Question 1

Provide a short and straight to the point comparison of Decision Tree and Random Forest Algorithms. Do so by addressing the following points:

1. **Derive the splitting criterion for a Decision Tree:** Consider a classification problem where the goal is to split the dataset based on features. Derive the Gini Impurity and show how it can be used to choose the best split for a decision tree.
2. **Explain how Random Forest improves over a single Decision Tree:** Discuss the concept of bootstrapping and feature bagging, and explain why Random Forests generally perform better than a single Decision Tree in terms of variance and overfitting.

Hints: The Gini Impurity for a node with classes $C_1, C_2, ..., C_k$ is defined as:

$$\text{Gini}(p) = 1 - \sum_{i=1}^{k} p_i^2$$

where p_i is the proportion of observations of class C_i in that node.

Question 2

Apply Decision Trees, Random Forests, and Support Vector Machines (SVM) to the **Iris Dataset**, a classic dataset for classification tasks, which is available via `sklearn.datasets`.

1. **Download the dataset:** Load the Iris Dataset from the `sklearn` library.
2. **Apply Decision Tree, Random Forest, and SVM:**
    - Split the dataset into a training set and a test set.
    - Train a Decision Tree classifier, a Random Forest classifier, and an SVM classifier on the dataset.
    - Use cross-validation to tune the hyperparameters for each model (e.g., max depth for Decision Trees, number of trees for Random Forests, and kernel for SVM).
3. **Evaluate the Models:**
    - Report the accuracy, precision, recall, and F1-score for each model on both the training and test sets.
    - Compare the results and explain the differences in performance between the models.