iJRASET

International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: XI Month of publication: November 2021

DOI: <https://doi.org/10.22214/ijraset.2021.39046>

www.ijraset.com

Call: 08813907089

E-mail ID: ijraset@gmail.com

Predictive Analytics with Machine Learning for Fraud Detection

A Sampath Abhishek¹, K. Krishna Prasad², Dr. M Manickam³

^{1,2}Student SRM Institute of Science and Technology

³Assistant Professor SRM Institute of Science and Technology

Abstract: The popularity of online shopping is growing day by day. In financial year 2021, over 40 billion digital transactions worth more than a quadrillion Indian rupees were recorded across the country. As the number of credit card users rise world- wide, the opportunities for attackers to steal credit card details and subsequently, commit fraud are also increasing. Since humans tend to exhibit specific behavioristic profiles, every cardholder can be represented by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent etc. So these frauds can be detected through various algorithms mainly random forest and logistic regression. To enhance the boost and build model with much more efficiency adaboost is also added.

Keywords: Fraud detection, behavioristic profile, random forest, logistic regression, adaboost.

I. INTRODUCTION

Due to the growing age of internet, wide set of people started utilizing the services of various products through Internet. With this growing demand in the market for online services, the chances of data theft and fraud have been increasing eventually. The Credit card fraud is vital among them as credit card transaction has become the most convenient way of transaction, whether online or offline. Financial fraud of this type can greatly affect corporate, organizational and government sectors of a country. So detection of such fraud involves monitoring the activities of population of users In order to estimate or avoid such objectionable behavior. This type of problem is particularly challenging as it is characterized by various factors such as fraudulent Behavior and class imbalance. So, this is a problem which demands attention from various fields of Computer science communities especially from the field of data science and machine learning. When a fraud detection system is implemented in a real world massive stream of payment transactions are quickly scanned by automatic tools that determines which transactions to be authorized. Machine learning algorithms such as random forest and logistic regression are then employed to analyze these authorized transactions and report suspicious ones. So, these reports are then provided to data analysts who cumulate fraudulent transactions and card holder's information to banking professionals. They check and confirm with the users and feedback report is then provided back to the model to eventually improve fraud detection performance over time.

II. LITERATURE REVIEW

Many researches are conducted for credit card fraud detection. Numerous literatures have been published and are ready for public usage. According to the journal entitled Fraud detection using machine learning, it was mentioned that dataset has a class imbalanced which must be essentially considered while choosing algorithms. A Comparative analysis of Various Credit card fraud detection took this issue further and suggested to improves boosting efficiency, which is one one the reasons why we also added adaboost. Also according to survey published by international journal related to Collating ml algorithms is very much essential as well.

III. METHODOLOGY

In this paper we propose to use random forest and logistic regression to detect the anomalous activities and identify the fraudulent transaction.

The basic rough architecture diagram can be represented with the following figure.

It begins with the collection of historical or raw data. The dataset for the proposed model is taken from Kaggle, a data analysis website which provides datasets. The model procedure continues to follow a series of Data preprocessing steps.

A few very important data preprocessing steps include:

Importing libraries

Importing the dataset

Handling the missing data

Encoding the categorical Data

Splitting the dataset into test and train data

Feature scaling.

The historical data is then structured into an organized data and passed on to the model.

The model picks up the corresponding dataset, performing statistical building and applying algorithms

To train the model. The trained model is tested again test data set and is then deployed.

The corresponding algorithms used inside the proposed are stated as below:

A. Random Forest

This algorithm works with the help of decision trees by splitting dataset into n nodes.

```
Import numpy as np
Import pandas as pd
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix, accuracy_score
classifier = RandomForestClassifier(n_estimators = 641, random_state=0)
classifier.fit(x_train, y_train)
y_pred = classifier.predict(x_test)
n_errors = (y_pred != y_test).sum()
cm = confusion_matrix(y_test, y_pred)
sns.heatmap(cm, annot=True)
print(accuracy_score(y_test, y_pred))
```

B. Logistic Regression

The pseudocode for this algorithm is written as:

```
Import numpy as np
Import pandas as pd
From sklearn.linear_model import LogisticRegression
Lr_model = LogisticRegression()
Lr_model.fit(x_train, y_train)
Lr_pred_train = Lr_model.predict(x_train)
Lr_pred_test = Lr_model.predict(x_test)
```

1) Implementation

The idea to identify fraudulent transaction based on identifying specific behavioristic profiles

Is difficult to implement as it requires the cooperation from bank's, which aren't willing to share

Information due to market competition, also due to legal reasons and protection of data.

SO we researched few papers and gathered information on effective implementation of the fraud detection system.

In order to maximize time efficiency and overhead charges, we can include a new parameter

Like a c-code combination of first three digits of password and phone number of user.

IV. RESULTS

After preprocessing of data set we have applied different machine learning models to the data set to predict the fraud anomalies in credit card.

The following are the classification report which includes precision score, recall and f1-score.

<pre>[] from sklearn.metrics import classification_report</pre>					<pre>from sklearn.metrics import classification_report</pre>				
<pre>[] print(classification_report(ytest, lr_pred_test))</pre>					<pre>print(classification_report(y_test , y_pred))</pre>				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	1.00	1.00	1.00	85307	0	1.00	1.00	1.00	85149
1	0.86	0.60	0.71	136	1	1.00	1.00	1.00	85440
micro avg	1.00	1.00	1.00	85443	accuracy			1.00	170589
macro avg	0.93	0.80	0.85	85443	macro avg	1.00	1.00	1.00	170589
weighted avg	1.00	1.00	1.00	85443	weighted avg	1.00	1.00	1.00	170589

The final result is critically reviewed with the help of a confusion matrix.

V. CONCLUSION

The online frauds are prevalent all across the world. The increasing fraud activities could damage our economy. This paper focused on how effectively we can minimize the chances of credit card fraud with the help of machine learning algorithms.

While the algorithm does reach over 85.6% its precision remains only at 28% when a tenth of the data set is taken into consideration. However, when the entire dataset is fed into the algorithm, the precision rises to 33%. This high percentage of accuracy is to be expected due to the huge imbalance between the number of valid and number of genuine transactions.

REFERENCES

- [1] Adi Saputra¹, Suhajito²: Fraud Detection using Machine Learning in e-Commerce, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 9, 2019.
- [2] Heta Naik , Prashasti Kanikar: Credit card Fraud Detection based on Machine Learning Algorithms, International Journal of Computer Applications (0975 – 8887) Volume 182 – No. 44, March 2019.
- [3] Yashvi Jain, Namrata Tiwari, Shripriya Dubey, Sarika Jain: A Comparative Analysis of Various Credit Card Fraud Detection Techniques, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-5S2, January 2019 .
- [4] Navanshu Khare , Saad Yunus Sait: Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models, International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 825-838 ISSN: 1314-3395.
- [5] Yong Fang¹, Yunyun Zhang² and Cheng Huang¹, Credit Card Fraud Detection Based on Machine Learning, Computers, Materials & Continua CMC, vol.61, no.1, pp.185-195, 2019.

10.22214/IJRASET

45.98

IMPACT FACTOR:
7.129

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)