International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 08 | Aug 2021 www.irjet.net p-ISSN: 2395-0072

Comparative Analysis of Credit Card Fraud Detection Using Machine Learning and Deep Learning Techniques

Subhash P[1], K R Sumana[2]

[1]*PG Student, Department of MCA, The National Institute of Engineering, Mysuru, Karnataka, India*
[2]*Assistant Professor, The National Institute of Engineering, Mysuru, Karnataka, India*

Abstract - Credit Card fraud is a sort of identity theft where thieves acquire or receive cash advances from another user's credit card account. This may occur through the use of a user's current accounts, physical credit card robbery, account number or PINs, or through the opening of an unknown credit card account in the user name. The Credit Card fraud detection project identifies the fraudulent nature of the new transaction by shaping the credit card transactions with the knowledge of those which have been fraudulent. In order to detect, if a transaction is a normal payment or a fraud, we will employ several predictive models. The strategies for classification are promising ways to identify fraud and non-fraud transactions. Sadly, classifying techniques do not work well in certain circumstances when it comes to big disparities in data distribution. In our work, we will be applying Machine-Learning algorithms: Logistic Regression, SVM, Naive Bays, Decision Trees, Random Forests and Deep Learning algorithm to predict fraud through Artificial Neural Networks. Results are analysed and compared.

Key Words: Credit Card, Machine Learning(ML), Deep Learning(DL), Decision Tree(DT), Logistic Regression(LR), Support Vector Machine(SVM), Naive Bayes(NB), Artificial Neural Networks(ANN).

1. INTRODUCTION

Credit card theft is a growing big issue which it costs banks and card service providers a huge amount of money. Banking institutions incorporate a range of protection techniques to try to prevent account misuse. Fraudsters get more sophisticated as security solutions become much more complex, i.e. fraudsters alter their strategies over time. As an outcome, improving fraud detection and prevention procedures Security modules aiming at blocking fraud is vital. Fraud detection has become a critical step towards reducing the negative impact on the delivery of services, prices and reputation of the company of fraudulent transactions. There are different methods for detecting fraud, with a view to maintaining a high quality of service. Maximum service performance while reducing to a minimum the number of deaths. Fraud is expensive and fraud detection can save a lot of money before the information is captured. The system is highly accurate and features few false alerts. Edge and Falcone Pre-decisional argue that authentic proactive processing considerably minimises the time scope available in the computer analysis and the accurate decision taken in response to new transactions. The proactive approaches also boost possibilities for early fraud alerting.

The faster a system for detecting fraud, the better is Fraud detection systems are trained on past transactions to decide new ones. In most circumstances, this training process is often paralleled. The amount of previous transactions processed by bypassing a time frame with less difficult approaches can be reduced in order to save computing time. However, each of these approaches can occasionally lead to reduced accuracy, leading to the lack of more incidents of fraud and the creation of more false alarms. As a result, a powerful tool is needed to carry out and process transactions in the shortest possible time by the fraud detection system.

2. LITERATURE REVIEW

Fraud detection concerns a big number of financial organizations and banks as this crime costs them roughly $ 67 billion per year. There are several sorts of fraud: insurance per fraud, credit card fraud, statement fraud, securities fraud etc., Of all of them, credit card fraud is the most common type. It is defined as an unlawful use of a credit card account. It occurs when the cardholder and the card issuer are not aware that the card is being used by a third party [1]. The intelligent approach presented for the detection of fraudulent credit card transactions optimized tree-based light gradient boost frame algorithms of learning. In the proposal, a Bayesian approach the optimization hyper parameter algorithm is smart Integrated with the LightGBM algorithm parameters to be tuned. The LightGBM method can be used fast Manage massive quantities of data and process the distributed data. It has been developed by Microsoft as an open source project. A LightGBM technique that can integrate unique features into a single package, the proposed approach can then be regarded to produce the same-featured histograms based on feature bundles. This approach is based on a feature scan algorithm. Based on theoretical time complexity, the complexity of the computation approach proposed was computed as follows: D (n Todemm), where m indicates the number of samples of the datasets, and m indicates the number of bundles [2]. We

present a new way of detecting fraud in this research. We use similar cardholder activity patterns to develop a recent cardholder conduct profile. Thus, we suggest a manner of resolving the model's potential for adaptation. The True Label Info from the transactions can be used fully by a feedback mechanism to alleviate the problem of drift. According to a number of incoming transactions it will change its own rating score. This online approach of detection of fraud can modify its parameters dynamically, in order to respond to cardholder transactions in time. The performance and effectiveness of our technique are shown by experimental findings. All these can be 80 percent accurate when detecting transactions compared with two other approaches. But AggRF is only good in FFDR and (RawLR) only in the CDDR. Our proposed strategy can improve FFDR and CDDR performance, as well as enhance average reminder and precision [3].

3. PROPOSED SYSTEM

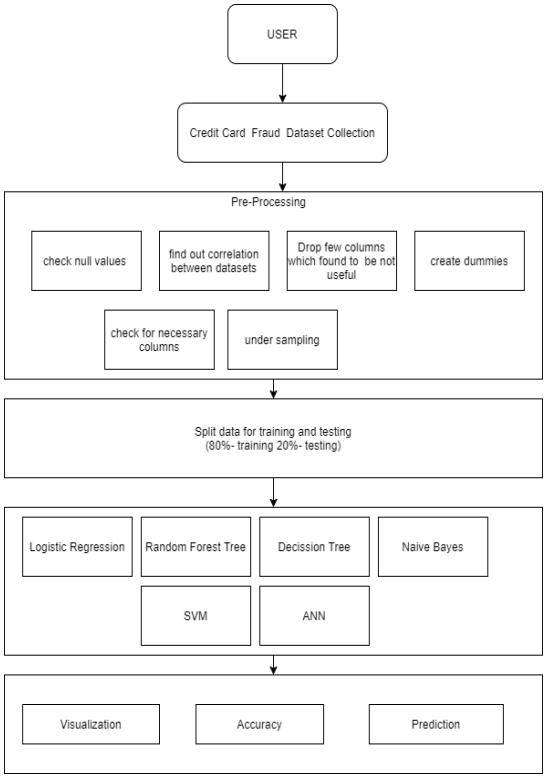

Fig-1 Block Diagram of the Proposed work

The workflow of the Detection of Credit Card fraud as depicted in Fig-1. The data set is gathered to create a prediction model. The first is the preprocessing of the dataset via preprocessing methods. Find the mean and mode of each attribute's total data, check in the dataset null values, missing values and irrelevant data. Use preprocessing technology and save the dataset in .csv file format using the preprocessing technique. Divided our data into workouts and tests, i.e. 80% of training data and 20% of test data, to construct the model.

4. IMPLEMENTATION

4.1 Dataset and Preprocessing

In the dataset, European cardholders are including credit card transactions done in September 2013. In this set of data, 492 scams from 284,807 transactions have occurred over the preceding two days. The transaction dataset is severely skewed and just 0.172% of the transactions are good (fraud).

It accepts just numerical input variables transformed by PCA. We cannot give the original data and other background material due to confidentiality problems. The key elements obtained by PCA are V1, V2,.....V28; 'Time' and 'Amount' are the only features that remain unaltered by PCA.

4.2 Data Analysis

Despite the anonymity of almost all predictors, we opt to concentrate our data analyses on unanonymised transaction time and amount predictors. There are 284.807 transactions in the data collection. The average value of this data set is 88.35 dollars, with a total of 25, 691,16 dollars for the larger transaction. On the other hand, the distribution of the money value of all the transactions, as we may assume on average and maximum, is notably rectangular. There is an insignificant majority of transactions, with just a minor part approaching the maximum.

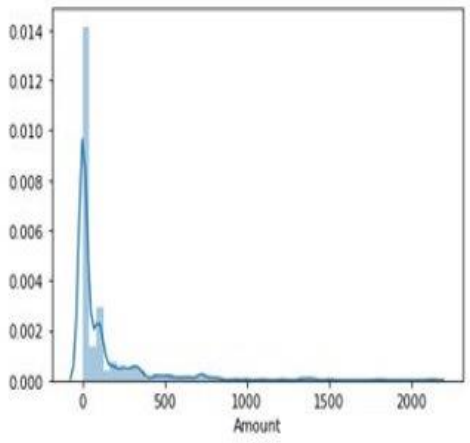

Fig-2 Distribution of Amount

The time was recorded in seconds from the first data gathering transaction. As a result, all transactions reported in two days can be determined by this dataset. The currency value of transactions is bimodal in contrast to the unimodal one. This results in a major reduction in volume of transactions around 28 hours following the first transaction. Although it is not possible to determine the original transaction time, the drop-in volume occurred at night is probable. The quantity and time distributions in the Credit Card Fraud data set are graphically displayed in Fig-2 and Fig -3.

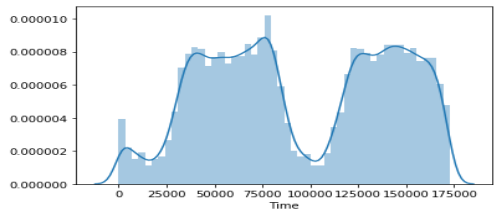

Fig - 3 Distribution of Time

What about the class distributions? What proportion of transactions are fraudulent, and what proportion are not? As one might expect, the vast majority of transactions are legal. In reality, only 0.17 percent of the transactions in this data set were fraudulent, leaving 99.83 percent of them non-fraudulent. This large contrast is seen in the Fig-4.

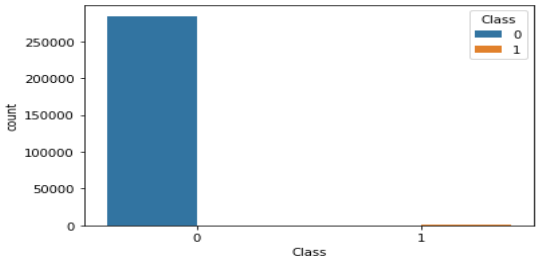

Fig-4 Fraud and Non-Fraud Count

Finally, it would be good to know if our predictors, especially as regards our class variable, had any significant associations. A heat map is one of the more visually appealing methods to look at this. As seen in Fig-5, some of our predictors appear to be linked to the class variable. However, for a large number of variables there appear to be very few significant connections.

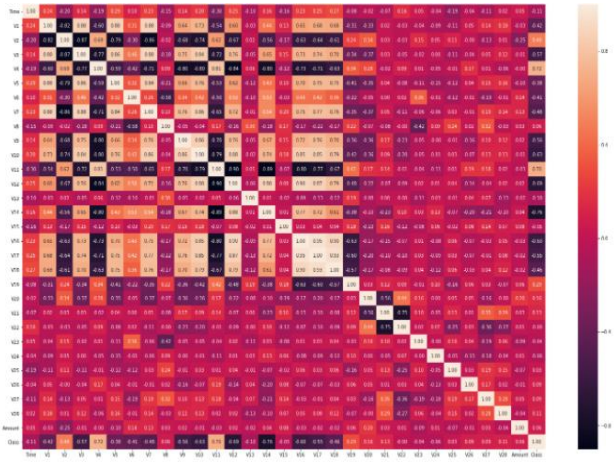

Fig-5 Co-Relation Heat Map

4.4 Training And Testing

The acquired data is divided into two portions using data mapping: 80% training data, and 20% test data. The data have been divided into training and testing sets to assign data points to the former and to the latter in the modelling data set. Therefore, a model is trained by means of a training set and used for a test set. This can be assessed in our application.

5. RESULT ANALYSIS

It is based on how accurate each algorithm is in detecting the fraud that results in the final results. Below Fig-6 shows the comparison chart of the all algorithms of ML/DL that are used in our work and observed the slight difference between the them on their accuracy.

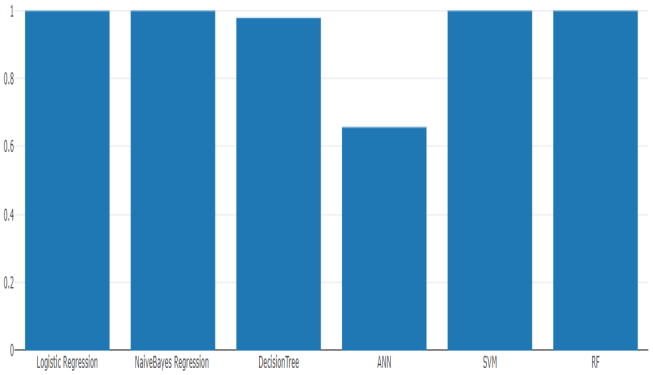

Fig-6 Comparison Chart of all Algorithms.

6. CONCLUSIONS

We have analysed Machine Learning and Deep Learning algorithms for the detection of fraud in Credit Card transactions. We started by comparing this with methods of Machine Learning, such as Logistical Regression, Support Vector Machine, Naive Bayes, Random Forest, and Decision Trees. Finally, we used Artificial Neural Network to identify fraud for Credit Card transactions that was challenging to train, but was good to recognise. The performance is assessed on the basis of precision, accuracy and reminder. The results for algorithms based on accuracy are like this Logistic Regression 99.91%, Decision Tree 99.93%, Naive Bayes 97.85%, SVM 99.93%, Random Forest 99.92%, ANN 65.67%.

REFERENCES

[1] An experimental study with imbalanced classification approaches for credit card fraud detection Author: S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. Hacid, H. Zeineddine Year: 2019

[2] An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine Author: A.A. Taha, S.J. Malebary Year: 2020

[3] Credit card fraud detection: a novel approach using aggregation strategy and feedback mechanism. Author: C. Jiang, J. Song, G. Liu, L. Zheng, W.Luan Year: 2018

[4]https://www.kaggle.com/mlg-ulb/creditcardfraud