Date of publication 30th June 2024

Digital Object Identifier

Credit Card Fraud Detection Using Machine Learning Techniques

Javeria Rizwan¶ (f2021266098@umt.edu.pk)

ABSTRACT Nowadays, Usage of Credit cards have increased all over the world as it's a more easy and efficient way to make payments, but with the increase in usage of credit cards, the misuse of them has also increased. These frauds result in huge loss for the consumers as well as the companies. This research is done to detect those frauds happening in the world using different Machine learning approaches.The research contains the process of data collection, data processing techniques, feature scaling of the data-set, model training, model testing, evaluation measures etc. The challenges we faced during this were an imbalanced data set, getting the accurate split of the data. However after training and testing the models, The accuracy of Logistic Regression model was better than other implemented models, it gained 95% accuracy.This research paper provides the important information about the methods to detect credit card fraud detection and other useful techniques related to fraud detection.

INDEX TERMS Credit card fraud,Fraud detection,Identity theft,Machine learning,Deep learning,Data imbalance,Feature selection,Classification algorithms

I. INTRODUCTION

Credit Card Frauds (CCF) is known as an identity theft in which an individual other than the owner of the credit card uses others credit card information to pay for their purchases and to withdraw money from the accounts. The owners are unaware of all this that their information is being stolen and intercepted unless some purchases are made. In today's era e-commerce and other online shopping sites are increasing their production due to which online transactions are continuously increasing. Due to this the fraudulent activities are also increasing worldwide.

The aim of this study is to demonstrate legitimate and fraudulent transactions. For this we focus on supervised and unsupervised learning along with class imbalance problems. As the dataset we used in our research analysis is highly unbalanced so to handle this un-balancing of data is also a challenge.

II. METHODOLOGY

The following steps were followed in the research methodology:

A. DATA COLLECTION

The work to detect CCF includes the concept of ML, DL, CCN, ensemble feature ranking approaches. Nowadays data Imbalance presents various challenges to analysis of Data in the real world, especially credit card frauds due to online fraudulent transactions that have resulted in huge loss worldwide. For this all the top approaches of ML and deep learning are applied to detect CCF. The paper is concerned with these following approaches

A: Machine Learning Approaches:

ML has different frameworks. ML approaches provide a solution for CCF detection through its various algorithms like Random Forest (RF), Logistic Regression (LR), Naive Byes (NV), Support Vector Classifier (SVM), Decision trees and KNN. In all these algorithms, the researchers combine the techniques for balancing data, ensemble, and feature ranking to construct solid decision classifiers.

Research Methodology:

The baseline methodology concluded from research work is:

B. MODEL SELECTION

In our research, classification Models were used to train the models. The architecture of models is to predict the probability of frauds.

C. DATASET

The dataset was taken from kaggle. The following dataset used to detect CCF contains 284,807 transactions out of which 492 are fraudulent transactions. The dataset contains the continuous data. The dataset contains the transaction records of European card holders. If we talk about the balancing of the data, then the following dataset is highly imbalanced. Due to the confidentiality issues of customers PCA transformation based on 'Time' and 'Amount' feature is used in this dataset to prevent such issues. Due to the imbalanced nature of Dataset the accuracy is being measured by the logistic regression and random forest techniques.

Dataset Link:
https://drive.google.com/file/d/1gEZB4PFcAsIjE0bks_OtFRuMKJpYlnom/view?usp=sharing

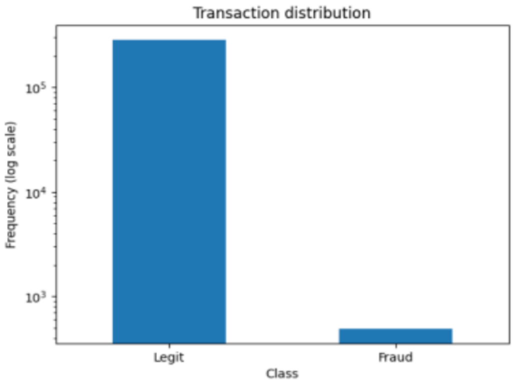

FIGURE 1: Transaction Distribution

D. DATA PREPROCESSING

The following data preprocessing steps were performed:

- **Data Cleaning**: Removing duplicate entries, handling missing values, and dealing with outliers.
- **Data Splitting**: Splitting the data into training and testing phases, using different ratios such as 70-30 and 60-40 to analyze accuracy.
- **Data Cleaning**: Handling missing values, deleting duplicate entries, and addressing outliers in the dataset.
- **Feature Scaling**: Removing irrelevant features and keeping only the relevant ones in the dataset.
- **Feature Engineering**: Creating new features or establishing patterns and relationships with existing features.

By performing data prepossessing, all the outliers in the data were removed, the data became more clean and clear for the model testing, the split method was used to train and test the data. Keeping only the relevant features to make the model more accurate.

E. MODEL TRAINING

Our models were trained on the data-set of Credit card fraud detection, it involved the data of the both transaction which were legit and fraud, the features included were location, time , date etc. The ultimate goal was to train the model to predict the fraud activity of the transactions, and the accuracy should be very better to check the model reliability.

F. MODEL TESTING

After training phase, the testing phase comes and we tested it on 70-30 and 60-40 data after that we calculated the accuracy of the data, confusion matrix, recall, precision and F1 scores. These evaluation measures were used to check the model effectiveness in the CCFD.

III. RESULTS AND DISCUSSION

A. EVALUATION MEASURES

These are some evaluation measures which are used to measure the performance and effective of the model. These help us with accuracy of the model, error rate etc.

- F-1 Score
- Accuracy
- Recall
- Precision

B. RESULTS

We will discuss the results of our model in this section. Our experiment was done on a Credit card fraud detection data set which we took from kaggle, it had over 284,807 transactions and 492 of them were considered fraudulent. We are gonna attach our results of each model below:

As we are getting same same values for precision, recall and F1 score so it means the model is consistent and more accurate.

TABLE 1: Evaluation Results of Credit Card Fraud Detection Models

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.95135	0.989	0.919	0.9528
Random Forest	0.924	0.967	0.88	0.9263
Decision Tree	0.924	0.920	0.939	0.93
Naive Byes	0.90	0.96	0.85	0.90
KNN	0.945	0.98	0.90	0.94
SVM	0.945	0.989	0.90	0.947

C. COMPARING MODELS RESULT AND DISCUSSION

These results show that both models have achieved high accuracy in the detection of fraud transactions. The LR model achieved around 95% accuracy with precision of 98% while other models also gained accuracy above 90% which shows they are also consistent

IV. CONCLUSION

In conclusion, This research paper focuses on identifying credit card fraud detection using different techniques of Machine learning algorithms.

Firstly, Machine learning algorithms like classifying models were used in the detection of Credit card fraud. These algorithms gave promising results with very high accuracy.This shows that the models we used were effective and efficient in identifying credit card fraud.

The evaluation results were quite impressive and it was a balanced performance in correctly classifying both legit and fraud transactions.

This research contributes to the field of credit card fraud detection by using different Machine learning and Deep learning algorithms. The importance of this is to frequently check the fraud transactions which will be detected by the model. This can be further enchanted by applying more techniques and making the models more accurate with no data biasedness

Overall, we came to this conclusion that Logistic Regression performed little better than other models in terms of accuracy and other evaluation measures. Although the other models accuracy was over 90 and according to the standards, it was very high and very consistent.

REFERENCES

[1] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, “Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms," *IEEE Access*, vol. 10, pp. 39700-39715, 2022, doi: 10.1109/ACCESS.2022.3166891.

[2] A. Kumar, D. Prusti, I. S. Purusottam, and S. K. Rath, “Real-time SOA based credit card fraud detection system using machine learning techniques," in *2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, doi: 10.1109/ICCCNT51525.2021.9579598.