**Project Proposal**

**Abstract**

Money laundering is the process of taking money obtained from illegal income like drugs, crimes, terrorist activities and make money appears to be earned from legal business. For decades, governments from around the world are trying to defeat the money laundry because of its harmful effect on the economics and because it helps criminals use their dirty money. This study aims to detect the money suspicions transactions, by identifying unusual and repeated patterns

**Dataset**

To achieve the goal of this study the dataset **Synthetic Financial Datasets For Fraud Detection** will be used. This dataset can be found at [Kaggle](https://www.kaggle.com/ealaxi/paysim1).

This dataset contains transactions with the follwing fields:

* step - maps a unit of time in the real world. In this case 1 step is 1 hour of time. Total steps 744 (30 days simulation).
* type - CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER.
* amount -  
  amount of the transaction in local currency.
* nameOrig - customer who started the transaction
* oldbalanceOrg - initial balance before the transaction
* newbalanceOrig - new balance after the transaction
* nameDest - customer who is the recipient of the transaction
* oldbalanceDest - initial balance recipient before the transaction. Note that there is not information for customers that start with M (Merchants).
* newbalanceDest - new balance recipient after the transaction. Note that there is not information for customers that start with M (Merchants).
* isFraud - This is the transactions made by the fraudulent agents inside the simulation. In this specific dataset the fraudulent behavior of the agents aims to profit by taking control or customers accounts and try to empty the funds by transferring to another account and then cashing out of the system.
* isFlaggedFraud - The business model aims to control massive transfers from one account to another and flags illegal attempts. An illegal attempt in this dataset is an attempt to transfer more than 200.000 in a single transaction.

The dataset is available as the .csv file. a sample of data is shown in the following table:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **step** | **type** | **amount** | **nameOrig** | **oldbalanceOrg** | **newbalanceOrig** | **nameDest** | **oldbalanceDest** | **newbalanceDest** | **isFraud** | **isFlaggedFraud** |
| **1** | PAYMENT | 9839.64 | C1231006815 | 170136.0 | 160296.36 | M1979787155 | 0.0 | 0.0 | 0 | 0 |
| **1** | PAYMENT | 1864.28 | C1666544295 | 21249.0 | 19384.72 | M2044282225 | 0.0 | 0.0 | 0 | 0 |

* Dataset is too large, I will use only enough part to get good results
* I believe I can get good results with a Neural Networks (NN) machine learning model. I will fit a model on the transactions dataset to identify the possible fraud transactions

**Tools**

There are tools that will be used to achieve the goal of this study, such as: TensorFlow, matplotlib, pandas for discovering the data and train a model. The work will be done through Jupyter notebook.