

★ Name: Tanaz Pathan
ID: 202318056

★ Example:

| | category | Documents |
|----------|----------|---------------------------------------|
| Training | - | just plain boring |
| | - | Entirely predictable and lacks energy |
| | - | no surprises and very few laughs |
| | + | very powerful |
| | + | The most fun film of the summer |
| Test | ? | predictable with no fun. |

⇒ vocabulary: { just, plain, boring, entirely, predictable, and, lacks, energy, no, surprises, very, few, laughs, powerful, the, most, fun, film, of, summer }

→ Prior from training:

$$P(c_j) = \frac{N_{c_j}}{N_{\text{total}}} \Rightarrow P(-) = \frac{3}{5} \text{ and } P(+) = \frac{2}{5}$$

→ Dropping with as it is not there in the training document.

⇒ Likelihoods:-

$$P(w_j | c) = \frac{\text{count}(w_j, c) + 1}{(\sum_{w \in V} \text{count}(w, c) + |V|)}$$

$$P(\text{'Predictable' | -}) = \frac{1+1}{14+20} = \frac{2}{34} = \frac{1}{17}$$

$$P(\text{'Predictable' | +}) = \frac{0+1}{9+20} = \frac{1}{29}$$

$$P(\text{'no' | -}) = \frac{1+1}{14+20} = \frac{2}{34} = \frac{1}{17}$$

$$P(\text{'no' | +}) = \frac{0+1}{9+20} = \frac{1}{29}$$

$$P(\text{'fun' | -}) = \frac{0+1}{14+20} = \frac{1}{34}$$

$$P(\text{'fun' | +}) = \frac{1+1}{9+20} = \frac{2}{29}$$

⇒ Scoring:-

$$P(-|S) = P(-) P(S|-)$$

$$= \frac{3}{5} \times \frac{2}{34} \times \frac{2}{34} \times \frac{1}{34} = 0.00006$$

$$P(+|S) = P(+) P(S|+)$$

$$= \frac{2}{5} \times \frac{1}{29} \times \frac{1}{29} \times \frac{2}{29} = 0.000032$$

Thus, we conclude that the test set is of negative category as

$$P(-|S) > P(+|S)$$