**IS Lab**

**Decision Tree based ID3 algorithm**

**Name- Adhyyan Tripathi**

**Sec/roll no. – C/8**

**Reg. no. – 201700403**

***Decision tree using ID3 Algorithm***

*The decision tree splits the nodes on all available variables and then selects the split which results in most homogeneous sub-nodes. ID3 stands for Iterative Dichotomiser 3. It is one of the many algorithms used to make decision trees. The Algorithm selection is based upon the type of target variables e.g.*

*#C4.5(successor of ID3)*

*#CART(Classification And*

*Regression Tree)*

*#CHAID(Chi-square automatic interaction detection Performs multi-level splits when computing classification trees)*

*#MARS(multivariate adaptive regression splines)*

*ID3 Algorithm iteratively divides the features into two or more groups at each step*

***Steps in ID3 algorithm:***

*#It begins with the original set S as the root node.*

*#On each iteration of the algorithm, it iterates through the very unused attribute of the set S and calculates Entropy(H) and Information gain(IG) of this attribute.*

*#It then selects the attribute which has the smallest Entropy or Largest*

*Information gain. #The set S is then split by the selected attribute to*

*produce a subset of the data.*

*#The algorithm continues to recur on each subset, considering only attributes never selected before.*



***Using a sample dataset of iris flower identification.***

*In [89]:*

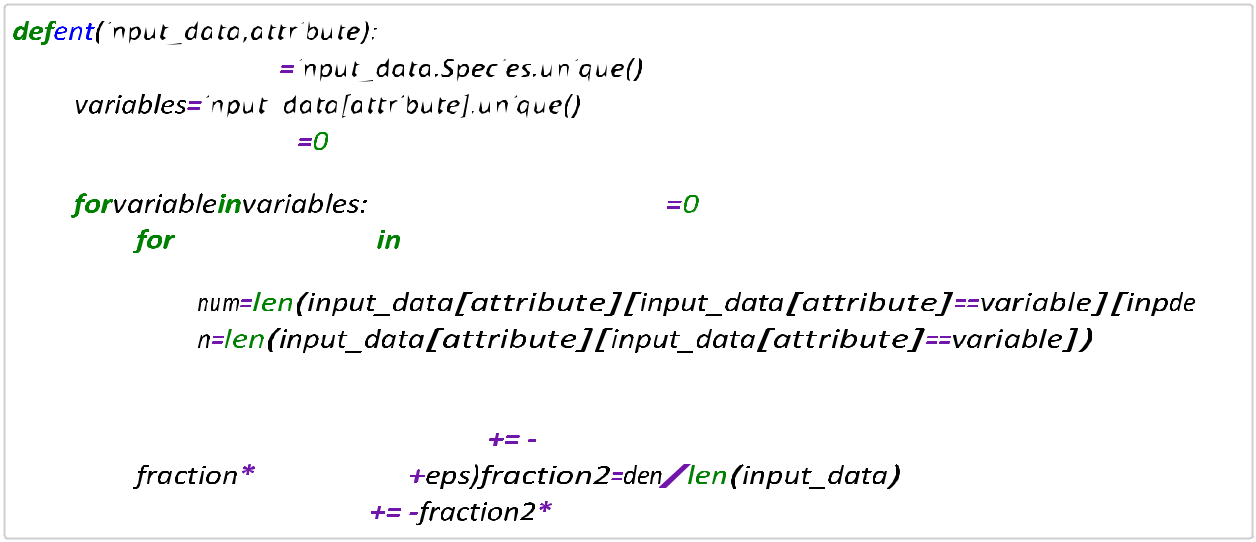


|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *In [90]:* | |  |  |  |  |  |  |  |  |  |  |  |
| *In [91]:* | |  |  |  |  |  |  |  |  |  |  |  |
| *Out[91]:* | |  |  |  |  |  |  |  |  |  |  |  |
|  |  | ***Id SepalLength SepalWidth*** | | | | |  | ***PetalLengt PetalWidth*** | | |  | ***Species*** |
|  |  |  |  |  | ***Cm*** | ***Cm*** |  | ***hCm*** |  | ***Cm*** |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| ***0*** | | *1* | | *5.1* | | *3.5* | *1.4* | | *0.2* | |  | *Iris-* |
|  |  |  |  |  |  |  |  |  |  |  |  | *setosa* |
| ***1*** | | *2* | | *4.9* | | *3.0* | *1.4* | | *0.2* | |  | *Iris-* |
|  |  |  |  |  |  |  |  |  |  |  |  | *setosa* |
| ***2*** | | *3* | | *4.7* | | *3.2* | *1.3* | | *0.2* | |  | *Iris-* |
|  |  |  |  |  |  |  |  |  |  |  |  | *setosa* |
| ***3*** | | *4* | | *4.6* | | *3.1* | *1.5* | | *0.2* | |  | *Iris-* |
|  |  |  |  |  |  |  |  |  |  |  |  | *setosa* |
| ***4*** | | *5* | | *5.0* | | *3.6* | *1.4* | | *0.2* | |  | *Iris-* |
|  |  |  |  |  |  |  |  |  |  |  |  | *setosa* |



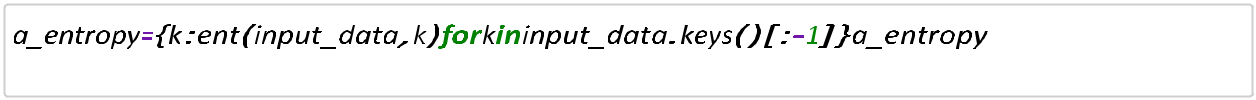
*In [92]:* *eps**eps*

***Calculating Entropy of each feature***



*In [93]:*

*eps*



*In [94]:*

*Out[94]: {'Id': 0.0,*

*'SepalLengthCm': 0.7080248798300978,*

*'SepalWidthCm': 1.0740925365975489,*

*'PetalLengthCm': 0.1386459770753558,*

*'PetalWidthCm': 0.14906466204571406}*



***Find Information Gain***

*In [95]:* 



*In [96]:* 

*In [97]:* *IG*



*Out[97]: {'Id': 1.584962500721156,*

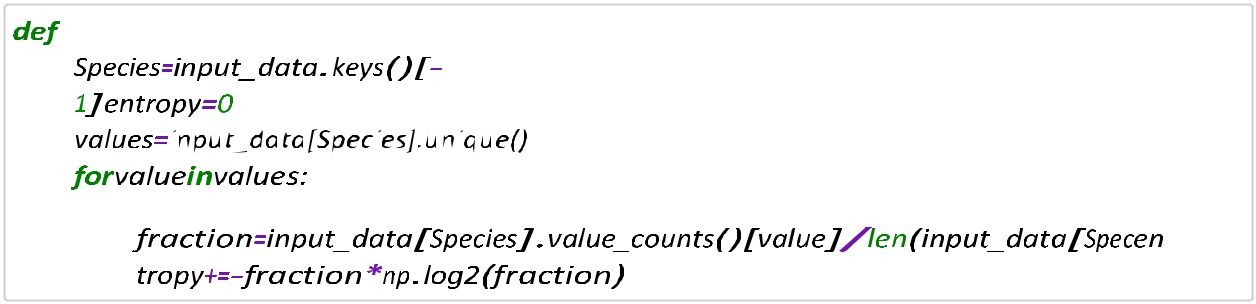
*'SepalLengthCm': 0.8769376208910583,*

*'SepalWidthCm': 0.5108699641236072,*

*'PetalLengthCm': 1.4463165236458002,*

*'PetalWidthCm': 1.435897838675442}*

*Find entropy of original dataset S*



*In [98]:*



*In [99]:*

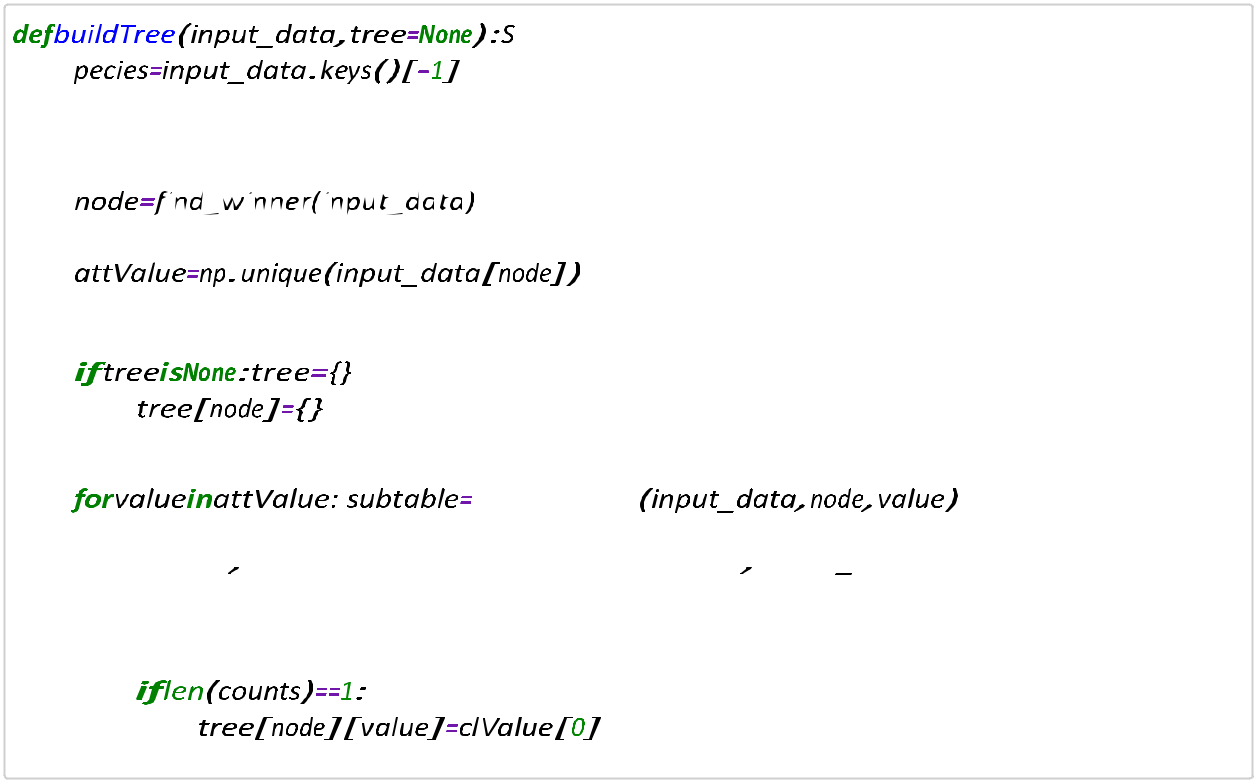
*key*

 *key*



*In [100]:*

***True***



*In [101]:*

***True***

*In [102]:t****=****buildTree(input\_data)*



*In [103]:* 

*In [104]:* 

