

In this work, we are going to classify whether an online transaction is fraudulent or not. It is a binary target, named: isFraud. A lot of us have been in a situation where a transaction has been canceled without our consideration. To have a better algorithm that only cancels the fraudulent transaction and doesn't just result in being embarrassed in a store. While it is often embarrassing at the moment, but this system is truly sparing consumers millions of dollars every year.

Most of the transactions are not fraud so in this case, we have unbalanced data. And there are 385 columns in this dataset with missing values. There are 41 columns with missing percent values than 90%. Machine learning algorithms do not work properly with an unbalanced dataset. Only 3.5% are fraudulent.

They are several methods to deal with an unbalanced dataset. Fortunately, Lightgbm by default thinks that the dataset is balanced. So we would need to change to the parameter as:

```
Is_unbalance = True
```

By doing this, the “unbalance” mode uses the values of the target to automatically adjust weights inversely proportional to class frequencies in the input data.

And lightgbm is a fast, distributed, high-performance gradient boosting framework based on decision tree algorithm.

Random Forest is a tree-based machine learning algorithm that leverages the power of multiple decision trees for making decisions. It combines the output of individual decision trees to generate the final output.

So, I ensemble the result of two model to improve the overall performance.