CIS 365 Entropy and Information Gain

Dr. Denton Bobeldyk

Given the dataset below:

Weather	Temperature	Humidity	Wind	Play Outside?
Sunny	Hot	High	Weak	Yes
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rainy	Mild	High	Weak	No
Rainy	Cool	Normal	Weak	No
Rainy	Cool	Normal	Strong	Yes
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	Yes
Sunny	Cool	Normal	Weak	Yes
Rainy	Mild	Normal	Weak	No

Determine the attribute (column) which provides the highest information gain in determining whether a data point from the dataset decides to play outside.

Create a decision tree with the above attribute placed at the root of the node. For each of the categories for that attribute you should create a branch. Calculate the information Gain for each of the remaining attributes and do one more branch. Stopping after 2 levels, have your decision tree predict whether you would play outside.

Hand-in:

- A word/pdf document containing a decision tree (2 levels) with your work in determining the information gain shown. If you wrote a program to find the solution, please include that as well.