

DA Assignment

2019120061 Pratham P Solanki

| | |
|----------|-----|
| PAGE NO. | 1 |
| DATE | / / |

1) Database Attributes:- 4

{ Day, Season, Fog, Rain }

Categories of classes:- 4

{ On time, Late, Very Late, Cancelled }

We used Naive Bayes classifier for foll. instance

A = { Week Day, Winter, High, None }

| Day | Ontime(14) | Late(2) | Very Late(3) | Cancelled(1) |
|----------|------------|---------|--------------|--------------|
| Weekday | 9/14 | 1/2 | 3/3 | 0 |
| Holiday | 2/14 | 1/2 | 0 | 0 |
| Saturday | 2/14 | 0 | 0 | 1/1 |
| Sunday | 1/14 | 0 | 0 | 0 |

| Season | Ontime | Late | Very Late | Cancelled |
|--------|--------|------|-----------|-----------|
| Spring | 4/14 | 0 | 0 | 1/1 |
| Summer | 6/14 | 0 | 0 | 0 |
| Autumn | 2/14 | 0 | 1/3 | 0 |
| Winter | 2/14 | 2/2 | 2/3 | 0 |

| Fog | Ontime | Late | Very Late | Cancelled |
|--------|--------|------|-----------|-----------|
| None | 5/14 | 0 | 0 | 0 |
| Normal | 5/14 | 1/2 | 2/3 | 0 |
| High | 4/14 | 1/2 | 1/3 | 1/1 |

| Rain | Ontime | Late | Very Late | Cancelled |
|--------|--------|----------------|-----------|-----------|
| None | 6/14 | 1/2 | 1/3 | 0 |
| Slight | 6/14 | 1/2 | 0 | 0 |
| Heavy | 2/14 | 2/0 | 2/3 | 1/1 |

Case 1:- Ontime.

$$\begin{aligned}
 P(\text{OnTime}, A) &= P(\text{ontime}) P(\text{Weekday, Ontime}) P(\text{Winter, Ontime}) \\
 &\quad P(\text{High, Ontime}) P(\text{None, Ontime}) \\
 &= \frac{14}{20} \times \frac{9}{14} \times \frac{2}{14} \times \frac{4}{14} \times \frac{6}{14} \\
 &= 7.871 \times 10^{-3}
 \end{aligned}$$

Case 2:- Late

$$\begin{aligned}
 P(\text{Late}, A) &= P(\text{Late}) P(\text{Weekday, Late}) P(\text{Winter, Late}) \\
 &\quad P(\text{High, Late}) P(\text{None, Late}) \\
 &= \frac{2}{20} \times \frac{1}{2} \times \frac{2}{2} \times \frac{1}{2} \times \frac{1}{2} \\
 &= 0.01250
 \end{aligned}$$

Case 3:- Very Late.

$$\begin{aligned}
 P(\text{V.Late}, A) &= P(\text{VLate}) P(\text{Weekday, VLate}) P(\text{Winter, VLate}) \\
 &\quad P(\text{High, VLate}) P(\text{None, VLate}) \\
 &= \frac{3}{20} \times \frac{3}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \\
 &= 0.01111
 \end{aligned}$$

Case 4:- Cancelled

$$\begin{aligned}
 P(\text{Cancel}, A) &= P(\text{Cancel}) P(\text{Weekday, Cancel}) P(\text{Winter, Cancel}) \\
 &\quad P(\text{High, Cancel}) P(\text{None, Cancel}) \\
 &= \frac{1}{20} \times 0 \times 0 \times \frac{1}{1} \times 0 \\
 &= 0
 \end{aligned}$$

As Case 2 (Late) has the highest probability out of the 4, the instance A will be classified as "Late".

2) To check if the given attributes "Gender" and "Reading" have no correlation. and are hence independent.

→ Null hypothesis:- "Gender" and "Reading" are uncorrelated and independent.

We will use χ^2 test.

$$\begin{aligned}
 \chi^2 &= \sum_{i=1}^m \sum_{j=1}^n \frac{(a_{ij} - e_{ij})^2}{e_{ij}} \\
 &= \frac{(250 - 90)^2}{90} + \frac{(50 - 210)^2}{210} + \frac{(200 - 360)^2}{360} \\
 &\quad + \frac{(1000 - 840)^2}{840} \\
 &= 570.93
 \end{aligned}$$

As $\chi^2 = 570.93 \gg 2.706$,

"Gender" and "Reading" are actually correlated
 So we can safely reject our null hypothesis.
 at confidence level 0.1.