Data Science methodology to credit cards

Using the topic that you selected, complete the Business Understanding stage by coming up with a problem that you would like to solve and phrasing it in the form of a question that you will use data to answer.

You are required to:

1. Describe the problem, related to the topic you selected.

As we are moving towards the digital world — cybersecurity is becoming a crucial part of our life. When we talk about security in digital life then the main challenge is to find the abnormal activity.

When we make any transaction while purchasing any product online — a good amount of people prefers credit cards. The credit limit in credit cards sometimes helps us me making purchases even if we don't have the amount at that time. but, on the other hand, these features are misused by cyber attackers.

To tackle this problem, we need a system that can abort the transaction if it finds fishy.

Here, comes the need for a system that can track the pattern of all the transactions, and if any pattern is abnormal then the transaction should be aborted.

Today, we have many machine learning algorithms that can help us classify abnormal transactions. The only requirement is the past data and the suitable algorithm that can fit our data in a better form.

2. Phrase the problem as a question to be answered using data.

The main challenges involved in credit card fraud detection are:

1. Enormous Data is processed every day and the model build must be fast enough to respond to the scam in time.

2. Imbalanced Data i.e most of the transactions *(99.8%)* are not fraudulent which makes it really hard for detecting the fraudulent ones

3. Data availability as the data is mostly private.

4. Misclassified Data can be another major issue, as not every fraudulent transaction is caught and reported.

5. Adaptive techniques used against the model by the scammers.

How to tackle these challenges?

1. The model used must be simple and fast enough to detect the anomaly and classify it as a fraudulent transaction as quickly as possible.

2. Imbalance can be dealt with by properly using some methods which we will talk about in the next paragraph

3. For protecting the privacy of the user the dimensionality of the data can be reduced.

4. A more trustworthy source must be taken that double-checks the data, at least for training the model.

5. We can make the model simple and interpretable so that when the scammer adapts to it with just some tweaks we can have a new model up and running to deploy

AI Fraud Detection System Implementation Steps:

- Data Mining. Implies classifying, grouping, and segmenting of data to search millions of transactions to find patterns and detect fraud.

- Pattern Recognition. Implies detecting the classes, clusters, and patterns of suspicious behavior. Machine Learning here represents the choice of a model/set of models that best fit a certain business problem. For example, the neural networks approach helps automatically identify the characteristics most often found in fraudulent transactions; this method is most effective if you have a lot of transaction samples.

For example, using the food recipes use case discussed in the labs, the question that we defined was, "Can we automatically determine the cuisine of a given dish based on its ingredients?".