

Fraud Transaction Detection Using Transactions Dataset

Objective

Build a system that can classify if a transaction is fraudulent or not.

Dataset

This dataset is a simulated dataset of original and fraudulent transactions, and only has the main core details of a transaction.

The simulated frauds are using the following scenarios:

1. Any transaction whose amount is more than 220 is a fraud. This is not inspired by a real-world scenario. It provides an obvious fraud pattern that should be detected by any baseline fraud detector. This will be useful to validate the implementation of a fraud detection technique.
2. Every day, a list of two terminals is drawn at random. All transactions on these terminals in the next 28 days will be marked as fraudulent. This scenario simulates a criminal use of a terminal, through phishing for example. You could add features that keep track of the number of fraudulent transactions on the terminal to help with this scenario.
3. Every day, a list of 3 customers is drawn at random. In the next 14 days, 1/3 of their transactions have their amounts multiplied by 5 and marked as fraudulent. This scenario simulates a card-not-present fraud where the credentials of a customer have been leaked. The customer continues to make transactions, and transactions of higher values are made by the fraudster who tries to maximize their gains. You could add features that keep track of the spending habits of the customer for this scenario.

The Fraud labels for the transactions have been simulated with the above conditions. Please take note of this as it will help you evaluate and diagnose your performance.

Description of main columns:

- TRANSACTION_ID: A unique identifier for the transaction
- TX_DATETIME: Date and time at which the transaction occurs
- CUSTOMER_ID: The identifier for the customer. Each customer has a unique identifier
- TERMINAL_ID: The identifier for the merchant (or more precisely the terminal). Each terminal has a unique identifier
- TX_AMOUNT: The amount of the transaction.
- TX_FRAUD: A binary variable, with the value 0 for a legitimate transaction, 1 for a fraudulent transaction.