

Showing 1 changed file with 111 additions and 0 deletions.

111 decision tree

@@ -0,0 +1,111 @@

```
1 + # Run this program on your local python
2 + # interpreter, provided you have installed
3 + # the required libraries.
4 +
5 + # Importing the required packages
6 + import numpy as np
7 + import pandas as pd
8 + from sklearn.metrics import confusion_matrix
9 + from sklearn.model_selection import train_test_split
10 + from sklearn.tree import DecisionTreeClassifier
11 + from sklearn.metrics import accuracy_score
12 + from sklearn.metrics import classification_report
13 +
14 + # Function importing Dataset
15 + def importdata():
16 +     balance_data = pd.read_csv(
17 +         'https://archive.ics.uci.edu/ml/machine-learning-'+
18 +         'databases/balance-scale/balance-scale.data',
19 +         sep= ',', header = None)
20 +
21 +     # Printing the dataset shape
22 +     print ("Dataset Length: ", len(balance_data))
```

78°F  
Partly cloudy



Showing 1 changed file with 111 additions and 0 deletions.

```
23 +     print ("Dataset Shape: ", balance_data.shape)
24 +
25 +     # Printing the dataset observations
26 +     print ("Dataset: ",balance_data.head())
27 +     return balance_data
28 +
29 + # Function to split the dataset
30 + def splitdataset(balance_data):
31 +
32 +     # Separating the target variable
33 +     X = balance_data.values[:, 1:5]
34 +     Y = balance_data.values[:, 0]
35 +
36 +     # Splitting the dataset into train and test
37 +     X_train, X_test, y_train, y_test = train_test_split(
38 + X, Y, test_size = 0.3, random_state = 100)
39 +
40 +     return X, Y, X_train, X_test, y_train, y_test
41 +
42 + # Function to perform training with giniIndex.
43 + def train_using_gini(X_train, X_test, y_train):
44 +
45 +     # Creating the classifier object
46 +     clf_gini = DecisionTreeClassifier(criterion = "gini",
47 +                                     random_state = 100,max_depth=3, min_samples_leaf=5)
```

78°F  
Partly cloudy



Showing 1 changed file with 111 additions and 0 deletions.

```
87 + # Driver code
88 + def main():
89 +
90 +     # Building Phase
91 +     data = importdata()
92 +     X, Y, X_train, X_test, y_train, y_test = splitdataset(data)
93 +     clf_gini = train_using_gini(X_train, X_test, y_train)
94 +     clf_entropy = train_using_entropy(X_train, X_test, y_train)
95 +
96 +     # Operational Phase
97 +     print("Results Using Gini Index:")
98 +
99 +     # Prediction using gini
100 +     y_pred_gini = prediction(X_test, clf_gini)
101 +     cal_accuracy(y_test, y_pred_gini)
102 +
103 +     print("Results Using Entropy:")
104 +     # Prediction using entropy
105 +     y_pred_entropy = prediction(X_test, clf_entropy)
106 +     cal_accuracy(y_test, y_pred_entropy)
107 +
108 +
109 + # Calling main function
110 + if __name__ == "__main__":
111 +     main()
```

78°F  
Partly cloudy

