# Pre-processing

#### Transforming raw data into an understandable format. There involves two steps which are na value checking, label encoding and feature scaling.

# 1.Check N/A Value

### The heatmap below represents all data between four columns whether empty(N/A value) or not, the scale of value on heatmap is from 0(pink red color) to 1(black color) which mean False and True for empty(N/A value) data.

# There has no N/A values since the whole image is pink red.

# 2.Label Encoding and Feature Scaling

#### Label Encoding refers to converting the labels into numeric form so as to convert it into the machine-readable form.

#### Feature scaling is a method used to normalize the range of independent variables or features of data since gradient descent converges much faster with feature scaling than without it. If we compare measurements that have different units. Variables that are measured at different scales do not contribute equally to the analysis and might end up creating a bias. Transforming the data to comparable scales can prevent this problem.

# Train/Test Split

### The data we use is usually split into training data and test data. The training set contains a known output and the model learns on this data in order to be generalized to other data later on. We have the test dataset (or subset) in order to test our model’s prediction on this subset.

# Modelling

#### The process of modelling in machine learning(ML) involves providing an ML algorithm with training data to learn from. The term ML model refers to the model artifact that is created by the training process.

# 2. One Hot Encoding and Feature Scaling

#### One Hot Encoding are used for categorical variables where no such ordinal relationship exists, there are N categories and therefore N binary variables are needed.

#### Feature scaling is a method used to normalize the range of independent variables or features of data since gradient descent converges much faster with feature scaling than without it. If we compare measurements that have different units. Variables that are measured at different scales do not contribute equally to the analysis and might end up creating a bias. Transforming the data to comparable scales can prevent this problem.

# Evaluation

#### Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

#### Recall is the ratio of correctly predicted positive observations to the all observations in actual class.

#### F1 Score is the weighted average of Precision and Recall.