# **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 Klastering:**
   * 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 Klastering:**
   * Data dimasukkan ke dalam model K-Means.
   * Proses klastering dilakukan dan hasilnya dievaluasi.
5. **Evaluasi dan Interpretasi:**
   * Cluster yang terbentuk dianalisis untuk memahami karakteristik masing-masing kelompok.
   * Rekomendasi diberikan berdasarkan hasil analisis cluster.

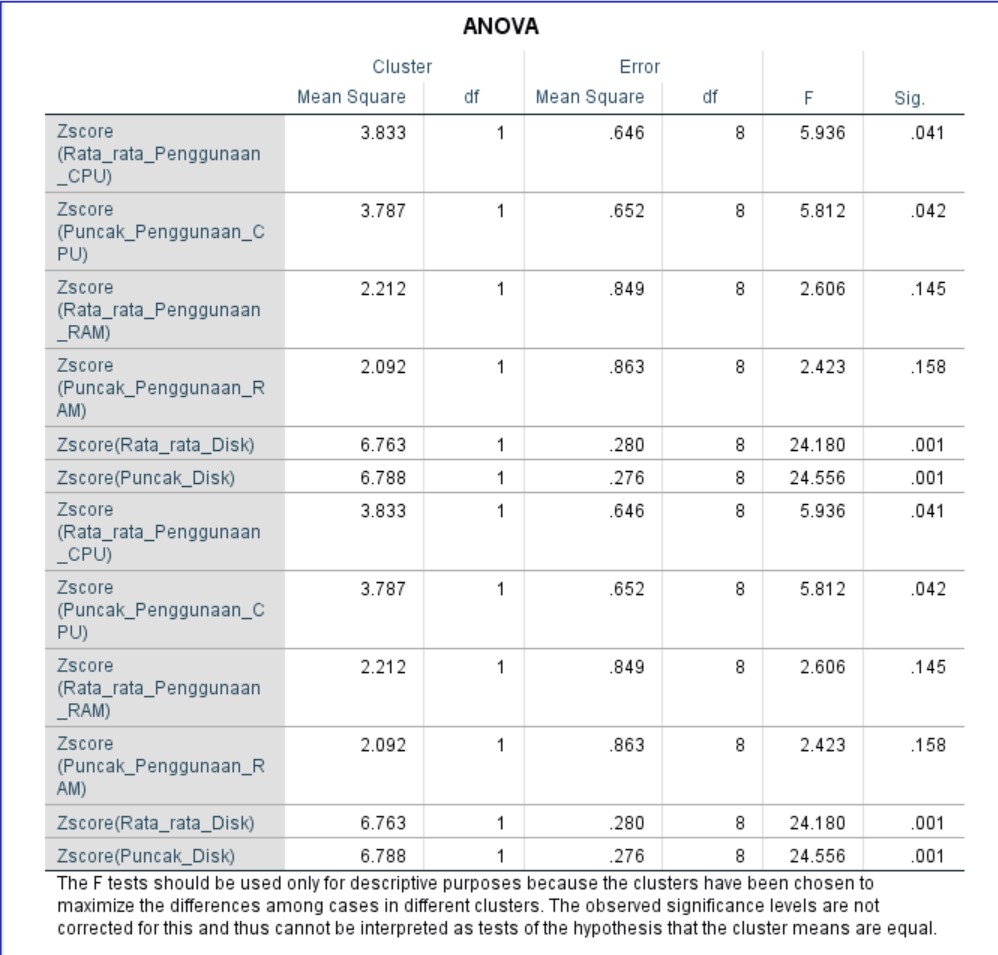
**TUGAS MANDIRI**

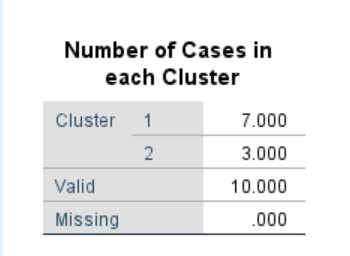
1. Dari studi kasus yang telah anda buat, Buatlah analisis clusternya?

**Non Hierarchical Cluster(K-Maean cluster)**

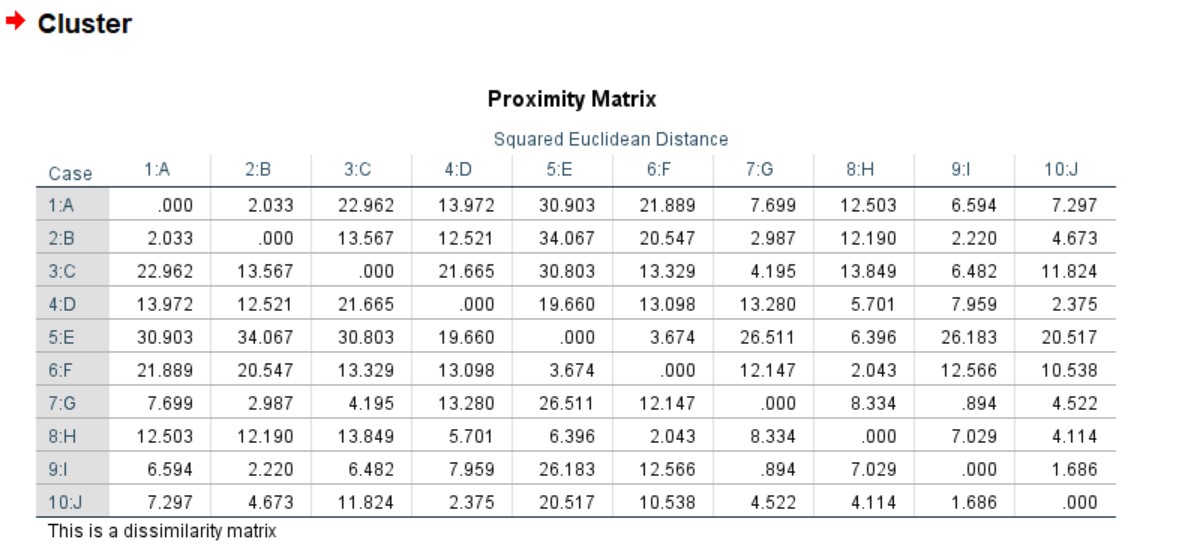


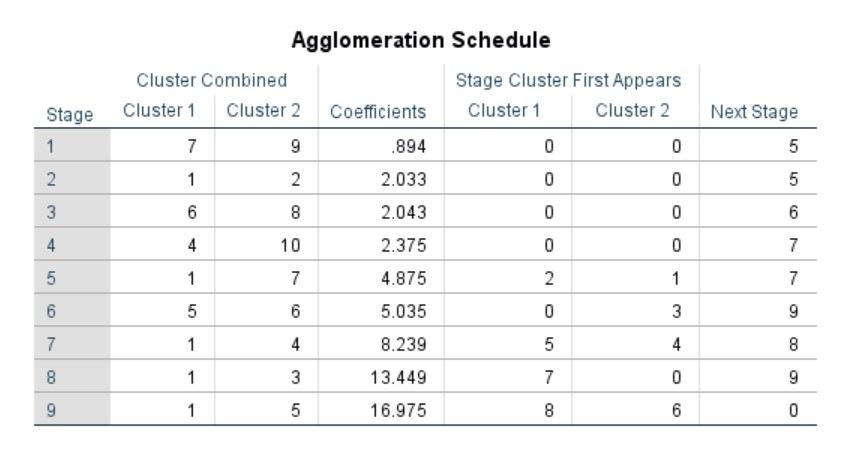


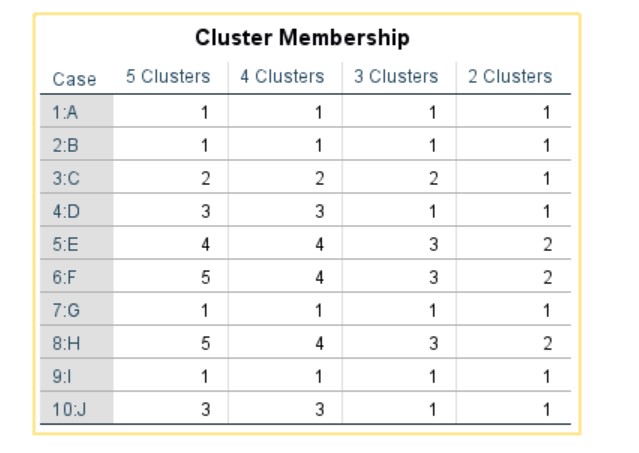


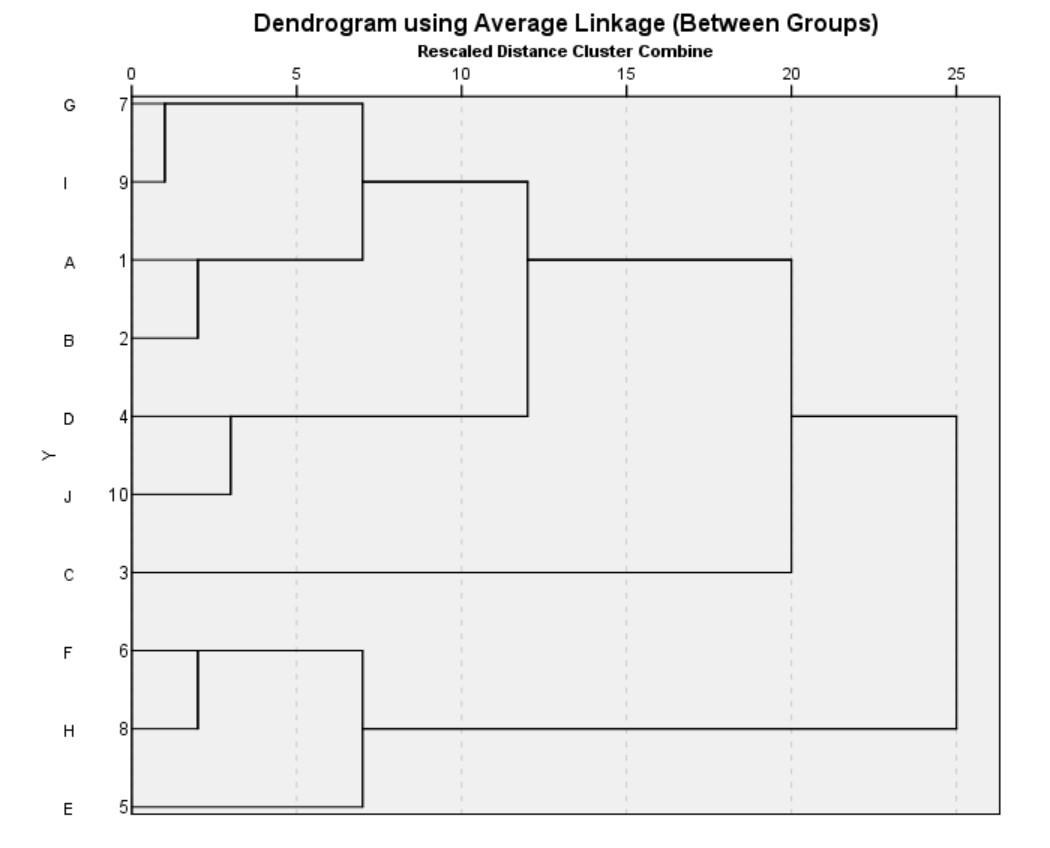


**Hierarchical Cluster**









1. 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.