CIT-223-011/2021

JOHN GICHARU

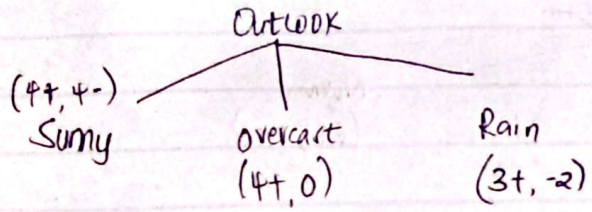
Question One

b) Machine translation - Use of computational techniques to translate text or speech from one natural language (the source language) into another (the target language).

Question two

Entropy

Outlook



$= -\frac{11}{17}\log_2 \frac{11}{17} - \frac{6}{17}\log_2 \frac{6}{17}$

$= 0.9365$

Sunny $= -\frac{4}{8}\log_2 \frac{4}{8} - \frac{4}{8}\log_2 \frac{4}{8} = 1$

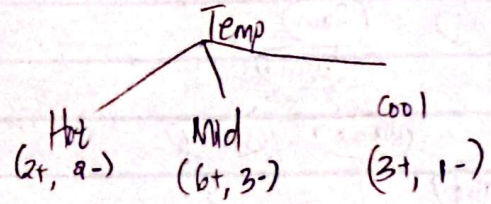
Overcast $= -\frac{4}{4}\log_2 \frac{4}{4} - \frac{0}{4}\log_2 \frac{0}{4} = 0$

Rain $= -\frac{3}{5}\log_2 \frac{3}{5} - \frac{2}{5}\log_2 \frac{2}{5} = 0.971$

Info gain $= 0.9365 - \left(\frac{8}{17}(1) + \frac{4}{17}(0) + \frac{5}{17}(0.971)\right)$

$= 0.1803$

Temparature



Hot $= \left(-\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4}\right) * \frac{4}{17}$ (1)

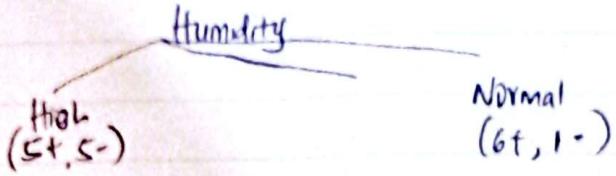
Mild $= \left(-\frac{6}{9}\log_2\frac{6}{9} - \frac{3}{9}\log_2\frac{3}{9}\right)\frac{9}{17} = \frac{9}{17}(0.9182)$

Cool $= \left(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}\right)\frac{4}{17} = \frac{4}{17}(0.8113)$

Hot + Mild + Cool = 0.928

Info Gain = 0.9305 - 0.928 = 0.0085

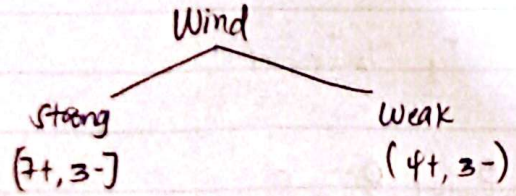
Humidity



High Entropy = High = 1

Normal $= -\frac{6}{7}\log\frac{6}{7} - \frac{1}{7}\log_2\frac{1}{7} = 0.5917$

Wind



Weak $= -\frac{7}{10}\times\log_2\frac{7}{10} - \frac{3}{10}\log_2\frac{3}{10} = 0.8813$

Strong $= -\frac{4}{7}\log_2\frac{4}{7} - \frac{3}{7}\log_2\frac{3}{7} = 0.9852$

Info Gain $= 0.9365 - \left(\left\{\frac{10}{17}\times 0.8813\right\} + \left(\frac{7}{17}\times 0.9852\right)\right) = 0.0101$

Decision tree