ASSIGNMENT 2

Name: 林柏均 Student ID: s107062240   05/01/2021

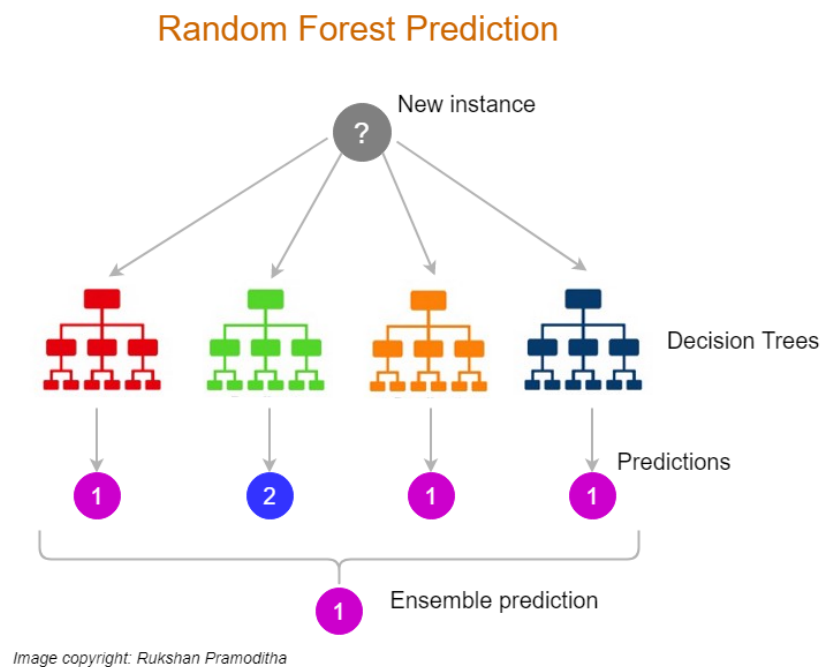## 1   Try to give more explanation and discussion for your assignment

Figure 1: Image of Random Forest Classifier

In this assignment, I choose Random Forest Classifier as my training model. And my Random Forest is constructed with 100 Decision trees. In the data preprocessing section, I combined all the files in NORMAL folder and PNEUMONIA folder as my training set and shuffled the collection so that it can improve the training quality. I also resized the chest scan image from different size to (512, 512) as my training input. And validation set is also combined from two validation folders.

2   List out what package you had used

- PIL: This package allows us to read out the image and convert it to array format

- joblib: This can help us save the Sklearn model

- alivebar: This can show the current loading or processing progress situation

- sklearn: Import Random Forest Classifier and other accuracy indicators

3 Training Result

3.1 Part 1

Below is the first chest scan image in the training NORMAL folder:

- sample image size: (2234, 2359)

- sample image min: 0

- sample image max: 255

- sample image mean: 124.39103010508906

- processed image size: (512, 512)

- processed image min: 0.0

- processed image max: 1.0

- processed image mean: 0.48801540299957885

3.2 Part 2

Random Forest Classifier Parameters:

- Decision Tree Count: 200

- Input array size: (512, 512)

- Random state: 10

  Random Forest Classifier Result:

- Accuracy: 0.6875

- Recall: 0.6875

- Precision: 0.808

4 Summary about what I have learned

From this assignment, I can review what I learned from the Machine Learning course content, and I can understand more about how Random Forest Classifier can do. In this assignment, the format of dataset is quite complex because of the different size of the image, therefore, it made me understand that data preprocessing is an essential process for Machine Learning.