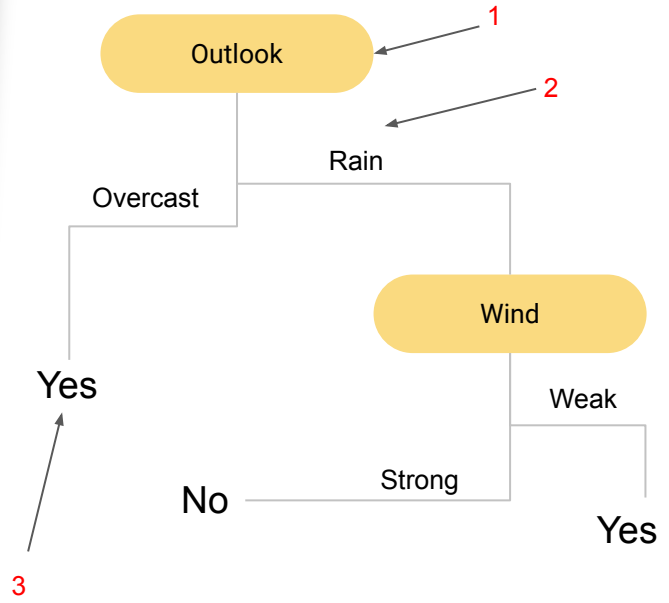
Create and describe the algorithm to automate the calculation of the Decision Tree for the Use Case “Playing Tennis” using ID3method

Homework H4.5 by Daniel Rück and Brian Brandner

Decision Tree

- Decision tree learning
- Predictive model
- used for data mining and machine learning
- node = feature(attribute)[1]
- link(branch) = decision(rule)[2]
- leaf = outcome (categorical or continues value)[3]



Playing Tennis

- Weather dataset for machine learning
- Playing or not playing a game based on weather condition
- Count the frequencies

	Outlook	Temperature	Humidity	Windy	Play
0	Sunny	Hot	High	F	No
1	Sunny	Hot	High	T	No
2	Overcast	Hot	High	F	Yes
3	Rainy	Mild	High	F	Yes
4	Rainy	Cool	Normal	F	Yes
5	Rainy	Cool	Normal	T	No
6	Overcast	Cool	Normal	T	Yes

ID3algorithm

- Iterative Dichotomizer
- Algorithm to build a decision tree
- uses **Entropy** function and **Information gain** as metrics

Root value

- classifies the training data the best
- highest Information Gain

Entropy formula

$$H(S) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i)$$

H - greek Eta, Entropy

S - Dataset

p(x_i) - Proportion of classification to results (Quantity of Yes or No)

Information Gain formula

$$IG(S, C) = H(S_{Total}) - \sum p(Z_{Column}) * H(S_{Column})$$

IG - Information Gain

S - Dataset

C - Column

H(S_Total) - Total entropy of the dataframe

p(Z_Column) - Value count of active column divided by max column length

H(S_Column) - Entropy of active column value

implementation with Jupyter Notebook