



UNIVERSITY OF INFORMATION
TECHNOLOGY AND SCIENCES (UITS)
DEPARTMENT OF INFORMATION TECHNOLOGY

LAB REPORT No.: 2

IT-452 : MACHINE LEARNING

Decision Tree Implementation

Submitted To:

Sumaiya Akhtar Mitu
Lecturer,
Department of IT ,
UITS
Email:sumaiya.akhtar@uits.edu.bd

Submitted By:

Name:Nazmul Zaman
Student ID:2014755055
Department of IT, UITS

Contents

1	Abstract	2
2	Advantage of Regression	2
3	Disadvantages of Regression	2
4	Code	2
5	Conclusion	5
6	References	5

1 Abstract

A decision tree is a decision support hierarchical model that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

2 Advantage of Regression

1. Some advantages of decision trees are:
2. Simple to understand and to interpret.
3. Requires little data preparation
4. The cost of using the tree (i.e., predicting data) is logarithmic in the number of data points used to train the tree.
5. Able to handle both numerical and categorical data.
6. Able to handle multi-output problems. Decision models can include all the variables that one wants to include in the model.

3 Disadvantages of Regression

The disadvantages of decision trees include: Decision-tree learners can create over-complex trees that do not generalize the data well
Decision trees can be unstable because small variations in the data might result in a completely different tree being generated.

4 Code

```
1 # -*- coding: utf-8 -*-
2 """lab2.ipynb
3
4 Automatically generated by Colaboratory.
5
6 Original file is located at
7     https://colab.research.google.com/drive/1
8     tkAV_1hy3FIFCv82FmltweJPY09krZ0h
9 """
10 import numpy as np
11 import pandas as pd
12
13 # Importing data
14
15 df = pd.read_csv("https://raw.githubusercontent.com/milaan9/
    Python_Decision_Tree_and_Random_Forest/main/dataset/
```

```
    playgolf_data.csv")
16 df
17
18 df.dtypes
19
20 df.info()
21
22 # Converting categorical variables into dummies/indicator
    variables
23
24 df_getdummy=pd.get_dummies(data=df, columns=['Temperature', '
    Humidity', 'Outlook', 'Wind'])
25 df_getdummy
26
27 # Separating the training set and test set
28
29 from sklearn.model_selection import train_test_split
30
31 X = df_getdummy.drop('PlayGolf',axis=1)
32 y = df_getdummy['PlayGolf']
33 X_train, X_test, y_train, y_test = train_test_split(X, y,
    test_size=0.30, random_state=101)
34
35 # importing Decision Tree Classifier via sklearn
36
37 from sklearn.tree import DecisionTreeClassifier
38
39 dtree = DecisionTreeClassifier(criterion='entropy',max_depth=2)
40 dtree.fit(X_train,y_train)
41 predictions = dtree.predict(X_test)
42
43 # visualising the decision tree diagram
44
45 from sklearn.tree import plot_tree
46
47 import matplotlib.pyplot as plt
48 fig = plt.figure(figsize=(16,12))
49 a = plot_tree(dtree, feature_names=df_getdummy.columns,
    fontsize=12, filled=True,
50               class_names=['Not_Play', 'Play'])
```

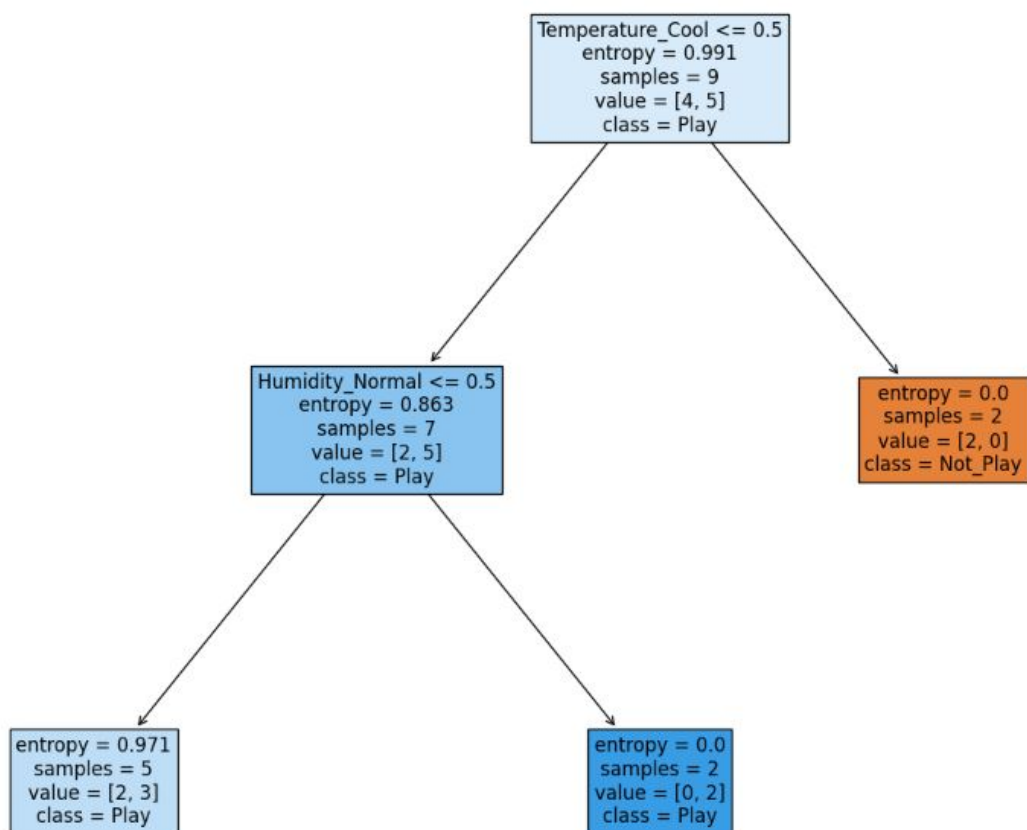


Figure 1:

5 Conclusion

Machine learning analytic uses machine decision tree algorithm mostly use to take a decision from different data set .

It's also called as a predictive model, it is used in many areas for its split approach which helps in identifying solutions based on different conditions by either classification or regression method.

6 References

2. <https://www.javatpoint.com/linear-regression-in-machine-learning>
3. <https://www.javatpoint.com/machine-learning-polynomial-regression>