**Report for ML Project [Supervised]**

**Team Members:**

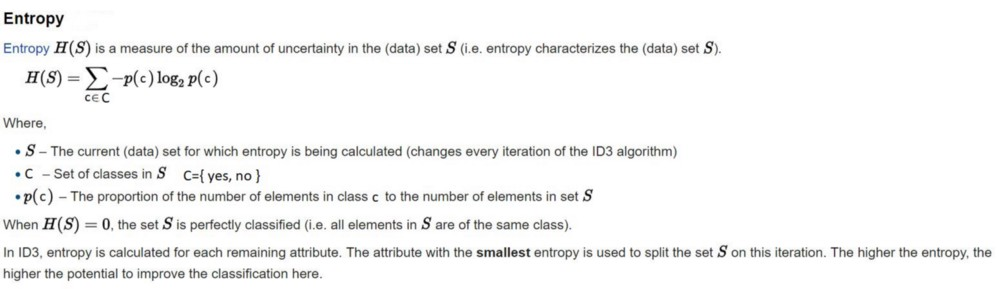
Reshma MR (1MS16IS070)

Rishabh Agarwal (1MS16IS071)

Disha Satheesh (1MS15IS033)

**Decision-Tree Algorithm**

1. Compute the entropy for data-set:



2) For every attribute/feature:

1. Calculate entropy for all categorical values

2. Take average information entropy for the current attribute

3. Calculate gain for the current attribute

3. Pick the highest gain attribute.

4. Repeat until we get the tree we desired.

**About Dataset:**

IRIS dataset is a dataset formed in 1936 and since then has been used for many academic research papers. The target value is divided into 3 class, namely 3 species types in which the flowers are to be classified. We chose this particular dataset for running our algorithm since it is a well-documented dataset, and also has become a standard for Machine Learning beginners.

The dataset has five attributes namely sepal length, sepal width, petal length, petal width and the species of the iris flower. It is a categorical value. The three categories of a species are

Iris-setosa, Iris-versicolor, Iris-virginica.

Given the dataset using decision trees we are trying to predict the class of the Iris species given the four attributes: sepal length, sepal width, petal length, petal width.

**OBSERVATIONS:**

1. The accuracy of the test dataset can change according to the max depth of the tree.

2. If the max depth is too less than the model could become underfit model.

3. If the max depth is too high then the model could become overfit model.

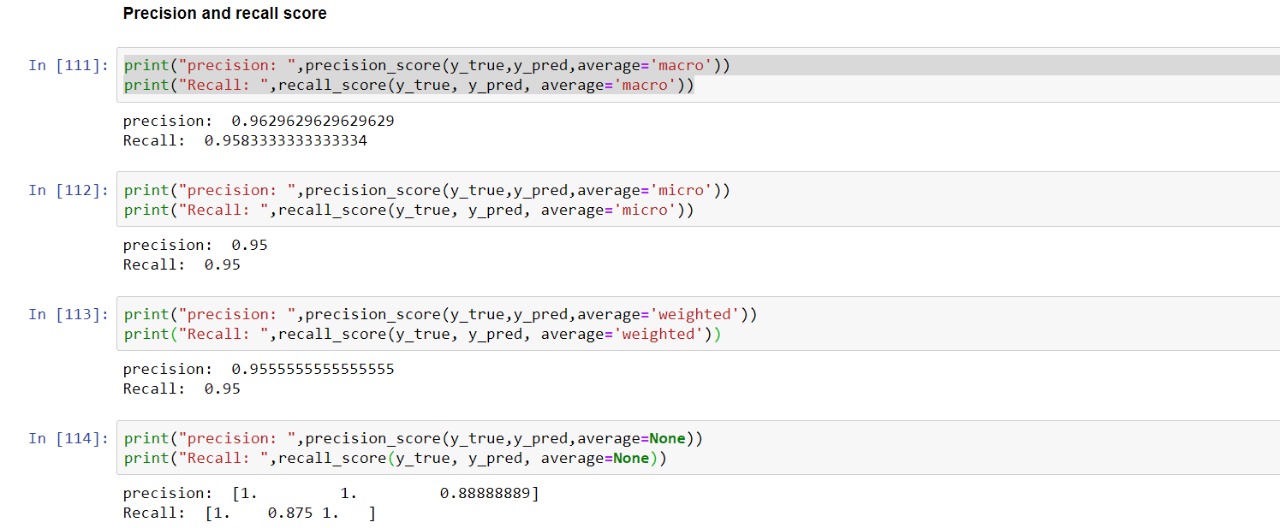
4. When the max depth is 5, the accuracy of the test dataset is about 95%.

5. When the max depth is 1, the accuracy of the test dataset is nearly 60%. From this we can say that the model is underfit. The model doesn’t perform well when exposed to new data.

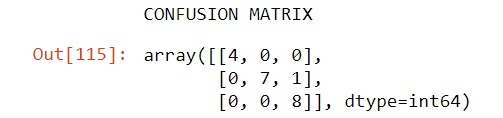
6. When the max depth is 10, the accuracy of the test dataset is about 90%. From this we can tell the model doesn’t perform well if we go in depth. We have to find the right value for max depth. Note that 5<10 and still when max depth = 5, the model gives accurate results compared to when max depth = 10.

**Result:**

**Precision and Recall:**

****

**Confusion Matrix:**



**Accuracy:**

