# * Naive Bayes

| Day | Outlook | Temperature | Humidity | Wind | Play Tennis |
|---|---|---|---|---|---|
| 1 | Sunny | Hot | High | Weak | No |
| 2 | Sunny | Hot | High | Strong | No |
| 3 | Overcast | Hot | High | Weak | Yes |
| 4 | Rain | Mild | High | Weak | Yes |
| 5 | Rain | Cool | Normal | Weak | Yes |
| 6 | Rain | Cool | Normal | Strong | No |
| 7 | Overcast | Cool | Normal | Strong | Yes |
| 8 | Sunny | Mild | High | Weak | No |
| 9 | Sunny | Cool | Normal | Weak | Yes |
| 10 | Rain | Mild | Normal | Weak | Yes |
| 11 | Sunny | Mild | Normal | Strong | Yes |
| 12 | Overcast | Mild | High | Strong | Yes |
| 13 | Overcast | Hot | Normal | Weak | Yes |
| 14 | Rain | Mild | High | Strong | No |

Given a new instance,

$x'$ = (Outlook = Sunny, Temperature = Cool, Humidity = High, Wind = Strong)

$$P(\text{Play} = \text{Yes}) = \frac{9}{14}$$

$$P(\text{Play} = \text{No}) = \frac{5}{14}$$

| Humidity | Yes | No |
|---|---|---|
| High | 3/9 | 4/5 |
| Normal | 6/9 | 1/5 |

| Outlook | Yes | No |
|---|---|---|
| Sunny | 2/9 | 3/5 |
| Overcast | 4/9 | 0/5 |
| Rain | 3/9 | 2/5 |

| Temp. | Yes | No |
|---|---|---|
| Hot | 2/9 | 2/5 |
| Mild | 4/9 | 2/5 |
| Cool | 3/9 | 1/5 |

| Wind | Yes | No |
|---|---|---|
| Strong | 3/9 | 3/5 |
| Weak | 6/9 | 2/5 |

$$P(C_i|x) = P(x_1|C_i) \times P(x_2|C_i) \times \dots\dots P(x_n|C_i) \times P(C_i)$$

$$\therefore P(\text{Yes}|x') = [P(\text{Sunny}|\text{Yes}) * P(\text{Cool}|\text{Yes}) * P(\text{High}|\text{Yes}) * P(\text{Strong}|\text{Yes})] * P(\text{Play} = \text{Yes})$$

$$= [2/9 * 3/9 * 3/9 * 3/9] * 9/14$$
$$= 0.0053$$

$$\therefore P(\text{No}|x') = [P(\text{Sunny}|\text{No}) * P(\text{Cool}|\text{No}) * P(\text{High}|\text{No}) * P(\text{Strong}|\text{No})] * P(\text{Play} = \text{No})$$

$$= [3/5 * 1/5 * 4/5 * 3/5] * 5/14$$
$$= 0.0206$$

⇝ we label $x'$ to be "No"

## * Gini Index

| RID | age | income | student | credit_rating | buys_computer |
|---|---|---|---|---|---|
| 1 | youth | high | no | fair | no |
| 2 | Youth | high | no | excellent | no |
| 3 | middle-age | high | no | fair | yes |
| 4 | senior | medium | no | fair | yes |
| 5 | senior | low | yes | fair | yes |
| 6 | senior | low | yes | excellent | no |
| 7 | middle-age | low | yes | excellent | yes |
| 8 | youth | medium | no | fair | no |
| 9 | youth | low | yes | fair | yes |
| 10 | senior | medium | yes | fair | yes |
| 11 | youth | medium | yes | exeellent | yes |
| 12 | middle-age | medium | no | excellent | yes |
| 13 | middle-age | high | yes | fair | yes |
| 14 | senior | medium | no | excellent | no |

$$Gini(D) = 1 - \left(\frac{9}{14}\right)^2 - \left(\frac{5}{14}\right)^2 = 0.4592$$

for
ncome •

| Distinct Values | Yes | No | Total |
|---|---|---|---|
| Low, Medium | 7 | 3 | 10 |
| High | 2 | 2 | 4 |

$$Gini_{income \in \{low, medium\}}(D) = \frac{10}{14} Gini(D_1) + \frac{4}{14} Gini(D_2)$$

$$= \frac{10}{14}\left(1 - \left(\frac{7}{10}\right)^2 - \left(\frac{3}{10}\right)^2\right) + \frac{4}{14}\left(1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2\right)$$

$$= 0.443$$

$$= \text{Gini}_{\text{income} \in \{\text{High}\}}(D)$$

| Distinct Values | Yes | No | Total |
|---|---|---|---|
| Low, High | 5 | 3 | 8 |
| Medium | 4 | 2 | 6 |

$$\text{Gini}_{\text{income} \in \{\text{low, High}\}}(D) = \frac{8}{14}\,\text{Gini}(D_1) + \frac{6}{14}\,\text{Gini}(D_2)$$

$$= \frac{8}{14}\left(1 - \left(\frac{5}{8}\right)^2 - \left(\frac{3}{8}\right)^2\right) + \frac{6}{14}\left(1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2\right)$$

$$= 0.4583$$

$$= \text{Gini}_{\text{income} \in \{\text{Medium}\}}(D)$$

| Distinct Values | Yes | No | Total |
|---|---|---|---|
| medium, High | 6 | 4 | 10 |
| Low | 3 | 1 | 4 |

$$\text{Gini}_{\text{income} \in \{\text{medium, High}\}}(D) = \frac{10}{14}\,\text{Gini}(D_1) + \frac{4}{14}\,\text{Gini}(D_2)$$

$$= \frac{10}{14}\left(1 - \left(\frac{6}{10}\right)^2 - \left(\frac{4}{10}\right)^2\right) + \frac{4}{14}\left(1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2\right)$$

$$= 0.4500$$

$$= \text{Gini}_{\text{income} \in \{\text{low}\}}(D)$$


Darshan UNIVERSITY

for age •

| Distinct Values | Yes | No | Total |
|---|---|---|---|
| Youth, middle-age | 6 | 3 | 9 |
| senior | 3 | 2 | 5 |

$$\text{Gini}_{age \in \{youth, middle\text{-}age\}}(D) = \frac{9}{14}\,\text{Gini}(D_1) + \frac{5}{14}\,\text{Gini}(D_2)$$

$$= \frac{9}{14}\left(1 - \left(\frac{6}{9}\right)^2 - \left(\frac{3}{9}\right)^2\right) + \frac{5}{14}\left(1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2\right)$$

$$= 0.4571$$

$$= \text{Gini}_{age \in \{senior\}}(D)$$

| Distinct Values | Yes | No | Total |
|---|---|---|---|
| Youth, Senior | 5 | 5 | 10 |
| middle-age | 4 | 0 | 4 |

$$\text{Gini}_{age \in \{youth, senior\}}(D) = \frac{10}{14}\,\text{Gini}(D_1) + \frac{4}{14}\,\text{Gini}(D_2)$$

$$= \frac{10}{14}\left(1 - \left(\frac{5}{10}\right)^2 - \left(\frac{5}{10}\right)^2\right) + \frac{4}{14}\left(1 - \left(\frac{4}{4}\right)^2 - \left(\frac{0}{4}\right)^2\right)$$

$$= 0.3571$$

$$= \text{Gini}_{age \in \{middle\text{-}age\}}(D)$$

| Distinct Values | Yes | No | Total |
|---|---|---|---|
| Senior, middle-age | 7 | 2 | 9 |
| Youth | 2 | 3 | 5 |

$Gini_{age \in \{senior,\ middle\text{-}age\}}(D) = \frac{9}{14} Gini(D_1) + \frac{5}{14} Gini(D_2)$

$= \frac{9}{14}\left[1 - \left(\frac{7}{9}\right)^2 - \left(\frac{2}{9}\right)^2\right] + \frac{5}{14}\left[1 - \left(\frac{2}{5}\right)^2 - \left(\frac{3}{5}\right)^2\right]$

$= 0.3937$

$= Gini_{age \in \{youth\}}(D)$

for Student •

| Distinct Values | Yes | No | Total |
|---|---|---|---|
| Yes | 6 | 1 | 7 |
| No | 3 | 4 | 7 |

$Gini_{student \in \{yes\}}(D) = \frac{7}{14} Gini(D_1) + \frac{7}{14} Gini(D_2)$

$= \frac{7}{14}\left[1 - \left(\frac{6}{7}\right)^2 - \left(\frac{1}{7}\right)^2\right] + \frac{7}{14}\left[1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2\right]$

$= 0.3674$

$= Gini_{student \in \{no\}}(D)$

for Credit-rating •

| Distinct Values | Yes | No | Total |
|---|---|---|---|
| fair | 6 | 2 | 8 |
| excellent | 3 | 3 | 6 |

$Gini_{credit\_rating \in \{fair\}}(D) = \frac{8}{14} Gini(D_1) + \frac{6}{14} Gini(D_2)$

$= \frac{8}{14}\left[1 - \left(\frac{6}{8}\right)^2 - \left(\frac{2}{8}\right)^2\right] + \frac{6}{14}\left[1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2\right]$

$= 0.429$

$= Gini_{credit\_rating \in \{excellent\}}(D)$

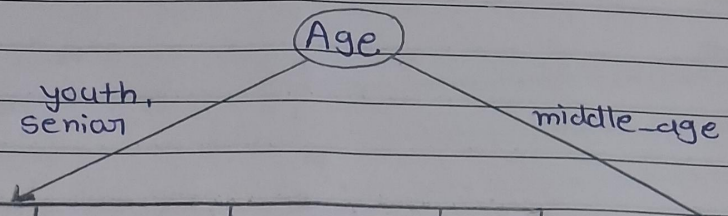
→ $Gini_{age \in \{youth, middle\_age\}}(D) = Gini_{age \in \{senior\}}(D) = 0.4571$ ✗
$Gini_{age \in \{youth, senior\}}(D) = Gini_{age \in \{middle\_age\}}(D) = 0.3571$ ✓
$Gini_{age \in \{senior, middle\_age\}}(D) = Gini_{age \in \{youth\}}(D) = 0.3937$ ✗

→ $Gini_{student \in \{yes\}}(D) = Gini_{student \in \{no\}}(D) = 0.3674$ ✓

→ $Gini_{credit\_rating \in \{fair\}}(D) = Gini_{credit\_rating \in \{excellent\}}(D)$
$= 0.429$ ✓

→ $Gini_{income \in \{low, High\}}(D) = Gini_{income \in \{medium\}}(D) = 0.4583$ ✗
$Gini_{income \in \{low, medium\}}(D) = Gini_{income \in \{high\}}(D) = 0.443$ ✓
$Gini_{income \in \{medium, high\}}(D) = Gini_{income \in \{low\}}(D) = 0.4500$ ✗

| Attribute | Gini | ΔGini |
|---|---|---|
| Age | 0.3571 | $0.4592 - 0.3571 = 0.1021$ ✓ |
| Income | 0.443 | $0.4592 - 0.443 = 0.0162$ |
| Student | 0.3674 | $0.4592 - 0.3674 = 0.0918$ |
| Credit_rating | 0.429 | $0.4592 - 0.429 = 0.0302$ |




| income | student | credit -rating | buys- computer | RID |
|---|---|---|---|---|
| high | no | fair | no | 1 |
| high | no | excellent | no | 2 |
| medium | no | fair | yes | 4 |
| low | yes | fair | yes | 5 |
| low | yes | excellent | no | 6 |
| medium | no | fair | no | 8 |
| low | yes | fair | yes | 9 |
| medium | yes | fair | yes | 10 |
| medium | yes | excellent | yes | 11 |
| medium | no | excellent | no | 14 |

| RID | income | student | credit _rating | buys- computer |
|---|---|---|---|---|
| 3 | high | no | fair | yes |
| 7 | low | yes | excellent | yes |
| 12 | medium | no | excellent | yes |
| 13 | high | yes | fair | yes |