Given the following three review texts and their class labels:

ID	Input review text	Class label
1	Good, thanks	Positive
2	No impressive, thanks	Negative
3	Impressive good	Positive

Determine the class label of the 4-th review text “No, thanks” using the Naïve Bayesian and *k*-NN (*k*=1) classifiers, respectively.

In the pre-processing step, all lower-case words were extracted, and all punctuations were discarded from all texts, as follows:

ID	good	thanks	no	impressive	Class label
1	1	1	0	0	Positive
2	0	1	1	1	Negative
3	1	0	0	1	Positive
4	0	1	1	0	?

(1) *P*(Class label=“Positive”|ID=4)

= *P*(Class label=“Positive”)*P*(ID=4|Class label=“Positive”) / *P*(ID=4)
= *P*(Class label=“Positive”)*P*(“thanks”, “no”|Class label=“Positive”) / *P*(ID=4)

According to the assumption of the Naïve Bayesian classifier,

P(“thanks”, “no”|Class label=“Positive”)
= *P*(“thanks” |Class label=“Positive”)*P*(“no”|Class label=“Positive”)

Based on the training set (three review text with ID equal to 1, 2, 3 and their class labels), we have:

P(Class label=“Positive”) = 2/3,
P(“thanks”|Class label=“Positive”) = 1/4,

P(“no”|Class label=“Positive”) = 0.

Thus,

P(Class label=“Positive”|ID=4) = (2/3) * (1/4) * 0 / *P*(ID=4) = 0 / *P*(ID=4)

Similarly,

P(Class label=“Negative”|ID=4)
= *P*(Class label=“Negative”)*P*(ID=4|Class label=“Negative”) / *P*(ID=4)
= *P*(Class label=“Negative”)*P*(“thanks”, “no”|Class label=“Negative”) / *P*(ID=4)
= (1/3) * (1/3) * (1/3) / *P*(ID=4) = (1/27) / *P*(ID=4)

Since *P*(ID=4) > 0,

P(Class label=“Negative”|ID=4) > *P*(Class label=“Positive”|ID=4)

Thus, we assign “Negative” to the review text with ID equal to 4.

(2) We can use the Euclidean distance to measure the dissimilarity between paired texts:

$$d(\text{ID=4, ID=1}) = \sqrt{(0-1)^2+(1-1)^2+(1-0)^2+(0-0)^2} = \sqrt{2}$$

$$d(\text{ID=4, ID=2}) = \sqrt{(0-0)^2+(1-1)^2+(1-1)^2+(0-1)^2} = 1$$

$$d(\text{ID=4, ID=3}) = \sqrt{(0-1)^2+(1-0)^2+(1-0)^2+(0-1)^2} = 2$$

For the review text with ID equal to 4, the review text with ID equal to 2 (whose class label is “Negative”) is the most similar text. Thus, we assign “Negative” to the review text with ID equal to 4 according to the *k*-NN (*k*=1) classifier.