



TEAM:

Nimisha - PES1201700083

Shashank - PES1201700667

Suhail - PES1201701420

Riya - PES1201701814

ASSIGNMENT 8

Problem statement:

Implement machine learning techniques to design a classifier using decision trees.

Dataset:

GALEX_data-extended-feats data

CODE:

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report, confusion_matrix
from sklearn.metrics import accuracy_score

data = pd.read_csv("GALEX_data-extended-feats.csv")
X=data.drop('class',axis=1)
y= data['class']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.15, random_state=23)

clf_entropy = DecisionTreeClassifier( criterion = "entropy", random_state = 100, max_depth = 6, min_samples_leaf = 6)
clf_entropy.fit(X_train, y_train)

y_pred = clf_entropy.predict(X_test)
print("Predicted values:")
print(y_pred)

print("Confusion Matrix: ", confusion_matrix(y_test, y_pred))

print ("Accuracy : ", accuracy_score(y_test,y_pred)*100)

print("Report : ", classification_report(y_test, y_pred))
```



RESULT:

Predicted values:

[0 1 1 ... 0 1 1]

Confusion Matrix: [[82 7 40]

[6 795 24]

[31 25 251]]

Accuracy : 89.45281522601111

Report : precision recall f1-score support

0 0.69 0.64 0.66 129

1 0.96 0.96 0.96 825

2 0.80 0.82 0.81 307

micro avg 0.89 0.89 0.89 1261

macro avg 0.82 0.81 0.81 1261

weighted avg 0.89 0.89 0.89 1261