Option II - A: Tree and Ensemble Learning

This option will feature components from decision trees, random forests, and ensemble learning.

Use the **multiclass** Abalone dataset from Option I, you need to apply CART for the sample problem and report the classification performance on the train and test set using the same train/test split. Abalone dataset from UCI ML repository: <https://archive.ics.uci.edu/ml/datasets/Abalone>.

**Part A**

1. Provide visualisation and analysis of your Abalone multi-class data (similar to first part of Assessment 2 )
2. Create a Decision Tree for the Abalone multi-class data and report train and test performance for multiple experimental (**can be 5 or more**) runs using different hyperparameters - i.e **tree depth or any other hyperparameter of your choice**. Take the best Tree and report the Tree Visualisation (show your tree and also translate few selected nodes and leaves into IF and Then rules): Note: Since Decision Trees give the same results for the same dataset, ensure that in different experimental runs, you create different set of train/test split as done in Week1 and Week 2 Exercise solutions.
3. Do an investigation about improving performance further by either pre-pruning or post-pruning the tree: <https://scikit-learn.org/stable/auto_examples/tree/plot_cost_complexity_pruning.html>
4. Apply either Random Forests and show performance (eg. accuracy score) as your number of trees in the ensembles increases.
5. Further, compare your results with XGBoost and Gradient Boosting and provide a discussion.
6. Compare results with **Adam/SGD (Simple Neural Networks)** and discuss them. You can use default hyper-parameters from the sklearn library - there is no need for extensive hyperparameter search as needed in Opton I.

Note that performance refers to accuracy which could be either classification accuracy, AUC or F1 score. You can report all or selected type of scores.

**Part B:** Apply the above steps for the Abalone regression problem. Compare results with linear regression model from earlier Assessment.

**Part C:** Apply the above steps to another dataset of your choice