**A Beginner’s guide to Regression based Machine Learning! (Using AML data)**

By: Gurteg Singh

**Prerequisites**

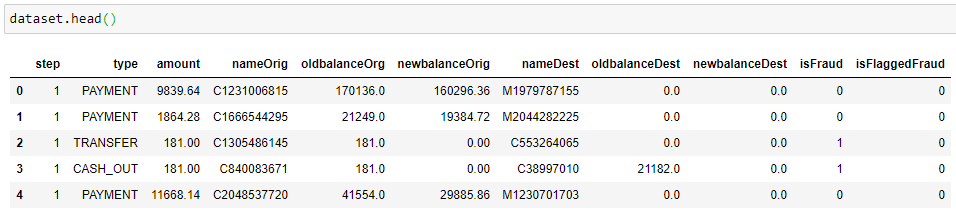
Downloaded data from <https://www.kaggle.com/ntnu-testimon/paysim1>

Downloaded Anaconda with Python 3.8 & launched Jupyter Notebook

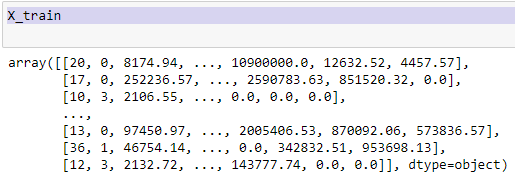
**Visualization**

I am looking to create a Binary model, to predict whether or not a particular transaction is Fraudulent or not.First I imported the libraries NumPy, Pandas, Statsmodel, & Matplotlib, as these were essential to the development of our models, while maintain graphical representation.

Original Data Snapshot



Then, I cleaned through the data, dropping variables that had little to no correlation with my dependent variable.I then assorted numbers to strings in my data.



Following this, I split 70% of my data to train the model and incorporated the remaining 30% to be the testing data. Because my data is binary, I chose a Linear Logit function to predict the cases, due to the Logit functions properties.

I then ran the model, and outputted the Y-intercept, as well as the coefficient of determination (R-squared).

**Results**

Finally I graphed the model as a binary false / true positive graph, and interpreted my results below.

