Zep Analytics

DATA SCIENCE INTERNSHIP

ANKITA DAS

Agenda

- 1. EDA (Exploratory Data Analysis)
- 2. Findings
- 3. Results

EDA (Exploratory Data Analysis)

I had downloaded my dataset from kaggle, dataset on Online payment fraud detection with 1048575 rows & 9 columns only. In every columns there have so many unique values but in the one column (isFraud) there have only two unique values (0 & 1) & this column would be my targeted column. Also there have no any null values. Good thing over this dataset is that there have less fraud case than non-fraud case.

Findings

My motive is to find fraud detection, accuracy of my model, type I & type II error.

Results

First, I have scaled my data by using StandardScaler, then train my data by x_train & y_train. Where X is my counting values & y is targeted value. Because this dataset is supervised, labeled & there have a targeted column so I had to use Logistic Regression. By using it I got my model accuracy 99% for both x_test & y_test. Using confusion matrix I got type I error is 63 & type II error is 168 & actual accuracy 99%. On the other hand, I evaluate classification report because if any error would have within the confusion matrix then it will correct those error, and now our accuracy getting higher.

Thank You