Module 5 - Bonus Challenges

§M5: Decision Trees

- Below are open-ended bonus challenges; solving them is not required but can help you better understand ML/AI in the context of engineering, and how to use them in practical cases.
- Bonus points earned in all homework assignments will be averaged (5 bonus points for each assignment) and then directly added to your final score to calculate your final letter grade.

Challenge 1.1. For this bonus question, you will explore one of the unique advantages of Random Forests: their ability to provide insights into feature importance. There are two commonly used quantitative measures for feature importance in Random Forests: Mean Decrease in Impurity (MDI) and Permutation Feature Importance (PFI). MDI is a built-in measure for random forest models in Scikit-learn, while PFI is a general method applicable to various models. Please review these two measures and how to use them in the `Scikit-learn documentation`. Then, complete the following tasks: (*5pts*)

1. Fit a Random Forest model with 5 trees on Scikit-learn's Iris dataset.Randomly split the dataset into 80% for training and 20% for testing, using the built-in function `train_test_split` in `sklearn`.Train your model on the training dataset and evaluate it on the testing set.
2. Extract MDI and PFI of the trained model, visualize them using the bar plot, and interpret the values. Your bor plot may look like that in Figure 1.
3. Adding an extra random feature with normal distribution to the original dataset. Train another random forest and extract its MDI and PFI values.
4. Adding an extra constant feature (all-ones) to the original dataset. Train another random forest and extract its MDI and PFI values.
5. Discuss the results you obtain, such as whether MDI and PFI provide the same ranking of feature importance, how random and constant variables affect the results for these two measures, and why these changes occur.

6. Submit your Jupyter notebook file with necessary comments.

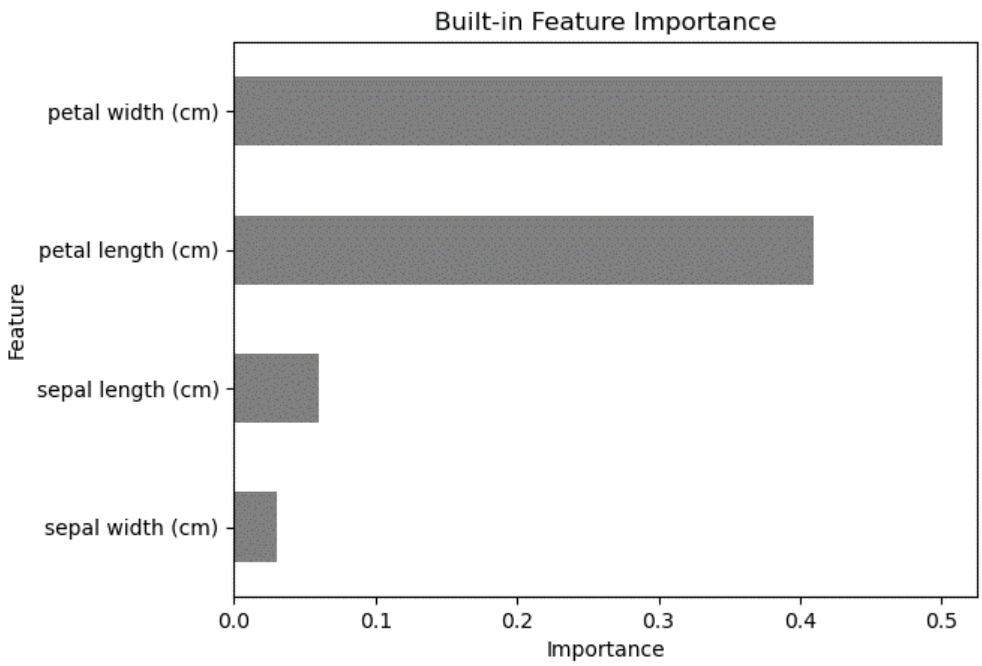

Figure 1: Bar plot of the extracted MDI