

## **Summary**

1. The first step is to import the libraries that will be used to run and load the dataset.
2. The data is then loaded and viewed. As an example, the data kinds and features information. The standard deviation and mean.
3. Then we do Exploratory Data Analysis (EDA) to get the “feel” of the data, we will be checking the distributions, the correlations etc of the different columns.
  - 3.1 We use seaborn to create a simple heatmap to see where we are missing data!
  - 3.2 We'll look at the data with our OUTCOME in mind in this part.
4. In this section we preprocess the data with data cleaning, slitting the data, feature scaling, and Sampling.
  - 4.1 Data cleaning refers to identifying and correcting errors in the dataset that may negatively impact a predictive model. SO we clean the data that contained Null information.
  - 4.2 Data splitting is commonly used in machine learning to split data into a train, test, or validation set. This approach allows us to find the model hyper-parameter and also estimate the generalization performance. We divided the data into two groups of 80/20. Training set for about 80% of the total. The remaining 20% will be used to test the data.
  - 4.3 Feature Scaling is a technique to standardize the independent features present in the data in a fixed range. It is performed during the data pre-processing to handle highly varying magnitudes or values or units. We scale the feature with StandardScaler().
  - 4.4 As our target data OUTCOME is imbalanced, we use oversampling method for balancing the data. We used SMOTE for sampling. SMOTE is an oversampling technique where the synthetic samples are generated for the minority class.
5. We tuned the HyperParameter of many machine learning models in this part to find the ideal parameter for the model. We used GridSearch for tuning the HyperParameter. GridSearchCV is a technique to search through the best parameter values from the given set of the grid of parameters. After that, we fit our model with the best parameter. Then we compare our model.