Transaction Fraud Detection

In this task you are predicting the probability that an online customer transaction is fraudulent, as denoted by the binary target isFraud(=1 or 0).

The data is broken into two files .identity and .transaction, which are joined by TransactionID. Not all transactions have corresponding identity information.

Categorical Features - Transaction

* ProductCD
* card1 - card6
* addr1, addr2
* P\_emaildomain
* R\_emaildomain
* M1 - M9

Categorical Features - Identity

* DeviceType
* DeviceInfo
* id\_12 - id\_38

The TransactionDT feature is a timedelta from a given reference datetime (not an actual timestamp).

Files

* train\_{transaction, identity}.csv - the training set

some data are missing in these files, you should be able to handle or decide how to drop missing data for your better modeling

Note: some features (columns) are hidden or anonymous because of security reason.