IBM Coursera Capstone Project

Credit Card Fraud

By:Tarushi Pathak

Introduction

Many of us fall prey to the fraud messages and calls sent to us and end up losing a huge amount of our savings. Sometimes the fraud messages are too obvious and other times they are extremely well planned and hard to figure out. Either way, the person ends up losing a huge amount of money.

It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase. Hence, in this project I'll be focusing on using the means of clustering the credit card purchase data in order to understand the fraud character traits which will further help the credit companies to recognize fraudulent credit card transactions faster.

Business Problem

The objective of this capstone project is to analyse and recognize the character traits of credit card fraud and cluster them on those basis. Using Data Science Methodology, we will aim to find answer to the question , Is the given data and trait of the corresponding transaction pointing to fraud or not? , using data analysis and machine learning techniques.

Target Audience

The credit card companies using this machine learning model will be able to assure protection from fraud to their customers to a certain extent. Customers ,also, will shift to those companies who have better results at detecting fraud. Hence,our target audience will be the credit card companies.

Data

We will be using kaggle's credit card fraud detection set.It can be find at this link: https://www.kaggle.com/mlg-ulb/creditcardfraud

As per the data provider,the datasets contains transactions made by credit cards in September 2013 by european cardholders.
The dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, … V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependent cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.