As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology, 7(05): 60-62, 2021

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0705009
Available online at:http://www.ijmtst.com/vol7issue05.html

Credit Card Fraud Detection

Ritik Kumar Singh[1] | Amit Kumar Singh[1] | Anuj Pratap Singh[1]

[1]B. Tech (Computer Science and Engineering), Lovely Professional University

To Cite this Article
Ritik Kumar Singh., Amit Kumar Singh., Anuj Pratap Singh. Credit Card Fraud Detection. *International Journal for Modern Trends in Science and Technology* 7, 60-62 (2021).

Article Info
Received on 07-April-2021, Revised on 17-April-2021, Accepted on 28-April-2021, Published on 08-May-2021.

ABSTRACT

This research paper proposes a solution that should be deployed to identify whether the transaction is fraud or not. Although we know that most of the transaction takes place online meaning that this transaction can be theft on the go and will create problem to user therefore this paper focus on some particular machine learning algorithm for example Random forest Algorithm, Decision Tree Algorithm, Logistic Regression, Support Vector Machine, K Nearest Neighbour, XGBoost .Which aims at solving such kind of real-world problem.

KEYWORDS: *Supervised Machine Learning Algorithms, UI, Data Preparation, Model Deployment.*

I. Introduction

The government of India launched the Digital India campaign to optimize the digital usage of technology that will empower the progress of our country. This will save the time and manpower to an immense stretch. And online transaction is an enormous part of digitization. This has massively saved times and effort of both consumers and the industries to save their heavy bucks by not paying bulky rentals to physical stores. This online transaction involves debit and credit cards and other online payment options which raises the issue of fraud transactions. This is a soaring problem where the account owner does not hold any control of the transaction and is carried out by the third party. Credit card transactions are supposed to be more secure than any other payment options, but the financial institutions are somehow facing huge challenges to protect it from frauds. Due to these credit card frauds, banking and insurance companies are going through the pain of losing their customers' trust and it is also creating a barrier in the process of revenue growth. So, its real crucial for these financial institutions to put a stop to these kinds of hoodwinking. The problem is that these institutions get to know about these fraudulences only after encountering extensive losses. It has to be identified beforehand. In this era of technologies with so much of secure ciphering, there are still 492 frauds out of 284,803 transactions in a year according to a survey done by the government of credit card fraud detection dataset this year. Now here another problem arises that it's not only the quantity of scam is boosting but also the quality of these fraudlance is surging. Improved techniques in card skimming, phishing and carding has taken this credit card trick to another level. This will ultimately lead to customer agitate.

STRUCTURE OF PAPER

The paper is organized as follows: In Section 1, the introduction of the paper is provided along with the structure, important terms, objectives andoverall description. In Section 2 we discuss related work. In Section 3 we have discuss about some fraud detection approaches. Section 4 is Data Transformation this explains about the splitting of data. Section 5tells us about the methodology and the process description. Section 6 tells us about the conclusion of the paper with acknowledgement and references.

OBJECTIVES

In this era with such a strong encryption technology, in this era with such a strong encryption technology, there are still 492 fraud cases in 284,803 transactions every year. This is the result of an investigation conducted by the government on credit card fraud since this year. In this project, we use algorithms and libraries, such as NumPy, Pandas, Seaborn, Matplotlib, Logistic Regression, Decision Tree, KNN, Random Forest, SVM. These algorithms and libraries can help us build fraud detection models. If a specific set of fraud detection data is provided to the system, the existing system can predict whether the current transaction is fraudulent. In this system, we use Python language to check whether transactions are fraudulent.

RELATED WORK

There are numerous works that have been done related to credit card fraud detection using machine learning.

In this project, we have a dataset to be imported using the Pandas library. Then, we check the data by checking the size of the data and various parameters available in the data set. Find the target parameter and delete it from the data. Since the number of fraud cases is very small, the number of fraud cases can be counted instead of the number of fraud cases. Therefore, it is very important to balance the data set. After balancing the data set, we can perform data conversion, that is, divide the data set into training and test data sets. Therefore, the challenge we face is to use different supervised machine learning algorithms to build our model. We transfer training data to various algorithms and calculate their accuracy. We will implement the model with the highest accuracy on the server so that users can use it. We are connected. When users submit their data to the UI and click the "Submit" button, the model will process the data and generate results (regardless of whether the transaction is fraudulent or not).

II. Fraud Detection Approaches

As a result, organizations began to pay more attention to fraudulent activities using modern Technology. Initially, the company hired employees specifically to investigate such cases or Fraudulent transactions. Fraud occurred. This requires a lot of manpower, energy and time. Second, the company uses manual processes to rule-based decision making. However, this Idea also led to the unfortunate result of fraud detection. Nowadays, people come up with New ideas about fraud every day. From the above situation, the system does not work in this Situation. This requires adding new fraud ideas to the system and running it again. The company has now begun migrating to other systems to detect fraud. This means that the Company has begun to relabel AI and ml algorithms to detect fraudulent transactions. Currently, among all available options, using the MLA algorithm is the best idea. It works by Processing previous data from various loans. Card transactions or experience in evaluating Fraudulent and non-fraudulent transactions. In this project, the dataset we used was loaded from Kaggle. Before we start building the model, let's talk about the data.

- Size of datasets.
- Type and no. Of parameters used in the dataset.
- Target values.
- No. of samples under target values.

To build a model, we need to research or obtain some information about our data in order to plan what we can do. You can explore the above points using the Python Pandas library.

III. Data Transformation

Since the quantity column does not have a value range like other columns, we must adjust the value of the quantity column to a value range that can be included in the form so that we can add columns with the following content. The rest of the columns, we must apply a certain percentage for this Characteristics. In terms of scaling, we have a new column naming rule. The value of this rule is scaled from the quantity column and scaled after the quantity column is deleted, because they no longer serve us.

Used Algorithm:

Logistic Regression: Logistic regression is a mathematical method for predicting binary groups, it is Logistic regression. The result or target variable is a binary variable. This is a variant of linear regression, where the target variable is categorical.

KNN Algorithm: The KNN algorithm is a supervised machine learning algorithm that can be used to solve classification and regression prediction problems. This algorithm is very useful because it changes with the size of the data. problem.

Decision Tree: Decision trees can be used for classification problems. The decision tree uses a tree diagram or decision model. In the decision tree, each internal node represents a test by attribute, each branch represents a test result, and each leaf node represents a class name. We have to draw the tree when the boundary is over and over again. Before proceeding, let's take a look at some of them the term.

- **Instance:** The vector object or attribute that characterizes the input area is called an instance.
- **Attributes:** a collection of descriptions.
- **Definition:** The operation of converting an input to an output.
- **The concept of purpose**: the function we are looking for, i.e. Real solution
- **Assumption category:** This is the set of all possible functions.

SVM: Support Vector Machine is a linear model for classification and regression problems. It can be used to solve linear and non-linear problems and it can work well for many different practical problems. This algorithm is simple because it creates a line or a plane which separates the data into classes.

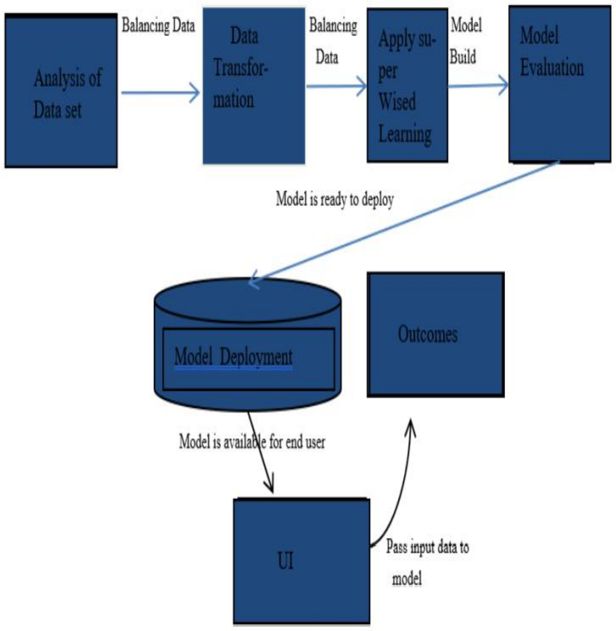
Random Forest Tree: Random forest, as the name suggests, is made up of a large number of individual decision trees that work together. Each tree in the random forest produces a class prediction, and the class with the most votes becomes the prediction of our model.

IV. METHODOLOGY

1. Importing the data with the help of pandas (python library).
2. After importing datasets, we tried to filter the fraud and Non-fraud cases and calculate the percentage of fraud cases.
3. Checked the accuracy with the help of ROC graph.
4. As the no. of non-fraud cases has more as compared to fraud cases, we need to balance the cases by inserting more fraud cases.
5. Again check the accuracy with the help of ROC graph after balancing the datasets whether it is increasing or not.
6. After that transform the data into training and testing data useful in building model.
7. Find the accuracy with different ML algorithms like (logistic, KNN, Decision tree, Random Forest, Support vector Machine).
8. Deploy the model with best accuracy with help of flask and connect it with user interface.
9. To cross check our model is working fine or not we passed the different inputs from testing datasets in the user interface and press submit button to check the output generated by model.

Process Description

The following block diagram makes it easier to understand how we proceed.



V. CONCLUSION

In this paper, Supervised Machine Learning Algorithm are used to find the problem of fraud detection. After building the best model by comparing the accuracy of models we are able to deploy it on the server and connect it with a user interface. To verify the model is working fine or not we build a user Interface and connect it with the server on which model has deployed we will pass the Data into the various parameters available on the user interface and press the submit button the model process the data and generates an output which shows weather the transaction is fraud or not.

References

[1] https://www.youtube.com/channel/UCh9nVJoWXmFb7sLApWGcLPQ

[2] https://www.youtube.com/user/krishnaik06

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

[4] https://www.geeksforgeeks.org/introduction-to-support-vector-machines-svm/

[5] https://www.geeksforgeeks.org/decision-tree/