Name: Prince Titiya
ID: 202318010
Subject: Big data processing

cat	Document
Training { −	Just plain boring
−	entirely predictable and lack energy
−	no surprises and very few laughs
+	Very powerful
+	the most fun film of the summer

| Test → ? | predictable with no fun |

① Prior from training:

$$P(G) = \frac{N_{cj}}{N_{total}} \implies \begin{array}{l} P(-) = 3/5 \\ P(+) = 2/5 \end{array}$$

② Likelyhood from training

$$P(V_i \mid c) = \frac{Count(V_i \mid c) + 1}{\left(\sum_{wev} Count(w, c)\right) + |V|}$$

$|V| = 20$ (unique words)

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

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

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

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

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

Here, we have drop 'with' because 'with' isn't present in the training set.

③ Scoring test set:

$$P(-) \cdot P(Prod|-) \cdot P(No|-) \cdot P(fun|-)$$

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

$$= 6.106 \times 10^{-5}$$

$$P(+) \cdot P(Prod|+) \cdot P(No|+) \cdot P(fun|+)$$

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

$$= 3.2 \times 10^{-5}$$