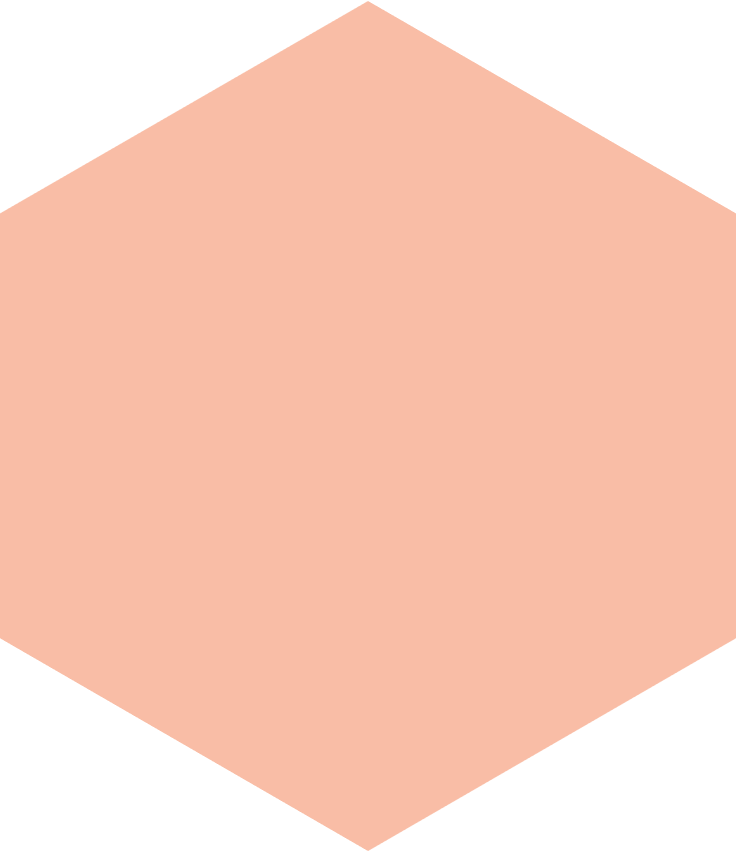


|  |
| --- |
| **Decision Tree** |
| ID3 Algorithem |
| A decision tree is a tree-like structure that is used as a model for classifying data. A decision tree decomposes the data into sub-trees made of other sub-trees and/or leaf nodes |
|  |





A decision tree is a graphical representation of possible solutions to a problem based on given conditions. It is called a tree because diagrammatically it starts with a single box (target variable) and ends up in numerous branches and roots (numerous solutions). It is a type of supervised learning algorithm that has target variables and in order to select solutions, it creates classifications. Based on classifications, however, it is applied to both categorical and continuous variables.

Using a decision tree, the population or samples can be split into two or more homogeneous sets. These homogeneous sets are constructed based on the most significant differentiator on input variables.

The dataset is used provided a 14 row of data with 4 attributes so can train a tree and construct it to use it later in prediction.

The method used to process the data is **Top-Down Induction of Decision Tree** algorithm .

Some of the result is Written down

Sample data used to train is countered

Name:

overcast

Group:

Outlook Temperature Humidity Wind PlayTennis

2 overcast hot high False yes

6 overcast cool normal True yes

Name:

rainy

Group:

Outlook Temperature Humidity Wind PlayTennis

3 rainy mild high False yes

4 rainy cool normal False yes

5 rainy cool normal True no

9 rainy mild normal False yes

Name:

sunny

Group:

Outlook Temperature Humidity Wind PlayTennis

1 sunny hot high True no

7 sunny mild high False no

8 sunny cool normal False yes

Find entropy of **Outlook**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_EntropyList\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Entropy Prop

Outlook

overcast 0.000000 0.222222

rainy 0.811278 0.444444

sunny 0.918296 0.333333

Counter({'yes': 6, 'no': 3})

9.0

no yes

0.3333333333333333 0.6666666666666666

gainForAll attribuite [0.2516291673878229, 0.029406945165600717, 0.07278022578373267, 0.17884894160409037]

**tree is :**

{'Outlook': {'overcast': 'yes',

'rainy': {'Wind': {False: 'yes', True: 'no'}},

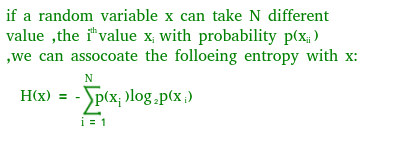
'sunny': {'Temperature': {'cool': 'yes',

'hot': 'no',

'mild': 'no'}}}}

**Discussion:**

Find **entropy** using



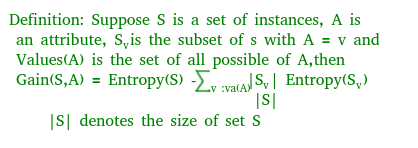
Entropy is the measure of uncertainty of a random variable, it characterizes the impurity of an arbitrary collection of examples. The higher the entropy the more the information content.

Using math library can find log and then sum the results

Example :

Entropy (overcase)= 0.000000 and it probability is 0.222222

**Information Gain**



 The entropy typically changes when we use a node in a decision tree to partition the training instances into smaller subsets. Information gain is a measure of this change in entropy.

Example ('Information Gain', 0.2516291673878229)

Using the information gain function repletely so can decide which attribute take as root node every level and which discard and that’s done in my code using ID3 function

**Accuracy score**

 Accuracy score is used to calculate the accuracy of the trained classifier.

This is done using accuracy\_score method provided by sklearn lib.

**Conclusion:**

Decision trees are able to generate understandable rules , perform classification without requiring much computation , it is able to handle both continuous and categorical variables and it is provide a clear indication of which fields are most important for prediction or classification.

All code can find it on : [**GitHub**](https://github.com/MujahedSaleem/Machinelearning) **(**[**https://github.com/MujahedSaleem/Machinelearning**](https://github.com/MujahedSaleem/Machinelearning)**)**

**Referencing:**

1. Tom mitchell, Machine learning, Chapter 3
2. Geek for geeks
3. skitiLearn
4. pandas