**TakenMind Global Managerial Internship**

**Assignment#4**

* Importing the required packages
* ***import pandas as pd***
* ***import sklearn.model\_selection as ms***
* ***from sklearn.preprocessing import LabelEncoder***
* ***import sklearn.metrics as metrics***
* ***from sklearn.tree import DecisionTreeClassifier***
* Read the data and print the header values
* ***data = pd.read\_csv("C:/Users/lenovo/Downloads/Iris.csv")***
* ***print data.head(10)***
* Transform the dependent variable into number
* ***enn = LabelEncoder()***
* ***data['Species\_n'] = enn.fit\_transform(data['Species'])***
* ***print data***
* Drop the dependent variable and new variable
* ***X = data.drop(['Species', 'Species\_n'], axis='columns')***
* ***print X***
* Separate the dependent value ‘y’
* ***y = data['Species\_n']***
* ***print y***
* Split the data into train and test in the ratio 7:3
* ***X\_train, X\_test, y\_train, y\_test = ms.train\_test\_split(X, y, test\_size=0.3, random\_state=100)***
* ***print X\_train.shape, X\_test.shape***
* ***print y\_train.shape, y\_test.shape***
* Made a model with gini Index.
* ***model\_gini = DecisionTreeClassifier(criterion="gini", random\_state=100, max\_depth=3, min\_samples\_leaf=5)***
* ***model\_gini.fit(X\_train, y\_train)***
* ***print model\_gini***
* Made a model with entropy Index.
* ***model\_entropy = DecisionTreeClassifier(criterion="entropy", random\_state=100, max\_depth=3, min\_samples\_leaf=5)***
* ***model\_entropy.fit(X\_train, y\_train)***
* ***print model\_entropy***
* Predict the values of two models
* ***pred\_gini = model\_gini.predict(X\_test)***
* ***print pred\_gini***
* ***pred\_entropy = model\_entropy.predict(X\_test)***
* ***print pred\_entropy***
* Calculate the accuracy of two models
* ***print ('Accuracy using Gini index', metrics.accuracy\_score(y\_test, pred\_gini))***
* ***print ('Accuracy using Entropy index', metrics.accuracy\_score(y\_test, pred\_entropy))***