

DecissionTree ML algo

February 5, 2021

```
In [ ]: #import necessary packages
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
from sklearn import tree

#loading data file
balance_data = pd.read_csv(r'C:\Users\Desktop\traffic uk\accidentjoin1.csv',sep =',',hea
print(balance_data)

In [ ]: print(len(balance_data))
print(balance_data.shape)
balance_data.head()

In [ ]: # sepration Target Variable
x=balance_data.values[:,1:20]
y = balance_data.values[:,0]
#splitting dataset into test and train
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.3,random_state=100)
# function to perform training with entropy
clf_entropy= DecisionTreeClassifier(criterion="entropy",random_state=100,max_depth=25,mi
clf_entropy.fit(x_train,y_train)

In [ ]: #function to make prediction
y_pred_en = clf_entropy.predict(x_test)
print(y_pred_en)

In [ ]: print("accuracy is",accuracy_score(y_test,y_pred_en)*100)
print(classification_report(y_test,y_pred_en))

In [ ]: clf = DecisionTreeClassifier(random_state=0)
clf= clf.fit(x_train,y_train)
print(dict(zip(balance_data.columns,clf.feature_importances_)))

In [ ]: # testing random data inputs
balance_test1 = pd.read_csv(r'C:\Users\Desktop\traffic uk\test2.csv',sep =',',header=0)
print(balance_test1)
```

```
In [ ]: testcriteria=balance_test1.values[:,1:21]
        testhead = balance_test1.values[:,0]

In [ ]: y_pred_en = clf_entropy.predict(testcriteria)
        print(y_pred_en)

In [ ]: print("accuracy is",accuracy_score(testhead,y_pred_en)*100)

In [ ]:

In [ ]:
```