on simply the nature of the neighbors. But KNN have some limitation which affect the learning of result. The main problem with the KNN is that it is a lazy learner So, these algorithm may affect the accuracy of the result and overall efficiency of program.

2. Logistic regression

This model is used generally when we deal with the binary output i.e.0 or 1. It differs from the linear regression as compared to its straight curve we get a curve in the shape of S in this. It is used in the tasks where probability is concerned. Sigmoid function is used to get the output in this algorithm.

3. Random forest

Random forests or decision forests are the algorithm in which we create several decision trees by taking in consideration various features of the trees on which we could split the tree nodes. Then after the trees are created we do a majority polling in which we try to determine which trees are giving maximum accuracy scores. Then these trees are used further for calculations.



4. Support Vector Machine

Support vector machines are supervised learning algorithms. This algorithm solves the linear as well as the non-linear problems. It simply creates a line or a hyperplane which act as a separator which separates the data into various categories. The points which are closer to hyperplane or separator are support vectors points and we use them to make better classifiers. Also removing these points will affect the position of the hyperplane. The main advantage of using it that it is capable of handling data which are not linearly separable by mapping data accordingly.

5. Decision tree

In Decision Trees, we create a tree by dividing the nodes based on different features. Whatever needs to be predicted is found by following the path along the edges based on features.

6. Artificial neural network

It is a computational model which is based on structure of biological neuron. The flow of information changes its structure and we try to find a mathematical relation or formula between the target and data by adjusting structure. Various weights and bias value is adjusted in order to achieve the required relationship to record and recall a pattern or relationship.

II. LITERATURE REVIEW

Aihua Shen and 2 others [1] have done a study on classification models such as logistic regression, decision tree on the credit card dataset.

Olawale Adepoju, Hemaint Jaiman and 2 other [2] have performed study on this and concluded that svm could be used as a technique to correctly identify the problem. W.Xu and Y.Liu[3] also have done the research on using svm as an algorithm and gotten favourable result as it was found capable in determining whether the transaction is fraud or not by creating appropriate support vectors and performing prediction as such.

N.Malini and M.Pushpa [5] has in their paper performed analysis based on outlier detection and Knn algorithm is used by optimizing it to minimize the false alarm rates and also to increase the fraud detection rate.

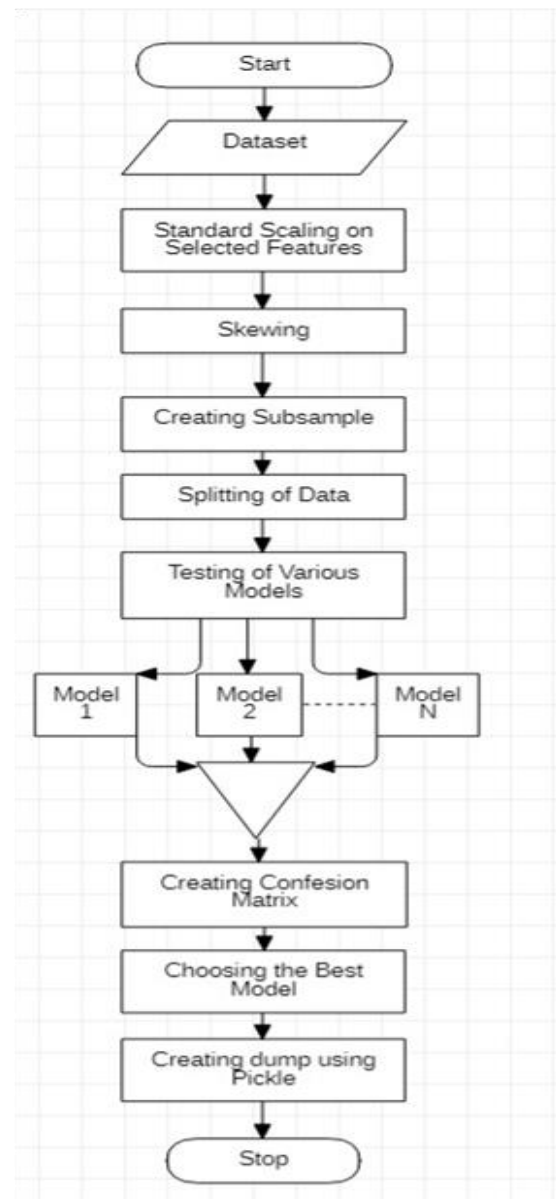
Hayat Nadim and 3 others [6] have tested various techniques on the data and concluded that the data is heavily rightly skewed and they analysed the use of oversampling and unersampling on the given dataset. Rimpal R. Popat and Jayesh Chaudhary [7]

analysed various ml as well as the use of fuzzy logic in the detection of fraud transactions.

Reshma Dayma and Krishna Modi [10] have worked on rule based mining and fuzzy clustering approach in their paper. Y. Sahin and E. Duman [12] have also compared the performance of logistic regression and the artificial neural network on the problem

Dejan Varmedja and 4 others [14] have used the smote technique for oversampling as they observed that the data was highly imbalanced. Moreover they also used multi layer perceptron to see the ability to use it on this kind of problems.

III. PROPOSED WORK FLOWCHART





IV. RESULTS

The performance of an algorithms could be evaluated on the basis of accuracy, precision and recall scores.

The results found after testing various techniques and tuning it to achieve maximum accuracy are as follows:

Technique name	Accuracy	Precision	Recall
Logistic Regression	0.9393	0.94	0.95
Svm	0.9510	0.95	0.95
Knn	0.9391	0.93	0.94
Random Forest	0.9527	0.96	0.95
Decision tree	0.9189	0.92	0.91
Ann	0.9391	0.92	0.93

As we could observe from the above results that the random forest technique is giving the maximum score of accuracy and precision then it could be considered a technique that could be employed in the prediction of the fraud transactions and since all these algorithms take less time to train and predict results then these are feasible to use. Knn and svm are excessive to train and therefore should be given less preference when creating a model. Moreover ann also gives good results but will take slightly more time to train but a model made with combining one or two models with ann or other structure would give better results.

Since all the rows of the transactions are not available and only the pca are available; this really is a problem in making a powerful technique but due to security and privacy purpose this data could not be released for the research.

Various algorithms performs differently in different environment and no single effective technique is there that we could say that this will work perfectly on this classification task. Moreover the parameters on which the performance of algorithms are somewhat limited mathematically but are not sometimes feasible enough when real life scenario

and exceptions get involved. This creates a loophole which the fraudsters often take advantage of.

V. CONCLUSIONS

In this paper, we conducted a study for credit card fraud detection by observing the behaviors of ml models on a Mlb-usb dataset after its suitable preprocessing and by tuning the parameters , the performance varies dramatically but sampling is done to improve the detection.

We have performed a comparative study to find out which technique would be used for maximum efficiency and the found results are presented. Here, random forest is found to be performing better than the others in these conditions.

This technique could be used as an added layer of security of the transaction mechanisms and is a fast and efficient way to increase the protection.

VI. REFERENCES

1. A. Shen, R. Tong and Y. Deng, "Application of Classification Models on Credit Card Fraud Detection," 10.1109/ICSSSM.2007.4280163.
2. O. Adepoju, J. Wosowei, S. lawte and H. Jaiman, "Comparative Evaluation of Credit Card Fraud Detection Using Machine Learning Techniques10.1109/GCAT47503.2019.8978372.
3. W. Xu and Y. Liu, "An Optimized SVM Model for Detection of Fraudulent Online Credit Card Transactions," , 10.1109/ICMeCG.2012.39.
4. W. Xu and Y. Liu, "An Optimized SVM Model for Detection of Fraudulent Online Credit Card Transactions," 10.1109/ICMeCG.2012.39.
5. N. Malini and M. Pushpa, "Analysis on credit card fraud identification techniques based on KNN and outlier detection",10.1109/AEEICB.2017.7972424.
6. A. H. Nadim, I. M. Sayem, A. Mutsuddy and M. S. Chowdhury, "Analysis of Machine Learning Techniques for Credit Card Fraud Detection," 10.1109/iCMLDE49015.2019.00019.
7. R. R. Popat and J. Chaudhary, "A Survey on Credit Card Fraud Detection Using Machine Learning," 10.1109/ICOEI.2018.8553963.
8. M. S. Kumar, V. Soundarya, S. Kavitha, E. S. Keerthika and E. Aswini, "Credit Card Fraud



Detection Using Random Forest Algorithm,"
10.1109/ICCCT2.2019.8824930.

9. S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang and C. Jiang, "Random forest for credit card fraud detection," 10.1109/ICNSC.2018.8361343.

10. K. Modi and R. Dayma, "Review on fraud detection methods in credit card transactions," 10.1109/I2C2.2017.8321781.

11. I. SADGALI, N. SAEL and F. BENABBOU, "Fraud detection in credit card transaction using machine learning techniques," 10.1109/ICSSD47982.2019.9002674.

12. Y. Sahin and E. Duman, "Detecting credit card fraud by ANN and logistic regression," 10.1109/INISTA.2011.5946108.

13. J. V. V. Sriram Sasank, G. R. Sahith, K. Abhinav and M. Belwal, "Credit Card Fraud Detection Using Various Classification and Sampling Techniques: A Comparative Study," 10.1109/CCES45898.2019.9002289.

14. D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic and A. Anderla, "Credit Card Fraud Detection - Machine Learning methods," 10.1109/INFOTEH.2019.8717766.