

CP3252 - Machine Learning Lab – Week 4

Date: 04.01.2025

TITLE

IMPLEMENTATION OF DECISION TREE ALGORITHM

TASK

- (1) The table below shows factors that would aid a player to play Golf or not. Build a Decision Tree model that could predict whether a player would play or not with implementation in Python. Build the model manually and verify the same with code.

Attributes				Classes
Outlook	Temperature	Humidity	Windy	Play Golf
Rainy	Hot	High	FALSE	No
Rainy	Hot	High	TRUE	No
Overcast	Hot	High	FALSE	Yes
Sunny	Mild	High	FALSE	Yes
Sunny	Cool	Normal	FALSE	Yes
Sunny	Cool	Normal	TRUE	No
Overcast	Cool	Normal	TRUE	Yes
Rainy	Mild	High	FALSE	No
Rainy	Cool	Normal	FALSE	Yes
Sunny	Mild	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes
Overcast	Mild	High	TRUE	Yes
Overcast	Hot	Normal	FALSE	Yes
Sunny	Mild	High	TRUE	No

- (a) When do we perform splitting of a Decision Tree Node and how?
(b) What can be understood from Entropy and Information Gain?

(2) Data Set Description: Data Filename: data4_19.csv The data set contains 150 data points, there are three classes where each class refers to a type of iris plant. The first four columns represent the attributes listed below. Note that only the first four columns should be used as attributes. The last column is the ground truth class name.

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. Ground truth class name: -- Iris Setosa -- Iris Versicolour -- Iris Virginica

Write Python code to implement Decision Tree for this dataset and find the species of the new flower given Sepal length= 5.2, Sepal width =3.1, Petal length= 1.4, Petal width =0.2.