Name: Patel Ayush
ID: 20231036

Big Data.

	Cat	Document
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.

① Prior from training.

$$P(c_j) = \frac{N_{c_j}}{N_{total}} \Rightarrow \begin{matrix} P(-) = 3/5 \\ P(+) = 2/5. \end{matrix}$$

② Likelihood from training

$$P(w_i|c) = \frac{count(w_i, c) + 1}{\left(\sum_{w \in V} count(w, c)\right) + |V|}$$

Here $|V| = 20$. (i.e unique words).

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

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

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

$$P(no/-) = \frac{1+1}{14+20} = \frac{1}{17}.$$

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

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

Here we have Drop 'with' because 'with' is not present in the training set.

(3) Scoring test set :

$$P(-) \cdot P(Pred/-) \cdot P(No/-) \cdot P(fun/-)$$

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

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

$$P(+) \cdot P(Pred/+) \cdot P(No/+) \cdot P(fun/+)$$

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

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