Name: Siddharth Kadam 2023IB015 BigData

Cat	Document	
-	just plain boring	Train
-	entirely predictable and lacks energy	
-	no surprises and very few laughs	
+	very powerful	
+	the most fun film of the summer.	
?	predictable with no fun	Test

1) prior from train: $P(C_j) = \frac{N_{cj}}{\sum N}$ => $P(-) = 3/5$, $P(+) = 2/5$

2) Likelihood from training:

$$P(w_i/c) = \frac{\text{count}(w_i, c) + 1}{\sum_{w \in V} \text{count}(w, c) + |V|}$$ * V is 20 unique words.

$P(\text{predictable}|+) = \frac{0+1}{9+20} = \frac{1}{29}$ $P(\text{predictable}|-) = \frac{1+1}{14+20} = \frac{2}{34} = \frac{1}{17}$

$P(\text{fun}|+) = \frac{1+1}{9+20} = \frac{1}{29}$ $P(\text{fun}|-) = \frac{0+1}{14+20} = \frac{1}{34}$

$P(\text{no}|+) = \frac{0+1}{9+20} = \frac{1}{29}$ $P(\text{no}|-) = \frac{1+1}{14+20} = \frac{1}{17}$

3. Scoring test set

$P(-) \cdot P(\text{pred}|-) \cdot P(\text{No}|-) \cdot P(\text{fun}|-)$

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

$P(+) \cdot P(\text{predictable}|+) \cdot P(\text{No}|+) \cdot P(\text{fun}|+)$

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