Estimating financial fraud through transaction-level features and machine learning

Ayed Alwadain [1], Rao Faizan Ali [2], Amgad Muneer [3*]

[1]Computer Science Department, Community College, King Saud University, Riyadh 145111, Saudi Arabia

[2]Department of Software Engineering, School of Systems and Technology, University of Management and Technology, Lahore 54400, Pakistan

[3]Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, USA

*Correspondence: muneeramgad@gmail.com;

Table S2: Details of variables

Variable	**Explanation**
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.