Week08: Regression Classification Evaluation

Given the dataset pid, the column label represents the class label of the record. This week, you need to apply three different classification techniques to predict the label of each record.

1. Split the dataset into training and testing (75/25 %).

2. Train the model using the training data.

3. Use the model to predict the label of the test data.

4. Compute and display the confusion matrix.

5. Evaluate each classifier using two evaluation metrics (e.g. precision and accuracy)

6. In columns C, D, E and F, the value 0 is not correct, remove those values and use one of the data imputation techniques that you learnt last week to fill the missing values that resulted from removing the value 0.

7. Use sampling with stratification to split the data such that each class gets represented by a ratio that reflects the ratio of the records that belong to the class in the dataset.

8. Check the performance of the model after the imputation.