

NAMA: NANDA PRAYOGI

NIM: A11.2021.13505

MATA KULIAH: SISTEM TEMU KEMBALI INFORMASI

KELOMPOK: A11.4703

STATUS: KUIS LATIHAN SOAL

DOKUMEN	TEKS
D1	PSIS berburu juara Liga Indonesia
D2	Hasil putusan Sidang Elit Politik
D3	Partai politik berebut suara
D4	Manchester United Juara Liga Inggris
D5	Timnas Indonesia juara Liga AFC

Kelompokan dokumen tersebut menjadi 2 cluster, dengan ketentuan:

- Menggunakan teknik k-means clustering.
- Similarity measure dengan Euclidean distance.
- Ditentukan Centroid D1 dan D3.
- Term weighting menggunakan TFIDF.
- Preprocessing: tokenization, stemming.

LANGKAH 1 PREPROCESSING

Hasil: D1: psis buru juara liga indonesia

D2: hasil putus sidang elit politik

D3: partai politik rebut suara

D4: manchest unit juara liga inggris

D5: timnas indonesia juara liga afc

LANGKAH 2 TERM WEIGHTING

psis buru juara liga indonesia hasil putus sidang elit politik partai rebut suara manchest unit inggris timnas afc

	D1	D2	D3	D4	D5	DF	IDF
psis	1					1	0,69897
buru	1					1	0,69897
juara	1			1	1	3	0,22185

liga	1			1	1	3	0,22185
indonesia	1				1	2	0,39794
hasil		1				1	0,69897
putus		1				1	0,69897
sidang		1				1	0,69897
elit		1				1	0,69897
politik		1	1			2	0,39794
partai			1			1	0,69897
rebut			1			1	0,69897
suara			1			1	0,69897
manchest				1		1	0,69897
unit				1		1	0,69897
inggris				1		1	0,69897
timnas					1	1	0,69897
afc					1	1	0,69897

LANGKAH 4 TFIDF

	D1	D2	D3	D4	D5
psis	0,69897				
buru	0,69897				
juara	0,22185			0,22185	0,22185
liga	0,22185			0,22185	0,22185
indonesia	0,39794				0,39794
hasil		0,69897			
putus		0,69897			
sidang		0,69897			
elit		0,69897			
politik		0,39794	0,39794		
partai			0,69897		
rebut			0,69897		
suara			0,69897		
manchest				0,69897	
unit				0,69897	
inggris				0,69897	
timnas					0,69897
afc					0,69897

LANGKAH 5 PROSES CLUSTER

D1 sebagai titik pusat kluster 1, dan D3 sebagai titik pusat kluster 2

	C1(D1)	C3(D3)
D1	0	1,690544291
D2	1,829344145	1,849299868
D3	1,690544291	0
D4	1,61280811	1,785539134
D5	1,397941344	1,690544291

HASIL

	C1(D1)	C3(D3)
D1	$\sqrt{}$	-
D2	$\sqrt{}$	-
D3	-	$\sqrt{}$
D4	$\sqrt{}$	-
D5	$\sqrt{}$	-