DETECTING FRAUDULENT TRANSACTIONS

Tanoo Bijapur
dept. Of Information Science
S.D.M College of engineering
Dharwad,India

Sushma Managundi
dept. Of Information Science
S.D.M College of engineering
Dharwad,India

Anusha Shetty
dept. Of Information Science
S.D.M College of engineering
Dharwad,India

Lydia Adina
dept. Of Information Science
S.D.M College of engineering
Dharwad,India

Under the guidance of ,

Dr.V K Parvati

dept. Of Information Science
S.D.M College of engineering
Dharwad,India

***Abstract*—**When constructing a credit card fraud detection model we extract the right features from the transactional dataset. In this paper we propose to create the respective models for Random forest, Logistic regression, Support vector machine(SVM) and Naïve Bayes algorithm. And find the accuracy rates of each algorithms and get the most accurate algorithm to detect fraudulent transaction.

I. INTRODUCTION

Due to advancement in E-commerce, the online transactions have increased which has caused increase in the fraudulent transactions.

In the era of digital world it is necessary to identify the frauds and provide appropriate solutions by understanding various technologies algorithms and types involved in detection.

The dataset of transactions is given to the machine, which understands the data by analysing. And using machine learning algorithms the respective models are built which help in detecting fraud or not.

Fraud can be defined as the intent of deception in order to obtain financial gain and without the knowledge of the card owner and the Bank.

Credit card Fraud can be of types:

1. CARD-NOT PRESENT(CNP)
2. LOST AND STOLEN CARDS
3. CARD ID THEFT
4. ACCOUNT TAKEOVER.

II. ALGORITHMS

1.Random Forest

Forest is a collection of decision trees.Decision tree are mainly indulged in overfitting of training data.They show high variation for small change in the input data,hence is sensitive to specific data.The input data can be trained to ensure that it

remains error-proned.This algorithm helps to grow many such decision trees and gives average of all classification trees,hence it results in the reduction of variance.The random forest model is built to which a specified dataset is given,it gives the appropriate resultant value.

2.SVM

ML feed the data to machines and then it learns from the data without explicitly programming it.It can be used for both classification and regression algorithms. It uses two hyperplane to separate two new classes and then check the distance between them.

3.Naïve Bayes

Naïve Bayes is based on the Assumption which has independence amongst the Predictors. In simple terms, this could be put up as Naive Bayes Classifier which assumes that a particular feature in a class is not exactly directly related to any other feature.

P(H/E) = P(E/H) / P(E)

Naive Bayes model isn't difficult to build and is really useful for very large datasets. Along with simplicity, Naive Bayes is also considered to have outperformed all the highly sophisticated classification methods.

4.Logistic Regression

A logistic regression model is termed as a probabilistic model. It helps in finding the probability that a new instance

belongs to a certain class. Since it is probability, the output lies between 0 and 1.

Whenever we are using the logistic regression as a binary classifier, we can consider the classes to be a positive class and a negative class. We then find the probability. Higher the probability (greater than 0.5), it is likelier that it falls into the positive class. Similarly, if the probability is low (less than0.5), we can classify this into the negative class.

III. COMPARISION BETWEEN THE ALGORITHMS

ALGORITHM	ACCURACY RATE	ADVANTAGES	DISADVANTAGES
Random forest	98.60%	Predictive performance can compete with the best supervised learning algorithms	Training a large number of deep trees can have high computational costs
Logistic Regression	97.50%	Performs well when dataset is linearly separable.	Assumption of linearity of dataset b/w independent and dependent variable.
Naïve Bayes	90.00%	Its easily trained with small dataset	It assumes every feature is independent which is not true always.
Support vector machine	97.70%	It is effective in high dimensional spaces	It doesn't provide probability estimates.

IV. CONCLUSION

In this paper we try to find whether its fraud or not using the algorithms. We calculate the accuracy rates of each algorithms and by comparing the accuracy rates obtained, the best algorithm be determined.

We extract the essential features from the transactional data and try to find the genuine feature from all these algorithms.

REFERENCES

[1] *University of lexembourg,"Deteceting fraud detection using periodic features"Aljendro Correa Bahnsen.*

[2] *Institute of Control Sciences" credit card fraud Detection system",Moscow Russia*

[3] *Karunya University "Analysis on credit card fraud detection methods" Asst professor Annie Portia.*

[4] *IEEE journal,"A novel approach using aggregation strategy and feedback mechanism" Changun jiang,2017*

[5] *Tongji University,"Random forest for credit card fraud detection",Shiyang xuon ,Shangai China.*

[6] *Tongji University,"Algorithms for detection"Guanjum liu,China.*