

FAHAD REHMAN

FA18-BLS-193

Intro to Data Science

Assignment-4

Group - II

Q #01

S1: [Sunshine, state, enjoy, sunshine]

S2: [brown, fox, jump, high, brown, fox,]

S3: [sunshine, state, fox, run, fast]

Bag of Words (frequency)

	Sunshine	state	enjoy	brown	fox	jump	high	run	fast
S1:	2	1	1	0	0	0	0	0	0
S2:	0	0	0	2	2	1	1	1	0
S3:	1	1	0	0	1	0	0	1	1

Term Frequency (frequency / Total terms)

~~tf-s1~~

	Sunshine	state	enjoy	brown	fox	jump	high	run	fast
H-S1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	0	0	0	0	0	0
H-S2	0	0	0	$\frac{2}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	0
H-S3	$\frac{1}{5}$	$\frac{1}{5}$	0	0	$\frac{1}{5}$	0	0	$\frac{1}{5}$	$\frac{1}{5}$

Term-Frequency x Inverse Document Frequency

$$Tf \cdot idf \Rightarrow Tf \times \log \left(\frac{\text{total number of Documents}}{\# \text{ of Documents containing the term}} \right)$$

	Sunshine state enjoy brown fox jump high run fast									
$Tf \cdot idf \cdot R1$	0.37	0.38	0.50	0	0	0	0	0	0	0
R2	0	0	0	0.61	0.51	0.33	0.33	0.25	0	0
R3	0.41	0.41	0	0	0.41	0	0	0.41	0.54	0