Name: Mahmood Tophiwala Student ID: 20231803O

BD Assignment

	Document	Category
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	?

A) 1. Priors from training

$$P(C_j) = \frac{N_{C_j}}{N_{total}} \quad ; \quad P(-) = \frac{3}{5}$$

$$P(+) = \frac{2}{5}$$

→ Dropping 'with' as isn't there in the training document.

2. Likelihoods

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

$$p('predictable'/-) = \frac{1+1}{14+20} = \frac{2}{34} = \frac{1}{17}$$

$$p('predictable'/+) = \frac{0+1}{9+20} = \frac{1}{29}$$

$$p('no'/-) = \frac{1+1}{14+20} = \frac{2}{34} = \frac{1}{17}$$

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

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

$$p('fun'/+) = \frac{1+1}{9+20} = \frac{2}{29}$$

3 Scoring

$$p(-)\,p(S/-) = \frac{3}{5} \times \frac{2 \times 2 \times 1}{34^3} \approx 0.00006$$

$$p(+)\,p(S/+) = \frac{2}{5} \times \frac{1 \times 1 \times 2}{29^3} \approx 0.000032$$

∴ Thus, we conclude, the test set is of Negative Category as,
$p(-)p(S/-) > p(+)\,p(S/+)$.