

| **Title:**  **Implementation of support vector machine** |
| --- |

**Objective:** To create a model for classification using support vector machine.

**Expected Outcome of Experiment:**

| **CO** | **Outcome** |
| --- | --- |
| **CO1** | Describe and apply supervised learning methods |

**Books/ Journals/ Websites referred:**

<https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm/>

**Theory of decision tree algorithm:**

Decision tree algorithms are a popular approach in machine learning for both classification and regression tasks. They model decisions as a tree-like structure, where each internal node represents a "test" on an attribute (e.g., whether a feature is less than a certain value), each branch corresponds to an outcome of the test, and each leaf node represents a class label or a continuous value.

1. **Splitting Criteria**: The decision tree starts with the entire dataset at the root node. It recursively splits the dataset into subsets based on the values of the attributes. The splitting is done using a metric such as Gini impurity or entropy (information gain) to maximize the homogeneity of the subsets with respect to the target variable.

2. **Tree Growing**: The tree continues to grow by splitting nodes until a stopping criterion is met. This criterion could be a maximum tree depth, minimum number of samples required to split a node, or other parameters to prevent overfitting.

3. **Pruning**: After the tree is fully grown, pruning techniques may be applied to reduce its size and complexity. This helps prevent overfitting by removing parts of the tree that are not statistically significant.

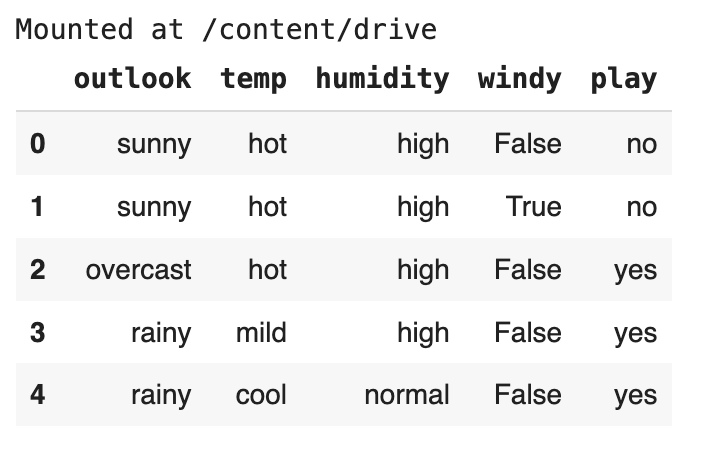
4. **Prediction**: To classify a new instance, the decision tree traverses from the root to a leaf node based on the attribute values of the instance, and assigns the majority class of the instances in that leaf as the predicted class.

5. **Regression**: For regression tasks, decision trees work similarly but predict a continuous value at each leaf node, typically the mean or median of the target variable for the instances in that leaf.

**Details of data set used:**

Tennis dataset

<https://www.kaggle.com/datasets/shreyakatukuri/tennis>

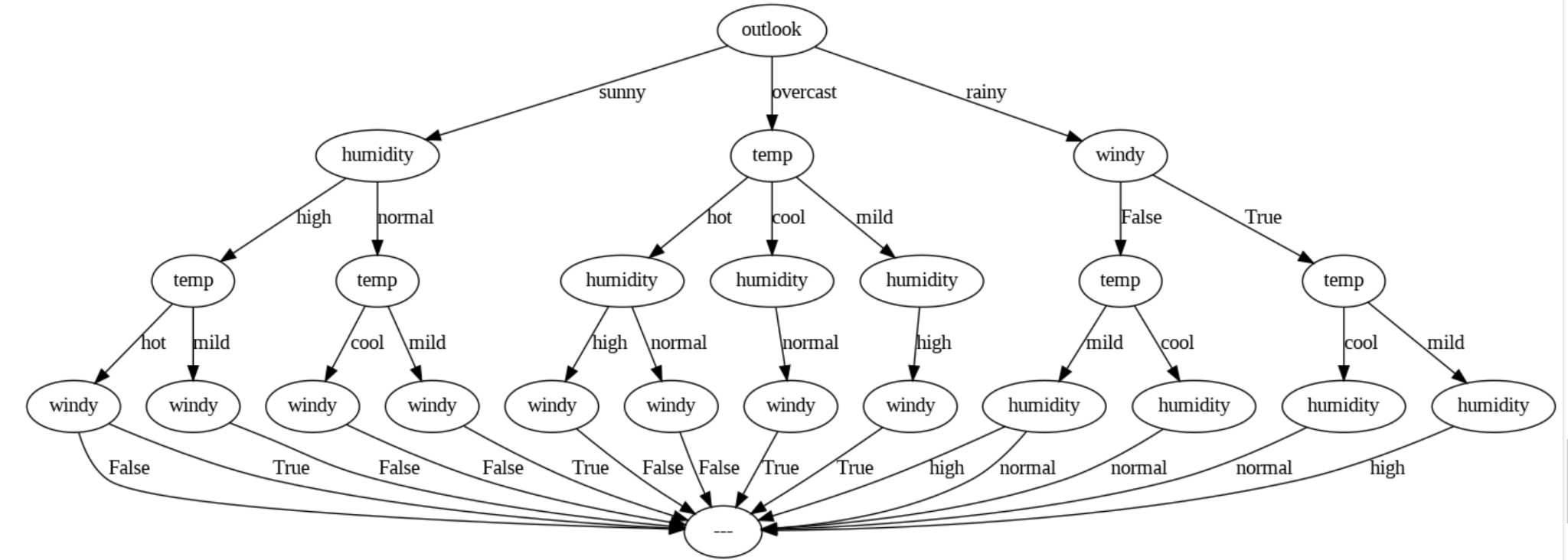


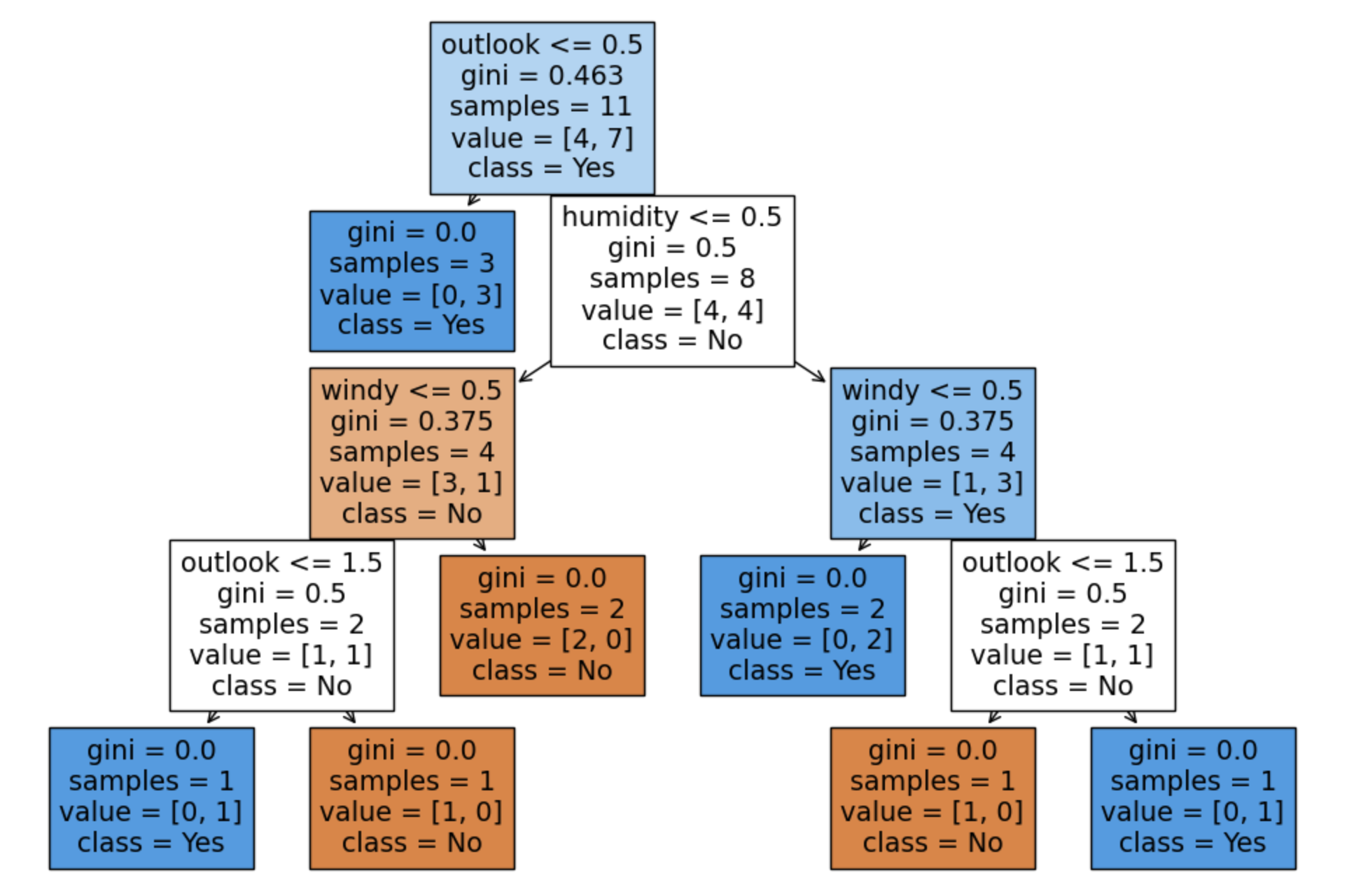
**Explanation of API/Tool used for implementation:**

Entropy based decision tree - didnt use any library

Gini indexing - used Sklearn

**Results:**





**Conclusion:** Thus we have implemented Decision trees in python. We have used entropy method as well as gini indexing for the Decision trees. Decision trees have several advantages, including their interpretability, ability to handle both numerical and categorical data, and ease of visualization. However, they can be prone to overfitting, especially with complex datasets. Ensemble methods like Random Forests and Gradient Boosting are often used to improve the performance of decision trees.