Registration complete. You are now logged in.

Dataset

492

Synthetic Financial Datasets For Fraud Detection

Synthetic datasets generated by the PaySim mobile money simulator

TESTIMON @ NTNU • updated 2 years ago (Version 2)

Usability 8.8 **License** CC BY-SA 4.0 **Tags** computing, finance, crime

Description

Context

There is a lack of public available datasets on financial services and specially in the emerging mobile money transactions domain. Financial datasets are important to many researchers and in particular to us performing research in the domain of fraud detection. Part of the problem is the intrinsically private nature of financial transactions, that leads to no publicly available datasets.

We present a synthetic dataset generated using the simulator called PaySim as an approach to such a problem. PaySim uses aggregated data from the private dataset to generate a synthetic dataset that resembles the normal operation of transactions and injects malicious behaviour to later evaluate the performance of fraud detection methods.

Content

PaySim simulates mobile money transactions based on a sample of real transactions extracted from one month of financial logs from a mobile money service implemented in an African country. The original logs were provided by a multinational company, who is the provider of the mobile financial service which is currently running in more than 14 countries all around the world.

This synthetic dataset is scaled down 1/4 of the original dataset and it is created just for Kaggle.

Headers

This is a sample of 1 row with headers explanation:

1,PAYMENT,1060.31,C429214117,1089.0,28.69,M1591654462,0.0,0.0,0,0

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

(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.

Past Research

There are 5 similar files that contain the run of 5 different scenarios. These files are better explained at my PhD thesis chapter 7 (PhD Thesis Available here http://urn.kb.se/resolve?urn=urn:nbn:se:bth-12932).

We ran PaySim several times using random seeds for 744 steps, representing each hour of one month of real time, which matches the original logs. Each run took around 45 minutes on an i7 intel processor with 16GB of RAM. The final result of a run contains approximately 24 million of financial records divided into the 5 types of categories: CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER.

Acknowledgements

This work is part of the research project "Scalable resource-efficient systems for big data analytics" funded by the Knowledge Foundation (grant: 20140032) in Sweden.

Please refer to this dataset using the following citations:

PaySim first paper of the simulator:

E. A. Lopez-Rojas , A. Elmir, and S. Axelsson. "PaySim: A financial mobile money simulator for fraud detection". In: The 28th European Modeling and Simulation Symposium-EMSS, Larnaca, Cyprus. 2016

Data (182 MB)

Data Sources

PS_20174392719_1... 11 columns

About this file

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".

Columns

step Maps a unit of time in the real world. In this case 1 step is 1 hour of time.

A type CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER

amount amount of the transaction in local currency

A nameOrig customer who started the transaction

oldbalanceOrg initial balance before the transaction

newbalanceOrig customer's balance after the transaction.

A nameDest recipient ID of the transaction.

oldbalanceDest initial recipient

Registration complete. You are now logged in.

isFraud identifies a fraudulent transaction (1) and non fraudulent (0)

isFlaggedFraud flags illegal attempts to transfer more than 200.000 in a single transaction.

PS_20174392719_1491204439457_log.csv (182 MB) 11 of 11 columns Views

	# step Maps a unit of time in the real world. In this case 1 step is 1 hour of time.	A type CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER	# amount amount of the transaction in local currency	A nameOrig customer who started the transaction	# oldbalanceOrg initial balance before the transaction	# newb custome the trans
1	1	PAYMENT	9839.64	C1231006815	170136.0	
2	1	PAYMENT	1864.28	C1666544295	21249.0	
3	1	TRANSFER	181.0	C1305486145	181.0	
4	1	CASH_OUT	181.0	C840083671	181.0	
5	1	PAYMENT	11668.14	C2048537720	41554.0	
6	1	PAYMENT	7817.71	C90045638	53860.0	
7	1	PAYMENT	7107.77	C154988899	183195.0	
8	1	PAYMENT	7861.64	C1912850431	176087.23	
9	1	PAYMENT	4024.36	C1265012928	2671.0	
10	1	DEBIT	5337.77	C712410124	41720.0	
11	1	DEBIT	9644.94	C1900366749	4465.0	
12	1	PAYMENT	3099.97	C249177573	20771.0	
13	1	PAYMENT	2560.74	C1648232591	5070.0	
14	1	PAYMENT	11633.76	C1716932897	10127.0	
15	1	PAYMENT	4098.78	C1026483832	503264.0	
16	1	CASH_OUT	229133.94	C905080434	15325.0	
17	1	PAYMENT	1563.82	C761750706	450.0	
18	1	PAYMENT	1157.86	C1237762639	21156.0	
19	1	PAYMENT	671.64	C2033524545	15123.0	
20	1	TRANSFER	215310.3	C1670993182	705.0	
21	1	PAYMENT	1373.43	C20804602	13854.0	
22	1	DEBIT	9302.79	C1566511282	11299.0	
23	1	DEBIT	1065.41	C1959239586	1817.0	
24	1	PAYMENT	3876.41	C504336483	67852.0	
25	1	TRANSFER	311685.89	C1984094095	10835.0	
26	1	PAYMENT	6061.13	C1043358826	443.0	
27	1	PAYMENT	9478.39	C1671590089	116494.0	
28	1	PAYMENT	8009.09	C1053967012	10968.0	

Similar Datasets


Credit Card Fraud

Google Play Store Apps


TMDB 5000 Movie

Global Terrorism


Bitcoin Historical Data

Registration complete. You are now logged in.

156,254 views · 21,952 downloads · 106 kernels · 16 topics

View more activity