Version: 1.0.0

Novelis Code Assessment

High-level problem statement

E-commerce websites often transact huge amounts of money. Whenever a huge amount of money is moved, there is a high risk of users performing fraudulent activities, e.g. using stolen credit cards, laundering money, etc.

Objective

The goal of this challenge is to build a machine learning model that predicts the probability that the first transaction of a new user is fraudulent.

Details

Electronica is an e-commerce site that sells wholesale electronics. You have been contracted to build a model that predicts whether a given transaction is fraudulent or not. You only have information about each user's first transaction on Electronica's website. If you fail to identify a fraudulent transaction, Electronica loses money *equivalent* to the price of the fraudulently purchased product. If you incorrectly flag a real transaction as fraudulent, it inconveniences the Electronica customers whose valid transactions are flagged—a cost your client values at $8.

Python Environment with Package Versions

Please use a Python 3.7+ environment with the following packages:

pandas==1.2.4
scikit-learn==0.24.2
numpy==1.17.4
matplotlib==3.1.1

```
"Fraud_Data" - information about each user first transaction
```

Columns:

- **user_id** : Id of the user. Unique by user
- **signup_time** : the time when the user created her account (GMT time)
- **purchase_time** : the time when the user bought the item (GMT time)
- **purchase_value** : the cost of the item purchased (USD)
- **device_id** : the device id. You can assume that it is unique by device. I.e., 2 transactions with the same device ID means that the same physical device was used to buy
- **source** : user marketing channel: ads, SEO, Direct (i.e. came to the site by directly typing the site address on the browser).
- **browser** : the browser used by the user.
- **sex** : user sex: Male/Female
- **age** : user age
- **ip_address** : user numeric ip address
- **class** : this is what we are trying to predict: whether the activity was fraudulent (1) or not (0).

- **country** : country of IP address