

## Case Study: Synthetic Financial Domain: Financial

There is a lack of publicly available datasets on financial services and specially in the emerging mobile money transactions domain. Financial datasets are important for many researchers at performing research in the domain of fraud detection.

### Tasks:

Now with data pipeline ready, you are required to develop the model and predict the fraud using spark streaming.

1. Explore the dataset and develop a model to predict the fraud.
2. Develop the application to train the model and persist the model to disk.
3. Create a new spark streaming application for the same
4. Application will connect to the flume to retrieve the data
5. Load the model
6. Predict the fraud and print the result to the logs
7. Test the application by sending dummy data rows from the consumer

Data: Paysim synthetic dataset of mobile money transactions. Each step represents an hour of simulation. This dataset is scaled down 1/4 of the original dataset which is presented in the paper "**Paysim : A financial mobile money simulator for fraud detection**".

### Solution:

**All Tasks have been performed .**

**Following items have been attached with this submission:**

1. Jupyter Notebook – code , output , explanation of Data exploration , feature engineering etc. are included
2. Flume Configuration with Http Source as a source and Custom sink for pulling Data by Spark Streaming

**1. Input Command used for transfer/streaming of messages was :**

```
curl -X POST -H 'Content-Type: application/json; charset=UTF-8' -d '{"headers" :  
{"a":"b"},"body": {"type":"TRANSFER","amount": 181.0,"oldbalanceOrig": 181  
,"newbalanceOrig": 0.0 , "oldbalanceDest": 0.0 , "newbalanceDest": 0.0}}'   
http://localhost:9000
```

**2. Flume agent was started :**

```
./flume-ng agent --conf conf --conf-file /mnt/home/edureka_960126/flume http.conf --name  
httpagent -Dflume.root.logger=INFO,console
```

**3. Output was logged onto console :**

```
type|amount|prediction|  
+---+-----+-----+  
+---+-----+-----+  
  
-----  
Time: 2020-07-21 22:51:00  
-----  
1  
  
+---+-----+-----+  
|   type|amount|prediction|  
+---+-----+-----+  
|TRANSFER| 181.0|      1.0|  
+---+-----+-----+  
  
-----  
Time: 2020-07-21 22:51:15  
-----
```

**Architecture of this application :**

**HTTP Source ( streaming JSON req.)-> Flume Agent -> Spark Streaming(Pull Based)-> Prediction Output**