3)

3. Given the below documents.

texts = [
    "good movie", "not a good movie", "did not like",
    "i like it", "good one"
]

Given the definition of TF and IDF, what is the **sum** of TF-IDF values for 1-grams in "good movie" text? Enter a math expression as an answer.

for the document 'good movie' there are 2 tokens → 'good' + 'movie'.

$$TF = \text{term frequency} = f_{t,d} / \sum f_{t',d}$$

$$idf = \text{inverse document frequency} = \ln \frac{N}{1 + |\{d \in D : t \in d\}|}$$

Token 'good'

$$TF = \frac{1}{2} = 0.5$$

$$idf = \ln(5/3) = 0.512$$

$$TF \cdot idf = 0.5 \cdot 0.512 = 0.256$$

Token 'movie'

$$TF = \frac{1}{2} = 0.5$$

$$idf = \ln(5/2) = 0.916$$

$$TF \cdot idf = 0.5 \cdot 0.916 = 0.458$$

$$\text{Sum of } TF \cdot idf \text{ values} = \boxed{\frac{1}{2} \cdot \ln(5/3) + \frac{1}{2} \cdot \ln(5/2) = 0.714}$$