

---

## **Text Mining Untuk Clustering Buku Di Perpustakaan Menggunakan Metode K-Means**

**Alfathan Anggi Riyanto<sup>1\*</sup>, Daryanto<sup>1</sup>, Ginanjar Abdurrahman<sup>1</sup>**

Universitas Muhammadiyah Jember.

email: anggiriyanto3@gmail.com

DOI: <https://doi.org/10.32528/nms.v1i6.239>

\*Correspondensi: Alfathan Anggi Riyanto

Email: anggiriyanto3@gmail.com

Published: November, 2022



**Copyright:** © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

**Abstrak:** Perpustakaan merupakan sarana untuk memberikan layanan informasi untuk semua kalangan dengan tujuan mencerdaskan masyarakat dan memudahkan dalam memberikan informasi. Selama ini perpustakaan pusat Universitas Muhammadiyah Jember mengelompokkan data tugas akhir secara manual, karena belum adanya sistem untuk pengelompokan tugas akhir berdasarkan jenisnya, sehingga terkadang ada kendala atau membutuhkan waktu yang lebih lama dalam proses pencarian informasi tentang data tugas akhir mahasiswa. Oleh karena itu, pengelompokan data tugas akhir membutuhkan teknologi sistem cluster yang lebih efektif dan efisien. Sistem cluster yang diterapkan untuk mengelompokkan data-data tugas akhir bersifat sederhana dengan menggunakan metode K-Means. Data tugas akhir yang diperoleh pada tahun 2018 berjumlah 95 dokumen. Berdasarkan validitas cluster dengan perbandingan hasil DBI menggunakan 2 cluster sampai dengan 8 cluster di peroleh hasil DBI tertinggi pada percobaan dengan menggunakan 4 Cluster dengan memperoleh nilai DBI 0,87.

**Keywords:** Perpustakaan, Tugas Akhir, *Text Mining*, *cluster*, metode *K-Means*

---

### **PENDAHULUAN**

Perpustakaan pusat di unit Universitas Muhammadiyah Jember merupakan sarana untuk memberikan layanan informasi dan fasilitas sumber referensi bagi mahasiswa-mahasiswi dan dosen. Unit perpustakaan memiliki tujuan menyediakan layanan sumber referensi dan informasi terkait aktivitas akademik di kampus. Selama ini perpustakaan pusat Universitas Muhammadiyah Jember masih menerapkan pengelompokan buku dengan cara manual, sehingga butuh waktu lebih lama dalam pencarian informasi data-data referensi buku. Faktor kuantitas buku yang dimiliki perpustakaan dan kendala dalam hal mengelompokkan data-data sumber referensi menyebabkan timbulnya kebutuhan teknologi sistem cluster yang lebih efektif dan efisien. Sistem pengelompokan tersebut nantinya akan berguna untuk memudahkan dalam menemukan informasi yang dibutuhkannya di dalam perpustakaan.

Teknik pengelompokan banyak digunakan untuk mengatasi problem pengelompokan dokumen yang memiliki kemiripan data dari dokumen tersebut. Tujuan dari metode pengelompokan adalah cluster data ke dalam kelompok yang memiliki kesamaan karakteristik. Pada pengelompokan sekumpulan data yang belum diberi label akan dikelompokkan sesuai dengan karakteristik data yang dimiliki setiap data tersebut. Salah satu metode yang tepat untuk sistem pengelompokan data di perpustakaan adalah metode K-Means yang sifatnya yang relatif cepat. Algoritma K-Means merupakan pengelompokan dengan partisi set data ke dalam kelompok atau cluster yang sudah ditetapkan. Pemilihan jumlah kelompok (*cluster*) sebagai nilai pusat cluster awal juga mempengaruhi hasil pengelompokan. Sedangkan menurut Alfiana, Santoso dan Ali Ridho B (2012) berpendapat bahwa metode K-means merupakan metode pengelompokan yang sederhana.

Berdasarkan uraian diatas dan hasil studi pustaka penelitian sebelumnya, maka peneliti memilih judul penelitian “Text Mining Untuk Clustering Buku Di Perpustakaan Menggunakan Metode K-Means”.

## METODE PENELITIAN

### A.Contoh Penerapan Algoritma

Tahapan pertama dalam penelitian ini yang harus dilewati adalah tahapan *preprocessing text*. Pada penelitian ini akan dilakukan perhitungan terhadap 10 data percobaan. Pada 10 judul buku yang di proses, terdapat 38 kata, setelah dilakukan tahapan preprocessing dokumen berupa *case folding*, *tokenization*, *filtering*, dan terakhir *stemming* total tetap 38 kata dengan total term yang akan diproses lebih lanjut adalah sebanyak 38 kata-kata yang unik.

NO	JUDUL
1	Alternatif Pemilihan Strategi Pemasaran Berdasarkan Analisis Swot Pada Cv. Bimo Krisno Tour <u>And</u> Travel Jember
2	Cluster Karakter Pengguna Batik Untuk Rekomendasi Motif Menggunakan Algoritma Naive Bayes
3	Sistem Pendukung Keputusan Penerimaan Pegawai Baru Menggunakan Metode Analytical Hierarchy Process
4	Menentukan Topik Dari Sebuah Abstrak Tugas Akhir Menggunakan Jaccard Measure
5	Pengaruh Harga, Citra Merek, Dan Atribut Produk Terhadap Minat Beli Produk Smartphone Oppo (Study Kasus Di Toko Bismar Ponsel Jember)
6	Analisis Pengaruh Kualitas Layanan, Komitmen Pelanggan Dan Kepercayaan Terhadap Loyalitas Pelanggan Swalayan Larisso Ambulu Jember

Tabel 1. Data Uji Coba a

NO	JUDUL
7	Analisis Pengaruh Ekuitas Merek Terhadap Keputusan Pembelian Sepeda Motor <u>Honda</u> (Study Kasus Pada Dealer Sjms Motor Pesanggaran)
8	Pengaruh Ukuran Perusahaan, Umur Perusahaan, Profitabilitas, Dan Solvabilitas Terhadap Audit Delay (Studi Pada Perusahaan Perbankan Yang Terdaftar Di Bei)
9	Respons Pertumbuhan Kurma Terhadap Berbagai Konsentrasi Ba Dan Ga3 Dalam Kultur In Vitro
10	Respon Produktifitas Okra (Abelmoschus Esculentus) Terhadap Pemberian Dosis Pupuk Petroganik Dan Pupuk N

[Sumber: Data Set]

Tabel 2. Hasil Tokenizing

No	Word	No	Word	No	Word
1	alternatif	28	sistem	55	produk
2	pemilihan	29	pendukung	56	terhadap
3	strategi	30	keputusan	57	minat
4	pemasaran	31	penerimaan	58	beli
5	berdasarkan	32	pegawai	59	produk
6	analisis	33	baru	60	smartphone
7	swot	34	menggunakan	61	oppo
8	pada	35	metode	62	study
9	cv	36	analytical	63	kasus
10	bimo	37	hierarchy	64	di
11	krisno	38	process	65	toko
12	tour	39	menentukan	66	bismar
13	and	40	topik	67	ponsel
14	travel	41	dari	68	jember
15	jember	42	sebuah	69	analisis
16	cluster	43	abstrak	70	pengaruh
17	karakter	44	tugas	71	kualitas
18	pengguna	45	akhir	72	layanan
19	batik	46	menggunakan	73	komitmen
20	untuk	47	jaccard	74	pelanggan
21	rekomendasi	48	measure	75	dan
22	motif	49	pengaruh	76	kepercayaan
23	menggunakan	50	harga	77	terhadap
24	algoritma	51	citra	78	loyalitas
25	na	52	merek	79	pelanggan
26	ve	53	dan	80	swalayan
27	bayes	54	atribut	81	larisso

No	Word	No	Word	No	Word
82	ambulu	104	umur	126	ba
83	jember	105	perusahaan	127	dan
84	analisis	106	profitabilitas	128	ga
85	pengaruh	107	dan	129	dalam
86	ekuitas	108	solvabilitas	130	kultur
87	merek	109	terhadap	131	in
88	terhadap	110	audit	132	vitro
89	keputusan	111	delay	133	respon
90	pembelian	112	studi	134	produktifitas
91	sepeda	113	pada	135	okra
92	motor	114	perusahaan	136	abelmoschus
93	honda	115	perbankan	137	esculentus
94	study	116	yang	138	terhadap
95	kasus	117	terdaftar	139	pemberian
96	pada	118	di	140	dosis
97	dealer	119	bei	141	pupuk
98	sjms	120	respons	142	petrogekanik
99	motor	121	pertumbuhan	143	dan
100	pesangaran	122	kurma	144	pupuk
101	pengaruh	123	terhadap	145	n
102	ukuran	124	berbagai		
103	perusahaan	125	konsentrasi		

[Sumber: Hasil Perhitungan]

Tabel 3. Hasil *Stopword*

No	Word	No	Word	No	Word	No	Word				
1	alternatif	13	travel	25	bayes	37	topik	66	pelanggan	95	solvabilitas
2	pemilihan	14	jember	26	sistem	38	abstrak	67	kepercayaan	96	audit
3	strategi	15	cluster	27	pendukung	39	tugas	68	loyalitas	97	delay
4	pemasaran	16	karakter	28	keputusan	40	akhir	69	pelanggan	98	studi
5	berdasarkan	17	pengguna	29	penerimaan	41	menggunakan	70	swalayan	99	perusahaan
6	analisis	18	batik	30	pegawai	42	jaccard	71	larisso	100	perbankan
7	swot	19	rekomendasi	31	menggunakan	43	measure	72	ambulu	101	terdaftar
8	cv	20	motif	32	metode	44	pengaruh	73	jember	102	bei
9	bimo	21	menggunakan	33	analytical	45	harga	74	analisis	103	respons
10	krisno	22	algoritma	34	hierarchy	46	citra	75	pengaruh	104	pertumbuhan
11	tour	23	na	35	process	47	merek	76	ekuitas	105	kurma
12	and	24	ve	36	menentukan	48	atribut	77	merek	106	konsentrasi
						49	produk	78	keputusan	107	ba
						50	minat	79	pembelian	108	ga
						51	beli	80	sepeda	109	kultur
						52	produk	81	motor	110	in
						53	smartphone	82	honda	111	vitro
						54	oppo	83	study	112	respon
						55	study	84	kasus	113	produktifitas
						56	kasus	85	dealer	114	okra
						57	toko	86	sjms	115	abelmoschus
						58	bismar	87	motor	116	esculentus
						59	ponsel	88	pesangaran	117	pemberian
						60	jember	89	pengaruh	118	dosis
						61	analisis	90	ukuran	119	pupuk
						62	pengaruh	91	perusahaan	120	petrogekanik
						63	kualitas	92	umur	121	pupuk
						64	layanan	93	perusahaan		
						65	komitmen	94	profitabilitas		

[Sumber: Hasil Perhitungan]

Tabel 4. Hasil Stemmer Nazief Adriani

No	Word	No	Word	No	Word
1	beri	6	komitmen	11	terima
2	citra	7	loyalitas	12	usaha
3	dukung	8	percaya	13	swot
4	guna	9	pilih	14	cv
5	kasus	10	putus	15	bimo

No	Word	No	Word	No	Word
16	krisno	27	jaccard	38	delay
17	tour	28	measure	39	ga
18	and	29	smartphone	40	in
19	travel	30	study	41	vitro
20	karakter	31	bismar	42	respon
21	algoritma	32	ponsel	43	produktifitas
22	na	33	larisso	44	okra
23	bayes	34	ambulu	45	abelmoschus
24	analytical	35	honda	46	esculentus
25	hierarchy	36	dealer	47	petrogeanik
26	process	37	sjms		

[Sumber: Hasil Perhitungan]

Tahapan selanjutnya adalah proses perkalian antara TF dengan nilai IDF. Berikut hasil proses perhitungan nilai Idf:

Term	Tf										DF	TF.IDF	
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10			
beri	0	0	0	0	0	0	0	0	0	0	1	1	1
citra	0	0	0	0	1	0	0	0	0	0	0	1	1
dukung	0	0	1	0	0	0	0	0	0	0	0	1	1
guna	0	2	1	1	0	0	0	0	0	0	0	3	0.5229
kasus	0	0	0	0	1	0	1	0	0	0	0	2	0.699
komitmen	0	0	0	0	0	1	0	0	0	0	0	1	1
loyalitas	0	0	0	0	0	1	0	0	0	0	0	1	1
percaya	0	0	0	0	0	1	0	0	0	0	0	1	1
pilih	1	0	0	0	0	0	0	0	0	0	0	1	1
putus	0	0	1	0	0	0	1	0	0	0	0	2	0.699
terima	0	0	1	0	0	0	0	0	0	0	0	1	1
usaha	0	0	0	0	0	0	0	3	0	0	0	1	1
swot	1	0	0	0	0	0	0	0	0	0	0	1	1
cv	1	0	0	0	0	0	0	0	0	0	0	1	1
bimo	1	0	0	0	0	0	0	0	0	0	0	1	1
krisno	1	0	0	0	0	0	0	0	0	0	0	1	1
tour	1	0	0	0	0	0	0	0	0	0	0	1	1
and	1	0	0	0	0	0	0	0	0	0	0	1	1
travel	1	0	0	0	0	0	0	0	0	0	0	1	1
karakter	0	1	0	0	0	0	0	0	0	0	0	1	1

Tabel 5. Hasil Bobot IDF

Term	Tf										DF	TF.IDF	
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10			
algoritma	0	1	0	0	0	0	0	0	0	0	0	1	1
na	0	1	0	0	0	0	0	0	0	0	0	1	1
bayes	0	1	0	0	0	0	0	0	0	0	0	1	1
analytical	0	0	1	0	0	0	0	0	0	0	0	1	1
hierarchy	0	0	1	0	0	0	0	0	0	0	0	1	1
process	0	0	1	0	0	0	0	0	0	0	0	1	1
jaccard	0	0	0	1	0	0	0	0	0	0	0	1	1
measure	0	0	0	1	0	0	0	0	0	0	0	1	1
smartphone	0	0	0	0	1	0	0	0	0	0	0	1	1
study	0	0	0	0	1	0	1	0	0	0	0	2	0.699
bismar	0	0	0	0	1	0	0	0	0	0	0	1	1
ponsel	0	0	0	0	1	0	0	0	0	0	0	1	1
larisso	0	0	0	0	0	1	0	0	0	0	0	1	1
ambulu	0	0	0	0	0	1	0	0	0	0	0	1	1
honda	0	0	0	0	0	0	1	0	0	0	0	1	1
dealer	0	0	0	0	0	0	1	0	0	0	0	1	1
sjms	0	0	0	0	0	0	1	0	0	0	0	1	1
delay	0	0	0	0	0	0	0	1	0	0	0	1	1
ga	0	0	0	0	0	0	0	0	1	0	0	1	1
in	0	0	0	0	0	0	0	0	1	0	0	1	1
vitro	0	0	0	0	0	0	0	0	1	0	0	1	1
respon	0	0	0	0	0	0	0	0	0	1	0	1	1
produktifitas	0	0	0	0	0	0	0	0	0	1	0	1	1
okra	0	0	0	0	0	0	0	0	0	1	0	1	1
abelmoschus	0	0	0	0	0	0	0	0	0	1	0	1	1
esculentus	0	0	0	0	0	0	0	0	0	1	0	1	1
petrogeanik	0	0	0	0	0	0	0	0	0	1	0	1	1

Term	Tf.idf										
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	
beri	0	0	0	0	0	0	0	0	0	0	1
citra	0	0	0	0	1	0	0	0	0	0	0
dukung	0	0	1	0	0	0	0	0	0	0	0
guna	0	1.05	0.52	0.52	0	0	0	0	0	0	0
kasus	0	0	0	0	0.7	0	0.7	0	0	0	0
komitmen	0	0	0	0	0	1	0	0	0	0	0

[Sumber:Hasil Perhitungan]

Term	Tf.idf									
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
loyalitas	0	0	0	0	0	1	0	0	0	0
percaya	0	0	0	0	0	1	0	0	0	0
pilih	1	0	0	0	0	0	0	0	0	0
putus	0	0	0.7	0	0	0	0.7	0	0	0
terima	0	0	1	0	0	0	0	0	0	0
usaha	0	0	0	0	0	0	0	3	0	0
swot	1	0	0	0	0	0	0	0	0	0
cv	1	0	0	0	0	0	0	0	0	0
bimo	1	0	0	0	0	0	0	0	0	0
krismo	1	0	0	0	0	0	0	0	0	0
tour	1	0	0	0	0	0	0	0	0	0
and	1	0	0	0	0	0	0	0	0	0
travel	1	0	0	0	0	0	0	0	0	0
karakter	0	1	0	0	0	0	0	0	0	0
algoritma	0	1	0	0	0	0	0	0	0	0
na	0	1	0	0	0	0	0	0	0	0
bayes	0	1	0	0	0	0	0	0	0	0
analytical	0	0	1	0	0	0	0	0	0	0
hierarchy	0	0	1	0	0	0	0	0	0	0
process	0	0	1	0	0	0	0	0	0	0
jaccard	0	0	0	1	0	0	0	0	0	0
measure	0	0	0	1	0	0	0	0	0	0
smartphone	0	0	0	0	1	0	0	0	0	0
study	0	0	0	0	0.7	0	0.7	0	0	0
bismar	0	0	0	0	1	0	0	0	0	0
ponsel	0	0	0	0	1	0	0	0	0	0
larisso	0	0	0	0	0	1	0	0	0	0
ambulu	0	0	0	0	0	1	0	0	0	0
honda	0	0	0	0	0	0	1	0	0	0
dealer	0	0	0	0	0	0	1	0	0	0
sjms	0	0	0	0	0	0	1	0	0	0
delay	0	0	0	0	0	0	0	1	0	0
ga	0	0	0	0	0	0	0	0	1	0
in	0	0	0	0	0	0	0	0	0	1
vitro	0	0	0	0	0	0	0	0	0	1
respon	0	0	0	0	0	0	0	0	0	1
produktifitas	0	0	0	0	0	0	0	0	0	1

  

Term	Tf.idf									
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
okra	0	0	0	0	0	0	0	0	0	1
abelmoschus	0	0	0	0	0	0	0	0	0	1
esculentus	0	0	0	0	0	0	0	0	0	1
petroganik	0	0	0	0	0	0	0	0	0	1

[Sumber: Hasil Perhitungan]

Penentuan cluster dokumen berdasarkan jenis “Pertanian, Teknik dan Ekonomi”, maka diperlukan ciri-ciri dari setiap cluster atau kata kunci dari cluster yang sudah ditentukan untuk keyword setiap cluster sebagai berikut:

Cluster(Pertanian)={ petroganik,produktifitas,okra,abelmoschus, esculentus }

Cluster(Teknik)={ analytical,hierarchy,process, citra, algoritma }

Cluster (Ekonomi) = { larisso, karakter, usaha, dukung, loyalitas }

Berdasarkan data kata kunci diatas, maka tahap berikutnya adalah proses menghitung jarak antar dokumen dengan menggunakan *Euclidean distance*. Berdasarkan kemunculan word pada tiap dokumen, maka dihasilkan seperti pada tabel 6 berikut:

Tabel 6. Cluster Dengan Rumus Eucledia Distance

No	Dokumen	C1	C2	C3	Hasil
1	D1	1	0	1	C2
2	D2	1	1	0	C3
3	D3	1	1	1	C1
4	D4	1	0	1	C2
5	D5	1	1	0	C3
6	D6	0	1	1	C1
7	D7	1	0	0	C2
8	D8	1	1	0	C3
9	D9	0	1	1	C1
10	D10	0	1	1	C1

[Sumber: Hasil Perhitungan]

Cluster (Pertanian) = {D3, D6, D9, D10}

Cluster (Teknik) = {D1, D4, D7}

Cluster (Ekonomi) = {D2, D5, D8}

## HASIL DAN PEMBAHASAN

### A. Hasil Cluster

Halaman cluster adalah halaman hasil dari setiap proses yang terdapat pada sistem clustering yaitu hasil *text preprocessing*, pembobotan dokumen, dan proses clusteringnya.

Gambar 1. Bagian Text Preprocessing Dan Pembobotan Dokumen Pada Halaman Cluster [Sumber: Hasil Halaman Cluster Pada Program]

Gambar 2. Bagian Proses Clustering Dan Hasil Pengelompokan Pada Halaman Cluster [Sumber: Hasil Halaman Cluster Pada Program]

Hasil bagian *text processing* dibagi menjadi 3, yaitu Tokenizing , Stopword Removal , dan Steaming

Tabel 7. Hasil Proses Tokenizing

No	Word
1	sumber
2	daya
3	manusia
4	adalah
5	salah

.	-
.	-
.	-
2933	keadaan

[Sumber: Hasil Perhitungan]

Tabel 8. Hasil Proses *Stopword Removal*

No	Word
1	daya
2	aset
3	terpenting
4	menjalankan
5	tugas
.	-
.	-
.	-
2405	operator

Sumber: Hasil Perhitungan

Tabel 9. Hasil Proses Stemming

No	Word
1	tuju
2	teliti
3	pola
4	salur
5	asar
.	-
.	-
.	-
10	banding

[Sumber: Hasil Perhitungan]

Proses selanjutnya adalah bagian hasil pembobotan ditunjukkan pada tabel 10 berikut:

Tabel 10. Hasil Pembobotan

No	Word	Doc							df	idf
		D1	D2	D3	D4	D5	.....	D95		
1	tuju	1	1	1	1	0	-	0	55	0,237
2	teliti	4	3	3	7	10	-	1	75	1.103
3	pola	0	0	0	0	0	-	0	7	2.133
4	salur	0	0	0	0	0	-	0	3	2.501
5	asar	0	1	0	0	0	-	0	5	2.279
;	-	-	-	-	-	-	-	-	-	-
;	-	-	-	-	-	-	-	-	-	-
;	-	-	-	-	-	-	-	-	-	-
110	banding	0	1	0	0	0	-	0	22	1.635

[Sumber: Hasil Perhitungan]

Contoh perhitungan idf:

$$idf_i = \log\left(\frac{D}{df_i}\right)$$

$$idf_{t_{uju}} = \log\left(\frac{95}{55}\right)$$

$$idf_{t_{uju}} = 0,237$$

dimana :

D adalah jumlah semua dokumen dalam koleksi

$Df_i$  adalah jumlah dokumen yang mengandung term  $t_i$

Berdasarkan hasil *text preprocessing* dan pembobotan dengan tf-idf, maka tahap selanjutnya adalah proses cluster dengan rumus *euclidean distance* seperti dibawah ini:

Tabel 11. Iterasi Ke-1 Metode K-means

No	Doc	C1	C2	C3	Cluster
1	D1	20.187	17.066	18.324	C2
2	D2	18.682	17.476	17.893	C2
3	D3	19.939	17.761	18.034	C2
4	D4	20.143	17.694	18.302	C2
5	D5	20.098	17.522	18.075	C2
:	-	-	-	-	-
:	-	-	-	-	-
<b>95</b>	<b>D95</b>	20.43	17.337	18.419	C2

[Sumber:Hasil Perhitungan]

Tabel 12. Iterasi Ke-2 Metode K-means

No	Doc	C1	C2	C3	Cluster
1	D1	20.187	17.066	18.324	C2
2	D2	18.682	17.476	17.893	C2
3	D3	19.939	17.761	18.034	C2
4	D4	20.143	17.694	18.302	C2
5	D5	20.098	17.522	18.075	C2
:	-	-	-	-	-
:	-	-	-	-	-
<b>95</b>	<b>D95</b>	20.43	17.37	18.49	C2

[Sumber:Hasil Perhitungan]

## B. Perhitungan DBI Dari Hasil Cluster

Perhitungan nilai Davies-Bouldin Index bertujuan untuk menganalisa kualitas clusters dari setiap proses clustering tersebut. Untuk mendapatkan nilai Davies-Bouldin Index, terlebih dahulu dihitung nilai Sum of Square Within-cluster (SSW), Sum of Square Between-cluster (SSB) dan Ratio.

$$SSW_1 = \frac{1}{1} * (0) = 0$$

$$SSW_2 = \frac{1}{86} * (1.304.830) = 15.172,4$$



$$SSW_3 = \frac{1}{8} * (106.721) = 13.340,125$$

Setelah diperoleh nilai SSW, maka langkah selanjutnya adalah menghitung nilai Sum of Square Between-cluster (SSB). Adapun nilai yang diperoleh setelah menghitung SSB adalah sebagai berikut:

$$C_1 = 1,458$$

$$C_2 = 96,651$$

$$C_3 = 4,079$$

$$SSB_{1,2} = \sqrt{(1,458 - 96,651)^2} = 95,193$$

$$SSB_{1,3} = \sqrt{(1,458 - 4,079)^2} = 2,621$$

$$SSB_{2,3} = \sqrt{(96,651 - 4,079)^2} = 92,572$$

Langkah selanjutnya setelah mendapatkan nilai SSW dan nilai SSB adalah menghitung nilai Ratio (Rasio).

$$R_{1,2} = \frac{0+15.172,4}{95,193} = 159,386$$

$$R_{1,3} = \frac{0+13.340,125}{2,621} = 5.089,708$$

$$R_{2,3} = \frac{15.172,4+13.340,125}{92,572} = 308,004$$

Setelah mendapatkan nilai rasio, maka langkah terakhir adalah menghitung nilai Davies-Bouldin Index (DBI). Nilai Davies-Bouldin Index (DBI) seperti dibawah ini.

$$DBI = \frac{1}{3} \max(159,386, 5.089,708, 308,004)$$

$$DBI = 1.696$$

### C. Validitas Cluster dengan Metode Davies Bouldin Index

Setelah proses clustering menghasilkan beberapa cluster, proses selanjutnya dalam penelitian ini adalah menghitung nilai validitas cluster menggunakan Davies Bouldin Index (DBI). Pada penelitian ini melakukan beberapa kali percoba dengan menggunakan 2 Cluster, 3 Cluster, 4 Cluster, 5 Cluster 6 Cluster, 7 Cluster dan 8 Cluster dari beberapa masing-masing cluster diperoleh nilai DBI berbeda berikut hasil DBI yang diperoleh:

Tabel 13. Hasil Davies Bouldin Index (DBI)

No	Jumlah Cluster	Nilai DBI
1	2	0.85
2	3	0.71
3	4	0.87
4	5	0.63
5	6	0.74
6	7	0.7
7	8	0.86

[Sumber:Hasil Perhitungan]

Berdasarkan hasil Tabel 13 dapat disimpulkan bahwa nilai DBI tertinggi pada percobaan dengan menggunakan 4 Cluster dan jumlah data yang digunakan 95 data dengan memperoleh nilai DBI 0,87.

---

## KESIMPULAN

Berdasarkan hasil analisa dan pembahasan pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut: hasil validitas cluster dengan perbandingan hasil DBI menggunakan 2 cluster sampai dengan 8 cluster di peroleh hasil DBI tertinggi pada percobaan dengan menggunakan 4 Cluster dengan memperoleh nilai DBI 0,87.

## REFERENSI

- Arlene G. Taylor, Wynar's Introduction to Cataloging and Classification 9<sup>th</sup> Edition (USA: Greenwood Publishing Group Inc, 2004)
- Batjo, A. A., *Klasifikasi Islam: Adaptasi Klasifikasi Persepuluh Dewey Dan Perluasan 297* (Jakarta: UI Press, 1985)
- Everitt, B.S., Landau, S., Leese, M., dan Stahl, D.. 2011. *Cluster Analysis*, 5<sup>th</sup> Edition. The Atrium, Southern Gate, Cichester, West Sussex, PO19 8SQ:John Willey & Sons, Ltd.
- Fauzi, M., 2017, *Optimasi Pusat Cluster Awal K-Means Dengan Algoritma Genetika Pada Pengelompokan Dokumen*, Medan, Universitas Sumatera Utara.
- Feldman, D.C. 2004, *Managing Individual Are Group. Behavioral in Organization*", New York: McGraw Hill.
- Giyanto, H., 2008. *Penerapan algoritma Clustering K-Means, K-Medoid, Gath Geva*. Tesis Tidak Terpublikasi. Yogyakarta: Universitas Gajah Mada.
- Goller, 2000, *Automatic Document Classification: A Thorough Evaluation of Various Methods, Proceedings of International Symposium on information Theory and Its Application*, pp. 145-162, USA.
- Gupta, V., 2009. *A Survey of Text Mining Techniques and Application*. *Journal of Emerging Technologies in Web Intelligence*. Vol. 1: 60-75 [2]
- Hair, J.F., Black, W.C., Babin, B.J., dan Anderson, R.E.. 2010. *Multivariate Data Analysis*, 7<sup>th</sup> Edition. New Jersey: Pearson Prentice Hall.
- Han, J., and Kamber, M., 2000, "Data mining: Concept and Technique", Morgan Kaufman. Johnson, R.A. dan Wichern, D.W.. 2007. *Applied Multivariate Statistical Analysis*, 6<sup>th</sup> Edition. Upper Saddle River, New Jersey 07458: Prentice Education, Inc.
- Johnson, R. A. & Wichern, D. W. 1996. *Applied Multivariate Stastistical Analysis*. 3<sup>th</sup>. New Delhi: Prentice-Hall.
- Ningrat, S, R., Maruddani, D, A, I., Wuryandari, T., 2016, *Analisis Cluster Dengan Algoritma K-Means Dan Fuzzy C-Means Clustering Untuk Pengelompokan Data Obligasi Korporasi*, Universitas Diponegoro, Semarang.
- Pressman, Roger S. 2012. *Rekayasa Perangkat Lunak – Buku Satu, Pendekatan Praktisi (Edisi 7)*. Yogyakarta: Andi.
- Priianti KR & Wijaya H. 2014. *Aplikasi Text Mining untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering*, *Jurnal Cybermatika*, Vol. 2(1).
- Sebastiani, F., 2002. *Machine Learning in Automated Text Categorization*. *ACM Computing Surveys*. Vol. 34(1): 1-47
- Simamora, Bilson. 2005. *Analisis Multivariat Pemasaran Edisi Pertama*. Jakarta: PT. Gramedia Pustaka tama
- Supriyanto, W., Muhsin, A., 2008, *Informasi Perpustakaan*", Yogyakarta, Kansius Anggota IKAPI

- Tawa, P., Hamakonda, Tairas, J, N, B., 2008, “Pengantar Cluster Persepuluhan Dewey”, Cetakan ke – 18. Jakarta.
- Triawati, C. 2009, Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia, Institut Teknologi Telkom, Bandung.
- Wiber, M., Welling, M., and Perona, P., 2000, “Unsupervised Learning of Models for Recognition”. In Proc. 6 th Euro. Conf,Comput, Vision, Dublin, Ireland.