The objective is to identify fraudulent credit card transactions so that credit card firms' consumers are not charged for products they did not purchase.

**Main challenges involved in credit card fraud detection are:**

Every day, massive amounts of data are gathered, and the model must be fast enough to respond to the fraud in time.

Imbalanced data, i.e. the majority of transactions (99.8%) are not fraudulent, making it extremely difficult to discover the fraudulent ones. Data availability, as the data is generally private.

Misclassified Data is another significant concern, as not every fraudulent transaction is detected and reported.

Scammers utilised adaptive tactics against the model.

**How to tackle these challenges?**

The model employed must be simple and fast enough to detect the abnormality and label the transaction as fraudulent as soon as feasible.

Imbalance may be dealt with appropriately by employing several approaches, which we shall discuss in the next paragraph.

The dimensionality of the data can be lowered to safeguard the user's privacy.

A more reliable source must be used to double-check the data, at the very least for training the model.

We can simplify and understand the model such that when the scammer adjusts to it, we can have a new model ready to deploy with just a few modifications.

**About the data**

The data for this article can be found [here](https://www.kaggle.com/mlg-ulb/creditcardfraud). This dataset contains the real bank transactions made by European cardholders in the year 2013. As a security concern, the actual variables are not being shared but — they have been transformed versions of PCA. As a result, we can find 29 feature columns and 1 final class column.