Sanjana Kothari
016721679

Vector Space Model

d1: User selected wedding gown
d2: User ordered online rose flowers
d3: User searched diamond ring
d4: User selected white wedding gown, online flowers, 3 carat diamond ring

→ Vocabulary: { user, selected, wedding, gown, ordered, online, rose, flowers, searched, diamond ring, white, 3, carat }

Term frequency (tf) = $\frac{\text{no. of times a word appears in a doc}}{\text{no. of words in the document}}$

Inverse document frequency (idf) = $\log\left(\frac{\text{no. of documents}}{\text{no. of times word appears across all docs}}\right)$

words	tf d1	tf d2	tf d3	tf d4	idf
user	1/4	1/5	1/4	1/11	$\log_2 4/4 = 0$
selected	1/4	0	0	1/11	$\log_2 4/2 = 1$
wedding	1/4	0	0	1/11	$\log_2 4/2 = 1$
gown	1/4	0	0	1/11	$\log_2 4/2 = 1$
ordered	0	1/5	0	0	$\log_2 4/1 = 2$
online	0	1/5	0	1/11	$\log_2 4/2 = 1$
rose	0	1/5	0	0	$\log_2 4/1 = 2$
flowers	0	1/5	0	1/11	$\log_2 4/2 = 1$
searched	0	0	1/4	0	$\log_2 4/1 = 2$ ✓
diamond	0	0	1/4	1/11	$\log_2 4/2 = 1$
ring	0	0	1/4	1/11	$\log_2 4/2 = 1$
white	0	0	0	1/11	$\log_2 4/1 = 2$
3	0	0	0	1/11	$\log_2 4/1 = 2$
carat	0	0	0	1/11	$\log_2 4/1 = 2$

Representation of vector model = tf * idf.

	user	selected	wedding	gown	ordered	online	rose	flowers	searched	diamond	ring	white	3	carat
d1	0	1/4	1/4	1/4	0	0	0	0	0	0	0	0	0	0
d2	0	0	0	0	2/5	1/5	2/5	1/5	0	0	0	0	0	0
d3	0	0	0	0	0	0	0	0	2/4	1/4	1/4	0	0	0
d4	0	1/11	1/11	1/11	0	1/11	0	1/11	0	1/11	1/11	2/11	2/11	2/11