CSE 438

Ehsanur Rahman Rhythm
20101298

## Types of Data:

BMI → Numerical ~~(Interval)~~

Smoking → Categorical ~~(Nominal)~~

PhysicalHealth → Numerical

Race → Categorical

AgeCategory → Numerical

SkinCancer → Categorical

## Calculating Entropy

For BMI, we can make data intervals:

$x < 20 \rightarrow 1$

$20 < x < 25 \rightarrow 22$

$25 < x < 30 \rightarrow 39$

$30 < x < 35 \rightarrow 25$

$35 < x < 40 \rightarrow 5$

$x > 40 \rightarrow 8$

$$\therefore \text{Entropy (BMI)} = -\frac{1}{100}\log_2\left(\frac{1}{100}\right) - \frac{22}{100}\log_2\left(\frac{22}{100}\right) - \frac{39}{100}\log_2\left(\frac{39}{100}\right) - \frac{25}{100}\log_2\left(\frac{25}{100}\right) - \frac{5}{100}\log_2\left(\frac{5}{100}\right) - \frac{8}{100}\log_2\left(\frac{8}{100}\right)$$

$$= 2.08441$$

Smoking:

$0 \rightarrow 53$

$1 \rightarrow 47$

$\therefore$ Entropy (Smoking) $= -\frac{53}{100}\log_2\left(\frac{53}{100}\right) - \frac{47}{100}\log_2\left(\frac{47}{100}\right)$

$= 0.9974$

Physical Health:

$x < 5 \rightarrow 31$

$5 < x < 10 \rightarrow 31$

$10 < x < 20 \rightarrow 11$

$x > 20 \rightarrow 27$

$\therefore$ Entropy (Physical Health) $= -\frac{31}{106}\log_2\left(\frac{31}{100}\right) - \frac{31}{100}\log_2\left(\frac{31}{100}\right)$

$- \frac{11}{100}\log_2\left(\frac{11}{100}\right) - \frac{27}{100}\log_2\left(\frac{27}{100}\right)$

$= 1.90789$

Race:

0 → 1

1 → 20

2 → 1

3 → 78

$$\text{Entropy(Race)} = -\frac{1}{100}\log_2\left(\frac{1}{100}\right) - \frac{20}{100}\log_2\left(\frac{20}{100}\right) - \frac{1}{100}\log_2\left(\frac{1}{100}\right) - \frac{78}{100}\log_2\left(\frac{78}{100}\right)$$

$$= 0.87685$$

Age Category:

$x < 30 \rightarrow 1$

$31 < x < 50 \rightarrow 5$

$51 < x < 70 \rightarrow 44$

$71 < x < 90 \rightarrow 50$

Entropy (Age Category) =

$$-\frac{1}{100}\log_2\left(\frac{1}{100}\right) - \frac{5}{100}\log_2\left(\frac{5}{100}\right)$$

$$-\frac{44}{100}\log_2\left(\frac{44}{100}\right) - \frac{50}{100}\left(\frac{50}{100}\right)$$

$$= 1.30368$$

Skin Cancer :

0 → 78

1 → 22

Entropy (skin Cancer) $= -\frac{78}{100}\log_2\left(\frac{78}{100}\right) - \frac{22}{100}\log_2\left(\frac{22}{100}\right)$

$= 0.76016$