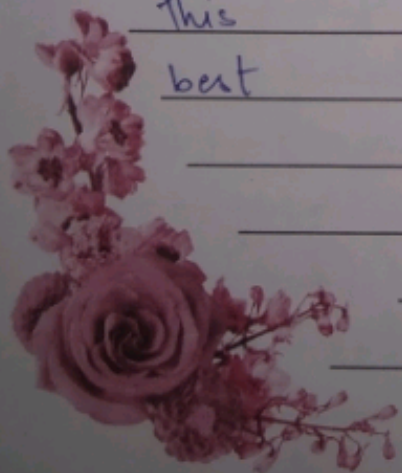


Date \_\_\_\_\_

## Bag of Word

Token	S1	S2	S3	Total
data	1	1	2	4
Science	2	1	0	3
is	1	1	0	2
one	1	1	0	2
of	1	1	0	2
the	1	1	1	3
most	1	0	0	1
important	1	1	0	2
courses	1	0	0	1
in	1	0	0	1
computer	0	0	1	1
scientist	0	0	1	1
perform	0	0	1	1
problems	0	1	0	1
analysis	0	1	0	1
this	0	1	0	1
best	0	1	0	1



# Term Frequency

	S1	S2	S3
data	$\frac{1}{2}$	$\frac{1}{9}$	$\frac{2}{6}$
science	$\frac{2}{12}$	$\frac{1}{9}$	0
is	$\frac{1}{12}$	$\frac{1}{9}$	0
one	$\frac{1}{12}$	$\frac{1}{9}$	0
of	$\frac{1}{12}$	$\frac{1}{9}$	0
the	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{1}{6}$
most	$\frac{1}{12}$	0	0
important	$\frac{1}{12}$	0	0
courses	$\frac{1}{12}$	$\frac{1}{9}$	0
in	$\frac{1}{12}$	0	0
computer	0	0	0
scientist	0	0	$\frac{1}{6}$
perform	0	0	$\frac{1}{6}$
analysis	0	$\frac{1}{9}$	$\frac{1}{6}$
this	0	$\frac{1}{9}$	0
best	0	$\frac{1}{9}$	0



Date \_\_\_\_\_

## Inverse Document frequency

Token

IDF

data

$$\log(3/3) = 0$$

science

$$\log(3/2) = 0.176$$

is

$$\log(3/2) = 0.176$$

one

$$\log(3/2) = 0.176$$

of

$$\log(3/2) = 0.176$$

the

$$\log(3/2) = 0.176$$

most

$$\log(3/1) = 0.477$$

important

$$\log(3/2) = 0.176$$

courses

$$\log(3/1) = 0.477$$

in

$$\log(3/1) = 0.477$$

computer

$$\log(3/1) = 0.477$$

scientist

$$\log(3/1) = 0.477$$

perform

$$\log(3/1) = 0.477$$

analysis

$$\log(3/1) = 0.477$$

this

$$\log(3/1) = 0.477$$

best

$$\log(3/1) = 0.477$$



# TF \* IDF

	S1	S2	S3
data	0	0	0
science	0.4981	0.3216	0
is	0.0146	0.3216	0
one	0.2496	0.3216	0
of	0.2496	0.3216	0
<del>best</del> the	0.1934	0.2498	0.2711
most	0.3274	0	0
important	0.3274	0	0
courses	0.2490	0.3216	0
in	0.3274	0	0
computer	0.3274	0	0
scientist	0	0	0.4591
perform	0	0	0.4591
analysis	0	0	0.4591
this	0	0.4229	0
best	0	0.4229	0