**Assignment 2 Report**

1. The dataset is split in a ratio of 70:30 between training and testing. 10,100 and 1000 number of component trees are tested to see the resulting change in classification performance. With a greater number of component trees, the accuracy of the model will be higher and stable. However, the running time of constructing the trees is longer too.
2. 10 component trees of the random forest model give the best classification performance which is around 83% accuracy. In those component trees, the first component tree gives the highest accuracy performance which is around 81%. After this, the eighth and tenth component trees gives the second highest accuracy performance which is around 79%. All these component trees’ classification performance is worser than the original random forest model.
3. For selected component trees, their relative importance of the attributes is different. For the first tree, the most important attribute is ‘grade of spondylolisthesis’ which contains around 0.56 importance value. Then the importance value order in descending order is ‘lumbar lordosis angle’, ‘pelvic radius’, ‘sacral slope’, ‘pelvic tilt’ and ‘pelvic incidence’. For the eighth tree, the most important attribute is ‘grade of spondylolisthesis’ which contains around 0.66 importance value. Then the importance value order in descending order is ‘pelvic radius’, ‘lumbar lordosis angle’, ‘pelvic incidence’, ‘pelvic tilt’ and ‘sacral slope’. For the tenth tree, the most important attribute is ‘lumbar lordosis angle’ which contains around 0.4 importance value. Then the importance value order in descending order is ‘grade of spondylolisthesis’, ‘pelvic radius’, ‘sacral slope’, ‘pelvic incidence’, and ‘pelvic tilt’.
4. The accuracy of the naïve Bayes classifier model is around 80% which is lower than the random forest model’s 83% accuracy. The precision of the ‘NO’ class is 100%. However, for the ‘DH’ and ‘SL’ class, the precisions are only 55% and 66% respectively.