**Quiz 4**

**Name:**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **abc123:**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. **(5 points):** Given the following dataset, what is the bag-of-words feature matrix X that would be generated? Assume the use of the “CountVectorizer()” with only unigrams. No code is required, simply draw the matrix and fill in the cells. Make sure the columns are clearly annotated.

**Data**

the cat is fat fat fat fat

the dog is cute

my dog is the best

1. **(5 points):** What is the relation between “term frequency” and “document frequency” with TF-IDF weighting? Why is it useful?