IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

Credit Card Fraud Detection

Prof. Radha Shirbhate[1], Somnath Borude[2], Vishal Hadke[3], Sushama Walunjkar[4], Shubham Bhujade[5]
Professor, Department of Artifcial Intelligence and Data Science[1]
Students, Department of Information Technology[2,3,4,5]
G H Raisoni College of Engineering and Management, Wagholi, Pune, Maharashtra, India

Abstract: *Now a days online transactions have become an important and necessary part of our lives. Credit card fraud detection is presently the most frequently occurring problem in the present world. This is due to the rise in both online transactions and ecommerce platforms. Credit card fraud generally happens when the card was stolen for any of the unauthorized purposes or even when the fraudster uses the credit card information for his use. As frequency of transactions is increasing, number of fraudulent transactions are also increasing rapidly. In order to reduce fraudulent transactions, machine learning algorithms like Naïve Bayes, Logistic Regression, and KNNetc. are discussed in this paper. The same set of algorithms are implemented and tested using an online dataset. Through comparative analysis it can be concluded that Logistic regression, naïve Bayes,and K-Nearest Neighbor (KNN) algorithms perform better in fraud detection.*

Keywords: Credit card, Fraud detection, Machine learning, supervised learning, Naïve Bayes, Logistic regression, andK-Nearest Neighbor (KNN).

I. INTRODUCTION

Due to rise and acceleration of E- Commerce &government offices, corporate industries, finance industries, and many other organizations. In the present world there has been a tremendous use of credit cards for online shopping which led to High amount of frauds related to credit cards. In the era of Digitalization, the need to identify credit card frauds is necessary.

The primary goal is to make a fraud detection algorithm, which finds the fraud transactions with less time and high accuracy by using machine learning based classification algorithms. As technology is advancing rapidly, the payment by cash is reduced and online payment gets increased, this paves way for the fraudsters to make anonymous transactions.

In some modes of online payments, only card number, expiration date, and cvv are required and that data may be lost without our presence, in some cases we don't even know our data is being stolen.

Fraud detection involves monitoring and analysing the behaviour of various users in order to estimate detect or avoid undesirable behaviour. In order to identify credit card fraud detection effectively, we need to understand the various technologies, algorithms and types involved in detecting credit card frauds.

1.1Types of Frauds

1. Online and Offline
2. Card Theft
3. Data phishing
4. Application Fraud
5. Telecommunication Fraud

II. RELATED WORK

Fraud in any way is a criminal activity and is an offence, credit card fraud is stealing money. There are many studies in which they tried to find whether a transaction is fraud or not. Still having many challenges and tries to overcome those problems. When done the literature survey on various methods of credit card fraud detection, we can conclude that to detect credit card fraud there are many other approaches in Machine Learning itself.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

In 2020 Ruttala Sailusha, V. Gnaneswar , R . Ramesh , .G. Ramakoteswara Rao have explained the work on Random Forest Algorithm and Adaboost Algorithm .They have taken the highly skewed dataset and worked on such type of dataset. The Logistic regression algorithm is similar to the linear regression algorithm. The linear regression is used for the prediction or forecasting the values. The Adaboost algorithm is fraud cases to classify the transactions which are fraud and non-fraud. The algorithms used are random forest algorithm and the Adaboost algorithm. The results of the two algorithms are based on accuracy, precision, recall, and F1score. The ROC curve is plotted based on the confusion matrix. The Random Forest and the Adaboost algorithms are compared and the algorithm that has the greatest accuracy, precision, recall, and F1-score is considered as the best algorithm that is used to detect the fraud. From their work they have concluded that the highest accuracy is obtained for both the Adaboost and Logistic Regression.

In 2021 D. Tanouz, R Raja Subramanian, D. Eswar, G V Parameswara Reddy, A. Ranjith kumar, CH V N M praneeth have explained the work on Decision Tree ,Random forest Logistic Regression ,Naïve Bayes Classification . Naïve Bayes on among the classification algorithm. This algorithm depends upon Bayes theorem. Bayes's theorem finds the probability of an event that is occurring is given. The Logistic regression algorithm is similar to the linear regression algorithm. The linear regression is used for the prediction or forecasting the values. The logistic regression is mostly used for the classification task. The J48 algorithm is used to generate a decision tree and is used for the classification problem. The J48 is the extension of the ID3 (Iterative Dichotomieser). J48 is one of the most widely used and extensively analysed areas in Machine Learning. This algorithm mainly works on constant and categorical variables. They have concluded that the Random Forest algorithm has the highest accuracy among the other algorithms and is considered as the best algorithm to detect the fraud. Random forest classifier performs best with having 96.7741% accuracy , 100% precision , 91.1111% recall , 95.3488% f1 scores and 95.5555 ROU-AUC score.

III. METHODOLOGY

3.1 Dataset

The dataset is sourced from ULB Machine Learning Group and description is found in [32]. The dataset contains credit card transactions made by European cardholders in September 2013. This dataset presents transactions that occurred in two days, consisting of 284,807 transactions. The positive class (fraud cases) make up 0.172% of the transactions data. The dataset is highly unbalanced and skewed towards the positive class. It contains only numerical (continuous) input variables which are as a result of a Principal Component Analysis (PCA) feature selection transformation resulting to 28 principal components. Thus a total of 30 input features are utilized in this study. The details and background information of the features cannot be presented due to confidentiality issues. The time feature contains the seconds elapsed between each transaction and the first transaction in the dataset. The 'amount' feature is the transaction amount. Feature 'class' is the target class for the binary classification and it takes value 1 for positive case (fraud) and 0 for negative case (non fraud).

3.2 Hybrid Sampling of Dataset

Data pre-processing is carried out on the data. A hybrid of under-sampling and over1sampling is carried out on the highly unbalanced dataset to achieve two sets of distribution for analysis. This is done by stepwise addition and subtraction of a data point interpolated between existing data points till over-fitting threshold is reached

3.3 Naïve Bayes Classifier

Naïve Bayes a statistical approach based on Bayesian theory, which chooses the decision based on highest probability. Bayesian probability estimates unknown probabilities from known values. It also allows prior knowledge and logic to be applied to uncertain statements. This technique has an assumption of conditional independence among features in the data. The Naïve Bayes classifier is based on the conditional probabilities and of the binary classes (fraud and non fraud)

3.4 K-Nearest Neighbour Classifier

The k-nearest neighbour is an instance based learning which carries out its classification based on a similarity measure, like Euclidean, Mahanttan or Minkowski distance functions. The first two distance measures work well with continuous variables while the third suits categorical variables. The Euclidean distance measure is used in this study for the kNN classifier.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

3.5 Logistic Regression Classifier

Logistic Regression which uses a functional approach to estimate the probability of a binary response based on one or more variables (features). It finds the best-fit parameters to a nonlinear function called the sigmoid. The sigmoid function (σ) and the input (x) to the sigmoid function

IV. ARCHITECTURE

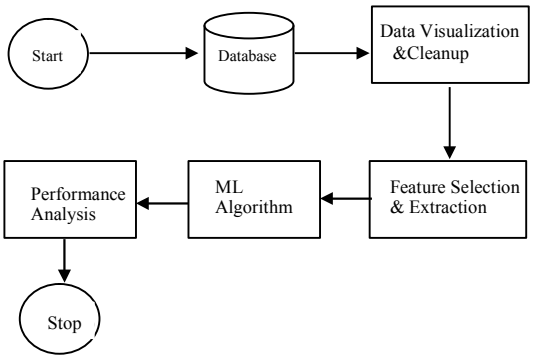

Figure: Architecture of credit card fraud detection system

4.1 System Implementations Plan

The fraud detection module will work in the following steps:

1. Upload the dataset which has the transactions and amount are treated as credit card transactions.
2. After dataset upload the feature selection and dimensionality reduction step is done.
3. We separate out the dataset as a train dataset and test dataset.
4. The train dataset are given to machine learning algorithms as an input.
5. We train the model using this dataset.
6. After that we test the model using testing dataset.
7. The valid transactions are treated as genuine transactions.
8. And fraudulent transactions are treated as Fraud transaction.
9. After that we check the accuracy.

4.2 Advantages:

1. Faster detection
2. Higher Accuracy
3. Improved efficient with larger data
4. Fewer false declines
5. Less manual work needed for additional verification
6. Ability to identify new patterns and adapt to changes.

V. CONCLUSION

Efficient credit card fraud detection system is an utmost requirement for any card issuing bank.Infiltrating merchant system can still millions of cards, cash out via underground marketplace online. Credit card fraud is without a doubt and Act of criminal dishonesty. Credit card fraud detection want to help the peoples from there wealth loss . Machine-learning techniques are mostly preferred in fraud detection, because of its high accuracy and detection rate. Still researchers are struggling to get more accuracy and detection rate.

REFERENCES

[1]. Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020) IEEE Xplore Part Number: CFP20K74-ART; ISBN: 978-1-7281-4876-2

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, December 2022
Impact Factor: 6.252

[2]. Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021) IEEE Xplore Part Number: CFP21K74-ART; ISBN: 978-07381-1327-2

[3]. Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018) IEEE Conference Record: # 42666; IEEE Xplore ISBN:978-1-53863570-4

[4]. Chaudhary, K. and Mallick, B., (2012). Credit Card Fraud: The study of its impact and detection techniques, International Journal of Computer Science and Network (IJCSN), Volume 1, Issue 4, pp. 31 – 35, ISSN: 2277-5420

[5]. 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, ISSN 2229-5518

[6]. N. Mahmoudi, E. Duman, "Detecting credit card fraud by Modified Fisher Discriminant Analysis", Elsevier Expert System with Application, 2015, pp. 2510-2516.

[7]. N. Halvaiee, M. Akbari, "A novel model for credit card fraud detection using Artificial Immune System", Elsevier Applied Soft Computing, 2014, pp. 40-49.