**Credit card fraud detection**

Dataset link: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud/

**Explanation**

This project is structured as below:-

1. Data Acquisition and Understanding

In this initial phase, our objective is to load the dataset and gain a comprehensive understanding of its features. This process allows us to identify the pertinent attributes that will contribute to our final model.

1. Exploratory Data Analysis (EDA)

During EDA, we typically conduct univariate and bivariate analyses of the dataset, with an option for feature transformations if required. While the data primarily consists of Gaussian variables, Z-scaling may not be necessary. Nevertheless, it's essential to assess data skewness and address it if present, as it could pose challenges in subsequent model development.

1. Train-Test Split

Once we become familiar with the dataset, we divide it into training and testing sets to see how our models perform on data they have not seen before. The k-fold cross-validation method is helpful for checking if something is valid. We just need to choose a good k value to make sure we include enough of the minority group in the tests.

1. Model Development

In this important stage, we are trying out different models and adjusting their settings to make sure they perform well on our dataset. We should try different ways of choosing our data to make our model results better.

1. Model Evaluation

Evaluation metrics are applied to assess model performance. Given the dataset's imbalanced nature, the accurate identification of fraudulent transactions takes precedence over non-fraudulent ones. Therefore, selecting an evaluation metric aligned with this business objective is imperative.