Infrastructure walkthrough on Machine Learning project.

Project:

Credit Card Fraud detection/Prediction WebApp Using Machine Learning

Presented By Pius Odhiambo

Project Summary:

This project is going to be about using different technologies and Tensorflow Linear classifier as the backend solution for detecting fraud in a credit card transaction.Preambles:

According to Forbes: Approximately 3.7 percent of individuals—or 8.3 million Americans—were victims of some sort of identity fraud within the past year. The average amount stolen is somewhere in the neighborhood of $1,882 per victim, not including time and money spent trying to remedy the identity fraud. In other words, identity thieves are stealing more than $15.5 billion each year.

This project will be focusing on credit card fraud detection using Machine Learning algorithm. The project will use different infrastructure within the context of technologies that has been learned.

About the Project

Machine Learning and its importance can never be referred to as being overrated. In this project, our target is to look into the past trends in our dataset and ensure that we train our data based on this to be able to help us detect/predict a fraudulent transaction when fed with another dataset. I will be studying the data, model it by using the Logistic Regression and Random Forest Algorithm. Since the Data set is highly unbalanced from the source (kaggle.com), I have done some data exploratory analysis to keep balanced and to improve model prediction accuracy level. Hence, the web application is more efficient in solving the problem.

In this project, I have tried to exemplify the moulding of the dataset making use of a machine learning classification, and my focus being the fraud detection in credit card transactions.

Classification is considered to be an instance of supervised learning, i.e. learning where a training set of correctly identified observations is available.

The outcome of our training is then used to classify other dataset fed into the WebApp..

The Dataset:

The Dataset is sourced from Kaggle.com

Context:

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.

Content:

The datasets contains transactions made by credit cards in September 2013 by european cardholders.

This 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.

For this project, I have balanced the data before training the dataset.

Project Infrastructures/ Technologies:

Data Storage:

- MySQL Database

User Interface (UI) Design

- HTML | CSS | Bootstrap
- Javascript libraries (d3.js)

Web Framework

- Python Flask

Machine Learning

- Tensorflow (CNN)
- Logistic Regression
- Random Forest
- Scikit Learn

Other Python Libraries:

- Pandas
- Numpy
- Matplotlib

Cloud Hosting:

- AWS

Thank You