

LAPORAN PRAKTIKUM STATISTIKA

(Dosen Pengampu : *Fachrul Madrapriya, S.T., M.PSDA.*)



Disusun oleh :

Nama : MUHAMMAD RIZAL NURFIRDAUS

NIM : 20230810088

Kelas : TINFC-2023-04

**TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS KUNINGAN**

PRE TEST

1. Bagaimana menurut pendapat perbedaan antara kedua jenis metode dalam analisis cluster, jelaskan dengan contoh-contoh?

K-Means Clustering dan Hierarchical Clustering adalah dua metode analisis cluster yang populer dengan pendekatan berbeda. K-Means Clustering membagi data menjadi k cluster berdasarkan kedekatan centroid, memerlukan penentuan jumlah cluster sebelumnya, dan lebih efisien untuk dataset besar. Contohnya, K-Means digunakan dalam segmentasi pelanggan dan optimasi penggunaan server. Meskipun mudah diimplementasikan, K-Means sensitif terhadap inisialisasi centroid dan hanya cocok untuk cluster berbentuk bulat dan berukuran sama.

Sebaliknya, Hierarchical Clustering membangun hierarki atau dendrogram yang menunjukkan hubungan antar data, tanpa memerlukan penentuan jumlah cluster sebelumnya. Metode ini cocok untuk data dengan bentuk dan ukuran cluster yang bervariasi dan memberikan visualisasi yang kaya melalui dendrogram. Contohnya termasuk analisis genetik dan organisasi dokumen. Namun, Hierarchical Clustering kurang efisien untuk dataset besar dan sensitif terhadap noise dan outlier. Pemilihan metode bergantung pada karakteristik data dan tujuan analisis yang diinginkan.

POST TEST

1. Buatlah studi kasus dari salah satu analisis cluster dalam bidang Teknik informatika?

Studi Kasus: Analisis Cluster untuk Optimalisasi Penggunaan Sumber Daya Server pada Perusahaan E-Commerce

Latar Belakang

Perusahaan e-commerce sering kali menghadapi tantangan dalam mengelola penggunaan sumber daya server mereka. Lonjakan lalu lintas web, terutama selama periode penjualan atau promosi khusus, dapat menyebabkan beban server yang tidak merata. Untuk mengoptimalkan penggunaan sumber daya dan memastikan performa yang stabil, perusahaan perlu menganalisis pola penggunaan server dan mengelompokkan aplikasi atau layanan yang memerlukan sumber daya serupa.

Tujuan

Studi ini bertujuan untuk menggunakan analisis cluster dalam mengelompokkan aplikasi atau layanan pada server berdasarkan pola penggunaan sumber daya mereka. Dengan mengidentifikasi cluster yang homogen, perusahaan dapat mengalokasikan sumber daya secara lebih efisien dan mengurangi risiko overload pada server tertentu.

Data dan Metodologi

1. Pengumpulan Data:

- Data penggunaan CPU, RAM, dan disk I/O dikumpulkan dari server perusahaan selama tiga bulan.

- Data ini mencakup metrik seperti rata-rata penggunaan, puncak penggunaan, dan variasi harian.
2. **Praproses Data:**
 - Data dibersihkan untuk menghilangkan nilai yang hilang atau tidak konsisten.
 - Standarisasi data dilakukan untuk memastikan semua metrik memiliki skala yang sama.
 3. **Pemilihan Algoritma Klustering:**
 - Algoritma K-Means dipilih karena kesederhanaannya dan kemampuannya untuk menangani data dalam jumlah besar dengan baik.
 - Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow dan Silhouette Score.
 4. **Implementasi Klustering:**
 - Data dimasukkan ke dalam model K-Means.
 - Proses klustering dilakukan dan hasilnya dievaluasi.
 5. **Evaluasi dan Interpretasi:**
 - Cluster yang terbentuk dianalisis untuk memahami karakteristik masing-masing kelompok.
 - Rekomendasi diberikan berdasarkan hasil analisis cluster.

TUGAS MANDIRI

1. Dari studi kasus yang telah anda buat, Buatlah analisis clusternya?

Non Hierarchical Cluster(K-Means cluster)

Initial Cluster Centers

	Cluster	
	1	2
Zscore (Rata_rata_Penggunaan _CPU)	1.76749	-1.25585
Zscore (Puncak_Penggunaan_C PU)	1.71756	-1.24375
Zscore (Rata_rata_Penggunaan _RAM)	-.58021	-1.40908
Zscore (Puncak_Penggunaan_R AM)	-.57827	-1.30111
Zscore(Rata_rata_Disk)	-.45990	1.94258
Zscore(Puncak_Disk)	-.47809	1.97365
Zscore (Rata_rata_Penggunaan _CPU)	1.76749	-1.25585
Zscore (Puncak_Penggunaan_C PU)	1.71756	-1.24375
Zscore (Rata_rata_Penggunaan _RAM)	-.58021	-1.40908
Zscore (Puncak_Penggunaan_R AM)	-.57827	-1.30111
Zscore(Rata_rata_Disk)	-.45990	1.94258
Zscore(Puncak_Disk)	-.47809	1.97365

Final Cluster Centers

	Cluster	
	1	2
Zscore (Rata_rata_Penggunaan _CPU)	.40533	-.94576
Zscore (Puncak_Penggunaan_C PU)	.40287	-.94003
Zscore (Rata_rata_Penggunaan _RAM)	.30787	-.71835
Zscore (Puncak_Penggunaan_R AM)	.29946	-.69874
Zscore(Rata_rata_Disk)	-.53835	1.25616
Zscore(Puncak_Disk)	-.53938	1.25856
Zscore (Rata_rata_Penggunaan _CPU)	.40533	-.94576
Zscore (Puncak_Penggunaan_C PU)	.40287	-.94003
Zscore (Rata_rata_Penggunaan _RAM)	.30787	-.71835
Zscore (Puncak_Penggunaan_R AM)	.29946	-.69874
Zscore(Rata_rata_Disk)	-.53835	1.25616
Zscore(Puncak_Disk)	-.53938	1.25856

ANOVA						
	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Zscore (Rata_rata_Penggunaan_CPU)	3.833	1	.646	8	5.936	.041
Zscore (Puncak_Penggunaan_CPU)	3.787	1	.652	8	5.812	.042
Zscore (Rata_rata_Penggunaan_RAM)	2.212	1	.849	8	2.606	.145
Zscore (Puncak_Penggunaan_RAM)	2.092	1	.863	8	2.423	.158
Zscore(Rata_rata_Disk)	6.763	1	.280	8	24.180	.001
Zscore(Puncak_Disk)	6.788	1	.276	8	24.556	.001
Zscore (Rata_rata_Penggunaan_CPU)	3.833	1	.646	8	5.936	.041
Zscore (Puncak_Penggunaan_CPU)	3.787	1	.652	8	5.812	.042
Zscore (Rata_rata_Penggunaan_RAM)	2.212	1	.849	8	2.606	.145
Zscore (Puncak_Penggunaan_RAM)	2.092	1	.863	8	2.423	.158
Zscore(Rata_rata_Disk)	6.763	1	.280	8	24.180	.001
Zscore(Puncak_Disk)	6.788	1	.276	8	24.556	.001

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Number of Cases in each Cluster

Cluster	1	7.000
	2	3.000
Valid		10.000
Missing		.000

Hierarchical Cluster

→ Cluster

Proximity Matrix

Case	Squared Euclidean Distance									
	1:A	2:B	3:C	4:D	5:E	6:F	7:G	8:H	9:I	10:J
1:A	.000	2.033	22.962	13.972	30.903	21.889	7.699	12.503	6.594	7.297
2:B	2.033	.000	13.567	12.521	34.067	20.547	2.987	12.190	2.220	4.673
3:C	22.962	13.567	.000	21.665	30.803	13.329	4.195	13.849	6.482	11.824
4:D	13.972	12.521	21.665	.000	19.660	13.098	13.280	5.701	7.959	2.375
5:E	30.903	34.067	30.803	19.660	.000	3.674	26.511	6.396	26.183	20.517
6:F	21.889	20.547	13.329	13.098	3.674	.000	12.147	2.043	12.566	10.538
7:G	7.699	2.987	4.195	13.280	26.511	12.147	.000	8.334	.894	4.522
8:H	12.503	12.190	13.849	5.701	6.396	2.043	8.334	.000	7.029	4.114
9:I	6.594	2.220	6.482	7.959	26.183	12.566	.894	7.029	.000	1.686
10:J	7.297	4.673	11.824	2.375	20.517	10.538	4.522	4.114	1.686	.000

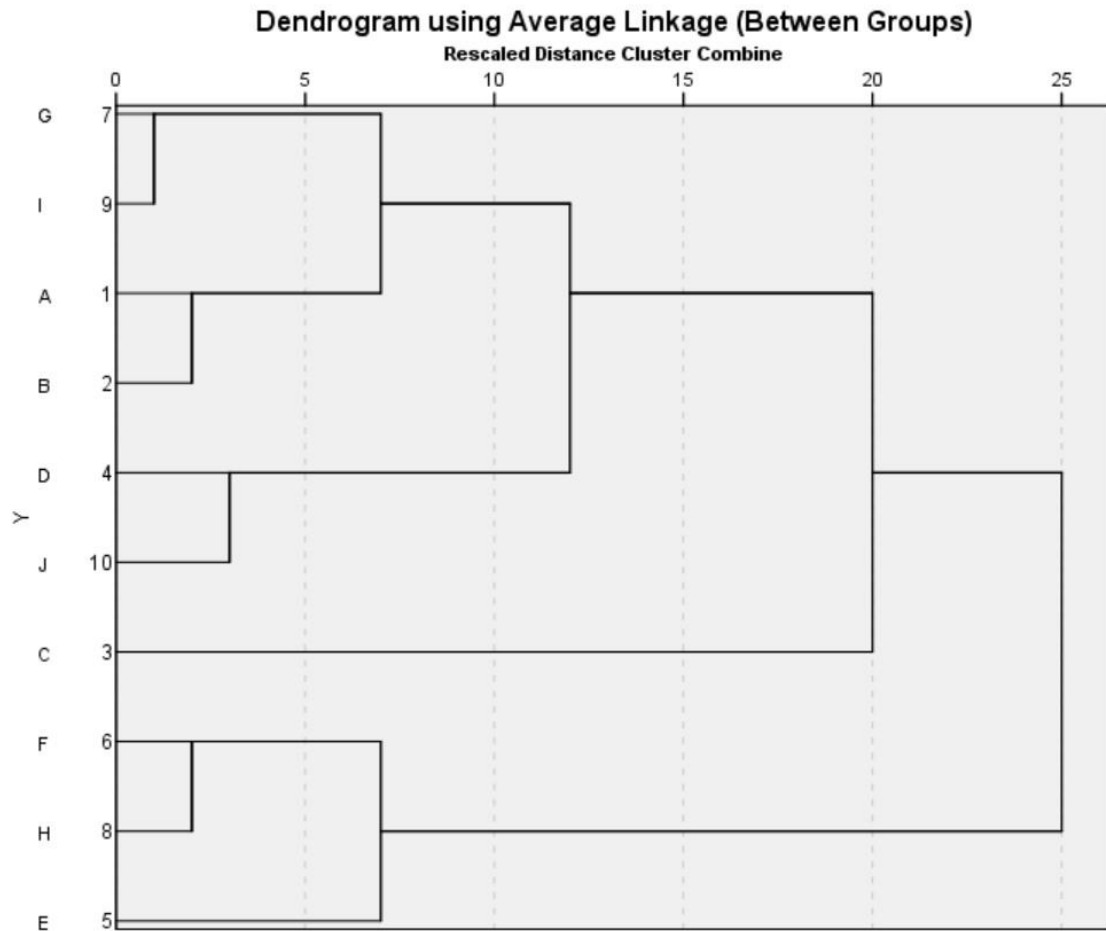
This is a dissimilarity matrix

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	7	9	.894	0	0	5
2	1	2	2.033	0	0	5
3	6	8	2.043	0	0	6
4	4	10	2.375	0	0	7
5	1	7	4.875	2	1	7
6	5	6	5.035	0	3	9
7	1	4	8.239	5	4	8
8	1	3	13.449	7	0	9
9	1	5	16.975	8	6	0

Cluster Membership

Case	5 Clusters	4 Clusters	3 Clusters	2 Clusters
1:A	1	1	1	1
2:B	1	1	1	1
3:C	2	2	2	1
4:D	3	3	1	1
5:E	4	4	3	2
6:F	5	4	3	2
7:G	1	1	1	1
8:H	5	4	3	2
9:I	1	1	1	1
10:J	3	3	1	1



2. Tarik sebuah Kesimpulan dengan sebuah pengelompokan dengan analisis cluster sesuai dengan objek yang diteliti.

Dari hasil uji ini, dapat disimpulkan bahwa terdapat perbedaan signifikan antara cluster dalam hal rata-rata dan puncak penggunaan CPU serta penggunaan Disk (baik rata-rata maupun puncak). Namun, perbedaan dalam penggunaan RAM (baik rata-rata maupun puncak) antar cluster tidak signifikan. Hal ini menunjukkan bahwa penggunaan CPU dan Disk adalah faktor utama yang membedakan cluster dalam analisis ini, sementara penggunaan RAM tidak memiliki perbedaan signifikan antar cluster.

Uji ini membantu mengidentifikasi metrik mana yang paling berpengaruh dalam membedakan cluster, yang selanjutnya dapat digunakan untuk mengoptimalkan alokasi sumber daya pada server sesuai dengan karakteristik penggunaan dari masing-masing cluster.