✖

Choosing features and metrics for nearest neighbor search

5 questions

1
point

1.

Among the words that appear in both Barack Obama and Francisco Barrio, take the 5 that appear most frequently in Obama. How many of the articles in the Wikipedia dataset contain all of those 5 words?

56066

1
point

2.

Measure the pairwise distance between the Wikipedia pages of Barack Obama, George W. Bush, and Joe Biden. Which of the three pairs has the smallest distance?

◯ Between Obama and Biden

◯ Between Obama and Bush

◯ Between Biden and Bush

1
point

3.

Collect all words that appear both in Barack Obama and George W. Bush pages.
Out of those words, find the 10 words that show up most often in Obama's page.
Which of the following is NOT one of the 10 words?

○   the

○   presidential

○   in

○   act

○   as

| 1 |
| point |

4.

Among the words that appear in both Barack Obama and Phil Schiliro, take the 5
that have largest weights in Obama. How many of the articles in the Wikipedia
dataset contain all of those 5 words?

14

| 1 |
| point |

5.

Compute the Euclidean distance between TF-IDF features of Obama and Biden.
Round your answer to 3 decimal places. Use American-style decimals (e.g.
110.921).

123.297

Submit Quiz