International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

A Learning Based Approach for Credit Card Fraud Detection

Siddamma Wadi[1], Dr. R N Yadawad[2]

[1]Student, Dept. of Computer Science, SDM College of Engineering and Technology, Dharwad, Karnataka, India
[2]Assistant Professor, Dept. of Computer Science, SDM College of Engineering and Technology, Dharwad, Karnataka, India

---

Abstract - *In the recent years, Finance fraud is a serious problem with far consequences in the financial industry. Data mining is widely used and applied to finance databases to automate analysis of huge volumes of complex data. Fraud detection in credit card is also a data mining problem. It becomes challenging due to two major reasons–first, the profiles of normal and fraudulent behavior change frequently and secondly due to reason that credit card fraud data sets are highly skewed. Here we investigate and check the performance of Naïve Bayes and KNN on highly skewed credit card fraud data. Credit card transactions data is collected from European cardholders consists of 284,786 transactions. These machine learning techniques are applied on the raw and pre-processed data. The performances of the techniques are evaluated based on accuracy, sensitivity, specificity and precision. The results indicates the optimal accuracy.*

Key Words: **Credit Card Fraud, Credir Card Fraud Detection, Naive BAyes, KNN.**

1. INTRODUCTION

Financial fraud is a growing serious concern with far consequences in the government, corporate organizations and finance industry. In Today's digital era, we are highly dependent on internet technologym and 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 and universally accepted mode of payment, focus has been given to recent computational methodologies to mitigate the credit card fraud problem. Fraud is as aged as humankind itself and can take carious forms and pose a problem. Moreover, the rise of new technologies makes way for fraudulent activities [2]. The use of credit card has been widespread in our society and credit card fraud is growing. These fraudulent activities are causing financial losses and affecting banks and individuals. The actions against fraud can be categorized as fraud prevention and fraud detection [6].

The design of fraud detection algorithms is the key for reducing losses caused every year due to fraudulent transactions. There are many fraud detection automated approaches which detects and avoids frauds in businesses such as credit card, retail, e-commerce, insurance, and industries. Data mining technique is one of the most commonly used and popular methods in solving credit fraud detection problem. It is impossible to predict the true intention and rightfulness behind an application or transaction. In reality, to gather possible evidences of fraud from the available data using mathematical algorithms is the best effective option. Fraud detection in credit card is a task 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 comparative analysis of KNN and Naïve Bayes are carried out.

From the experiments the result that has been concluded is that Naïve Bayes shows accuracy of 92% and KNN shows accuracy of 83%. The results obtained thus conclude that Naïve bayes forest shows the most precise and high accuracy of 92% in problem of credit card fraud detection with dataset provided by ULB machine learning.

The Paper is structured as follows: Section 1. of the paper provides a basic introduction about the Credit Card Fraud problems and Detection. Section 2. consists literture survey. Section 3 describes our proposed methodology. Section 4 gives our experimental results Finally, Section 5 concludes the paper.

2. LITERATURE SURVEY

A detailed research case study involving credit card fraud detection, where data normalization is applied before Cluster Analysis. The results obtained from the use of Cluster Analysis and Artificial Neural Networks on fraud detection has shown that by clustering attributes neuronal inputs can be minimized [1].

A new comparison measure that reasonably depicts the gains and losses due to fraud detection is proposed [2]. A cost sensitive method which is based on Bayes minimum risk is introduced using the proposed cost measure. Improvements up to 23% is obtained when this method and other state of art algorithms are compared.

Various techniques based on Sequence Alignment, Machine learning, Artificial Intelligence, Genetic Programming, Data mining etc. has been evolved since a last decade and is still evolving to detect fraudulent transactions in credit card [3]. In [4], A comparison has been done between models based on artificial intelligence along with general description of the developed fraud detection system are given in this paper such as the Naive Bayesian Classifier and the clustering model.

3. PROPOSED METHODOLOGY

Credit card fraud detection hinge on analysis of recorded transactions. Here we perform analysis on skewed data. Transaction data has number of attributes like time, amount and class. Automated analysis are needed as manual analysis would be time consuming and not at all an easy task as it involves huge number of transactions to be checked upon. Our goal is to implement machine learning models in order to classify, to the highest possible degree of accuracy, credit card fraud from a dataset gathered in Europe cardholders. After initial data exploration, we implement KNN and Naïve bayes. Some challenges we come across from the start were the huge imbalance in the dataset: frauds only account for 0.172% of fraud transactions. In this case, it is worst to have false negatives than false positives in our predictions because false negatives mean that someone gets away with credit card fraud. False positives, on the other hand, merely cause a complication and possible hassle when a cardholder must verify that they did, in fact, complete said transaction (and not a thief).

The dataset contains 31 numerical variables. The dataset obtained are normalized due to financial confidentiality. The 'time' feature specifies the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'amount' is the transaction amount. Feature 'class' is a response binary variable and shows value 1 in case of fraudulent activity, else 0. In a huge transaction of 284807, only 492 account for fraudulent transactions. The cardholder identifier is unavailable and hence each transactions are independent.

4. EXPERIMENTAL RESULTS

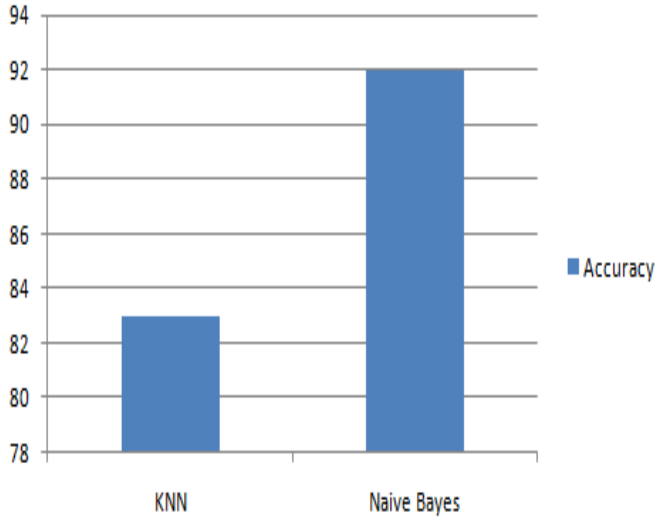
The dataset is obtained from kaggle website. The entire dataset here contains 284807 transactions. Each row contains 31 features and is normalized due to confidentiality issues. Following confusion matrix shows the number of correct and wrong predictions made using Naïve Bayes model. From the above table, we can see that total of 716 wrong predictions are made using Naïve Bayes algorithm. This model shows an accuracy score of 92%.

Class	Non-fraud	Fraud
Non-fraud	98853	657
Fraud	59	114

Following confusion matrix shows the number of correct and wrong predictions made using knn model. From the above table, we can see that total of 93 wrong predictions are made using knn algorithm. This model shows an accuracy score of 83%.

Class	Non-fraud	Fraud
Non-fraud	98459	44
Fraud	147	33

The below figure shows the comparison graph of both the implemented algorithms. It also proves that Naïve bayes algorithm is giving a better accuracy over the KNN algorithm.



5. CONCLUSION

From the experiments, the result that has been concluded that the Naïve bayes will perform better with a larger number of training data, but speed during testing and application will suffer. Application of more pre-processing techniques may also help and makes contribution in the detection. The other algorithms suffer from the imbalanced dataset problem and requires more preprocessing to give better results. The results shown by these algorithms are great but it could have been better if more preprocessing have been done on the data.

REFERENCES

[1] Raj S.B.E., Portia A.A., "***Analysis on credit card fraud detection methods***", Computer, Communication and Electrical Technology International Conference on (ICCCET) (2011), 152-156.

[2] 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).

[3] 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).

[4] Phua C., Lee V., Smith, Gayler K.R., "***A comprehensive survey of data miningbased fraud detection research***". arXiv preprint arXiv:1009.6119 (2010).

[5] 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.

[6] 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.