

Project 2

In your language of choice, implement the decision tree induction algorithm based on the information gain calculation as discussed for the ID3/C4.5 algorithm.

Define an input format for your program to provide it input data.

Transform the following example data (last column play represents the class) into your input format.

Provide 10 ten tuples as input to your program to induce the decision tree. Then output the induced tree in any readable format which explains the structure of the tree.

Use the last 4 tuples as test data, record your program output and compare with the actual class labels. What is the accuracy of your implementation of the decision tree?

Outlook	Temperature	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No