1. Calculate the bag of words (without TF-IDF, just word counts) matrix for the following 5 phrases.

• college student

• airport taxi driver

• Toronto art museum

• life art

• machine learning student

Hints: How many rows should be there? How many columns should be there?

5 rows and 11 columns.

2. Calculate the IDF for the words “college” and “art”.

IDF of ‘college’

IDF = ()

IDF = log(

IDF = (

=

= 0.698

IDF of ‘art’

IDF = ()

IDF = log(

IDF = (

= 0.397