**TUGAS BESAR DATA MINING**

* Judul dan link github project

Judul : Implementasi Clustering Menggunakan Kmeans pada Dataset Hardware Komputer

Link github : <https://github.com/Sapinah004/TB_DataMining>

* Data mahasiswa :

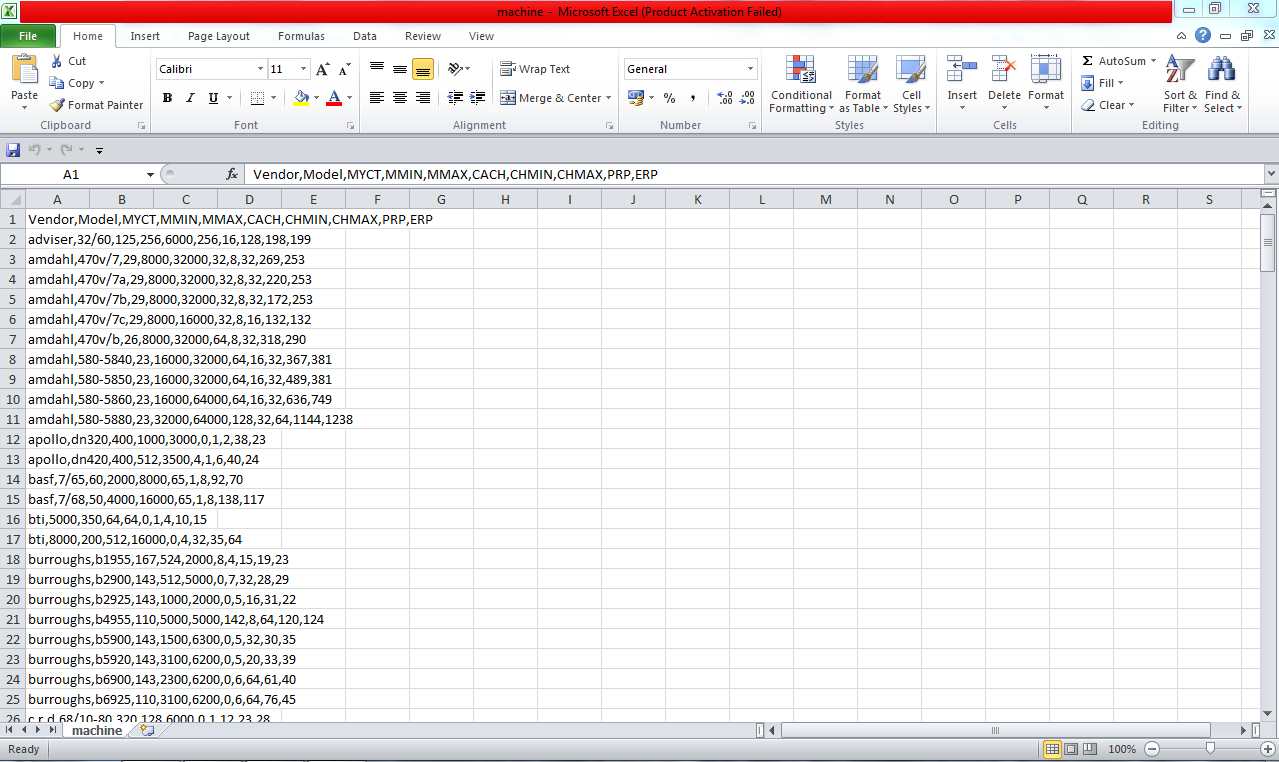
Sapinah-3311901004- <https://github.com/Sapinah004>

Arya Bima Hera Pradana – 3311901022 – <https://github.com/Arya022>

Dinda Alvianita – 3311911081 – [https://github.com/Dindaalvianitaa](https://github.com/dinda081)

* Dataset

Machine.csv



* Penjelasan dan analisa mengenai dataset

Dataset yang dipilih mengenai dataset kinerja CPU komputer

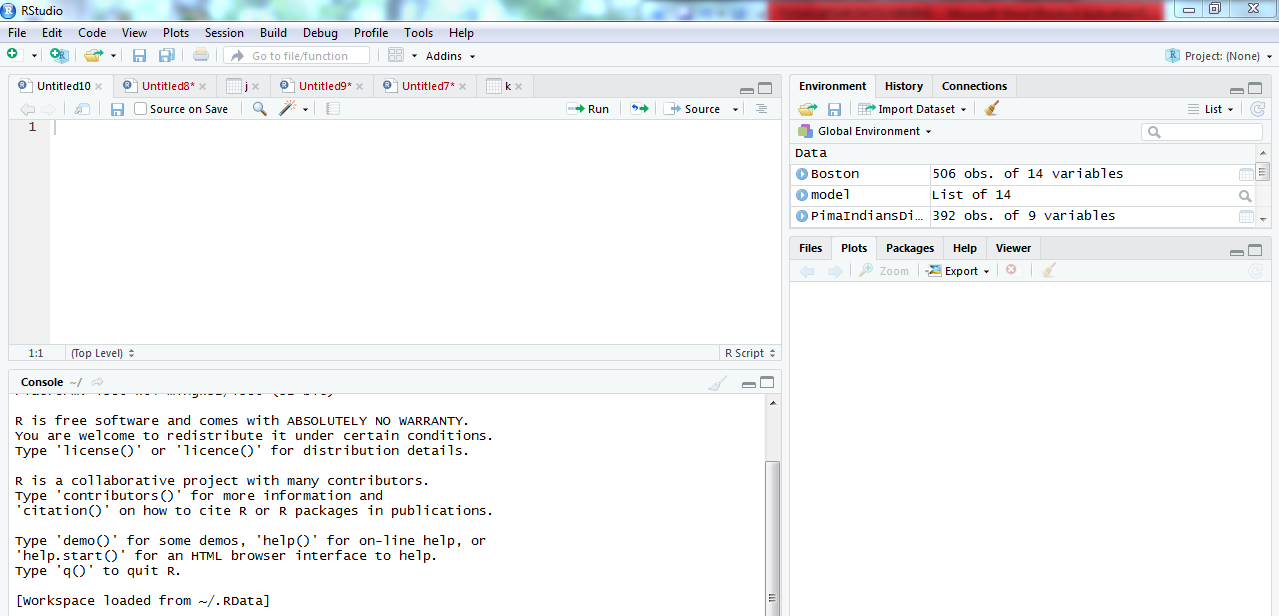
Adapun isi dataset terdiri dari :

1. Nama vendor
2. Jenis
3. MYCT – waktu siklus mesin dalam nano detik
4. MMIN – minimum memori utama
5. MMAX – maksimum memori utama
6. CACH – cache memori
7. CHMIN – minimum channel dalam satuan
8. CHMAX – maksimum channel dalam satuan
9. PRP – kinerja realtif yang dipublikasikan
10. ERP – perkiraan kinerja relatif yang dihasilkan

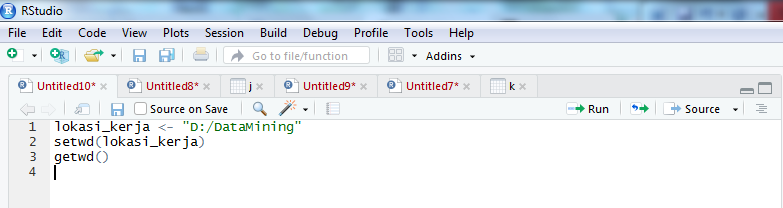
* Proses data mining yang dilakukan pada dataset tersebut

Proses yang dilakukan pada data mining yaitu proses clustering menggunakan kmeans

1. Buka aplikasi r studio



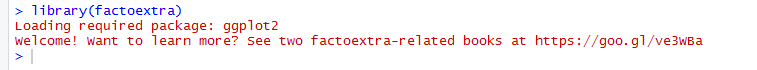
1. Atur lokasi kerja





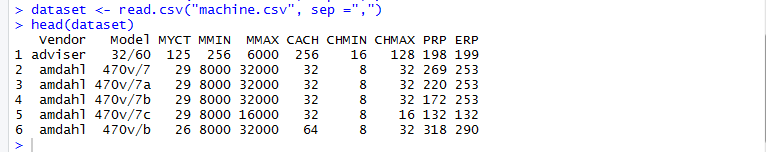
1. Gunakan library factoextra





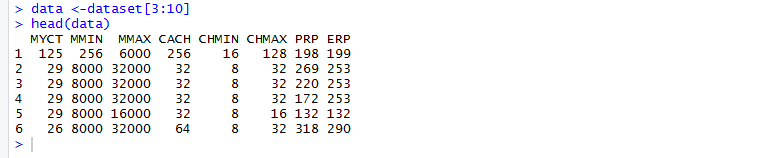
1. Import dan membaca dataset





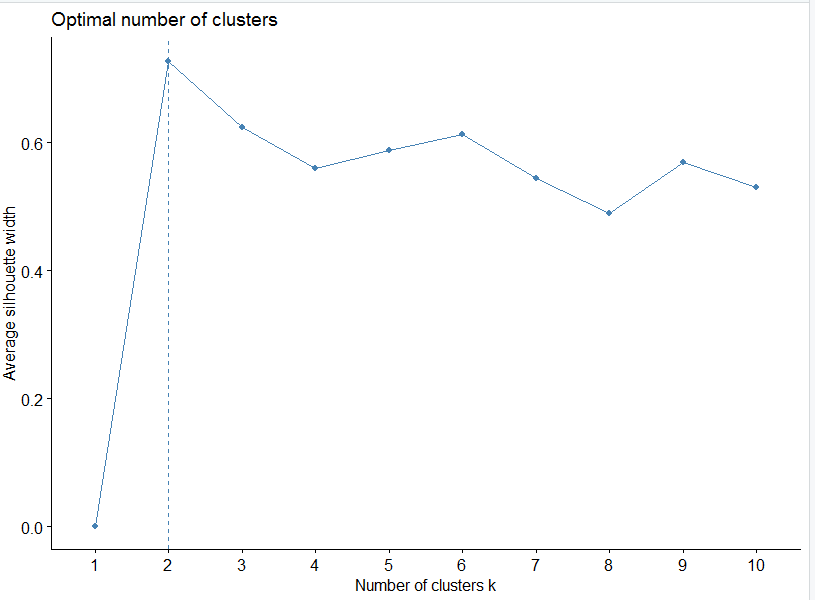
1. Karena pada cluster hanya menggunakan data numerik maka data akan diubah dan hanya berisi data numerik





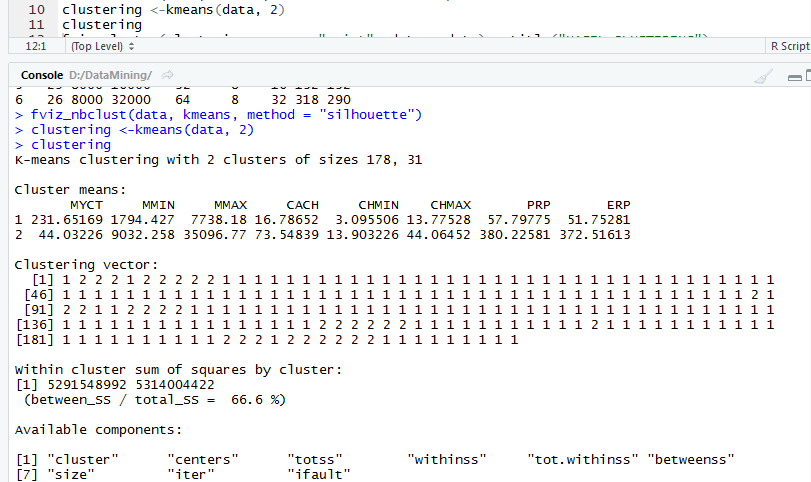
1. Selanjutnya, menentukan jumlah cluster optimal. Salah satu metode yang bisa digunakan untuk menentukan jumlah cluster yaitu menggunakan metode Silhoutte dengan menggunakan library factoextra.





Dari gambar diatas, rekomendasi cluster optimal adalah k=2 yang ditandai dengan garis vertikal yang berada pada angka 2

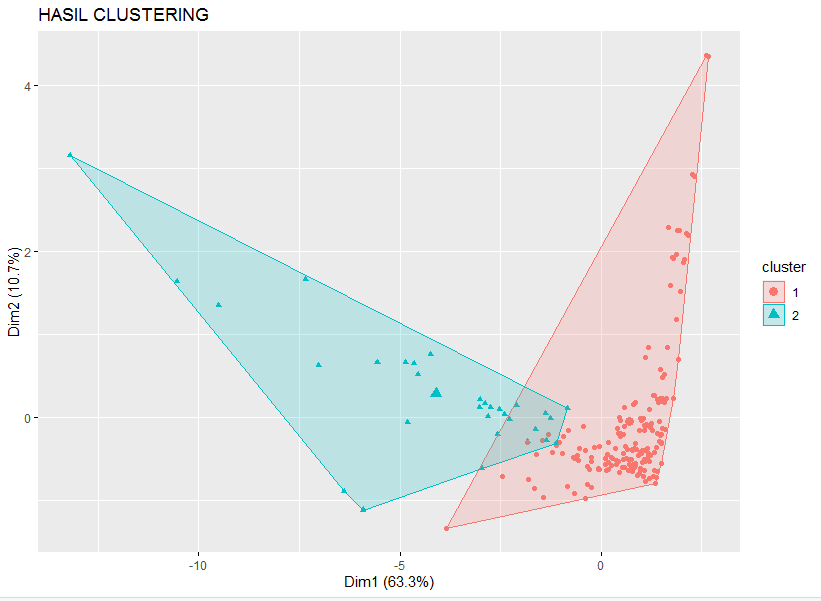
1. Clustering dengan metode kmenas dengan 2 buah titik cluster



Berdasarkan hasil diatas, pada cluster 1 terdapat 178 hardware komputer dan cluster 2 terdapat 31 hardware komputer

1. Untuk melihat visualisasi cluster gunakan syntax seperti berikut :

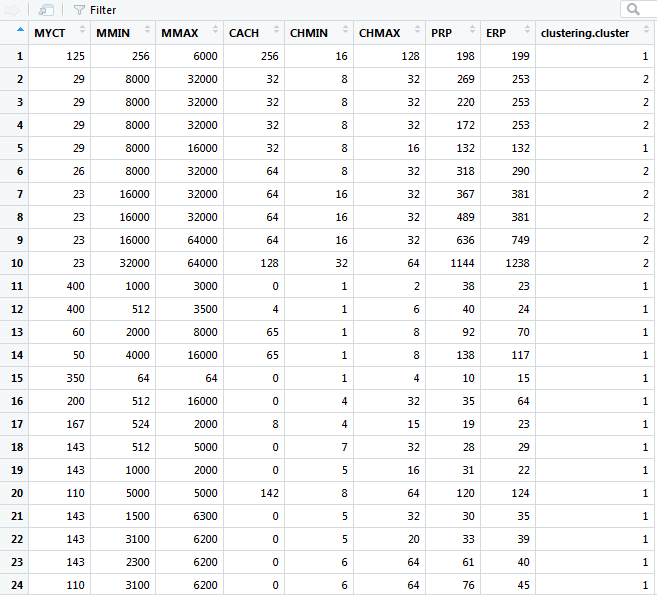


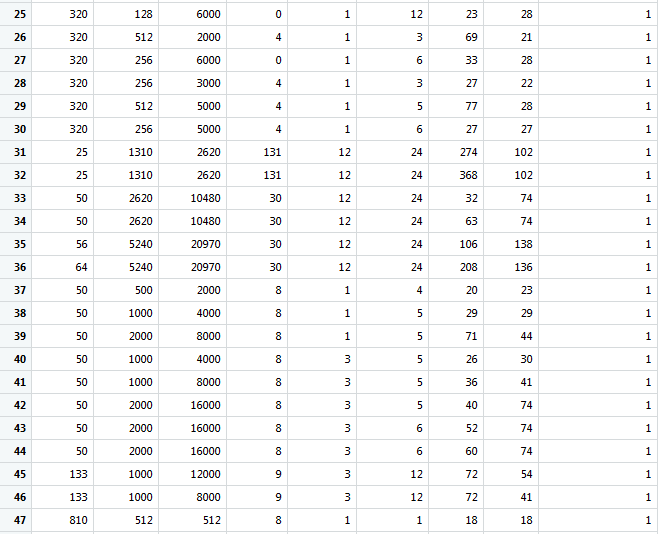


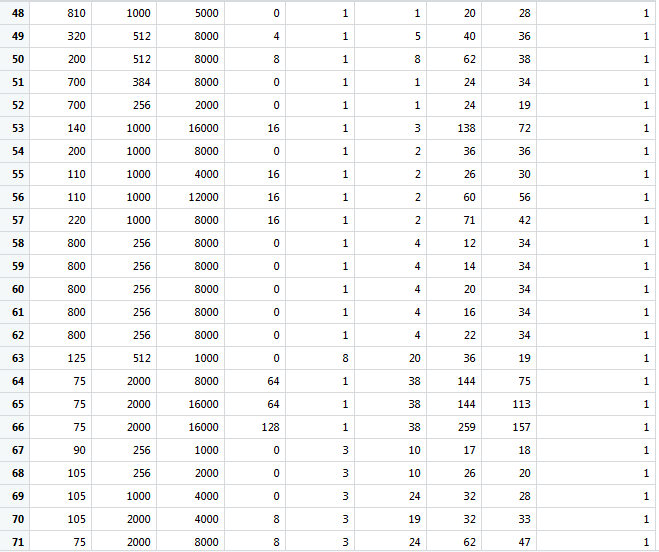
1. Buat sebuah data frame untuk melihat lebih jelas data clustering

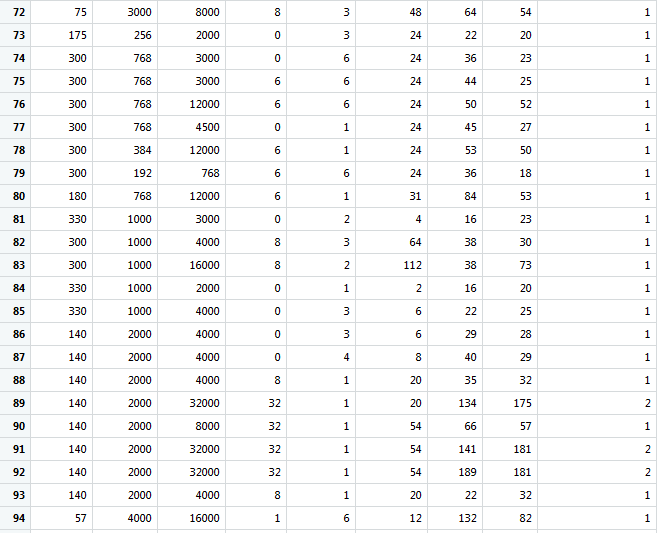


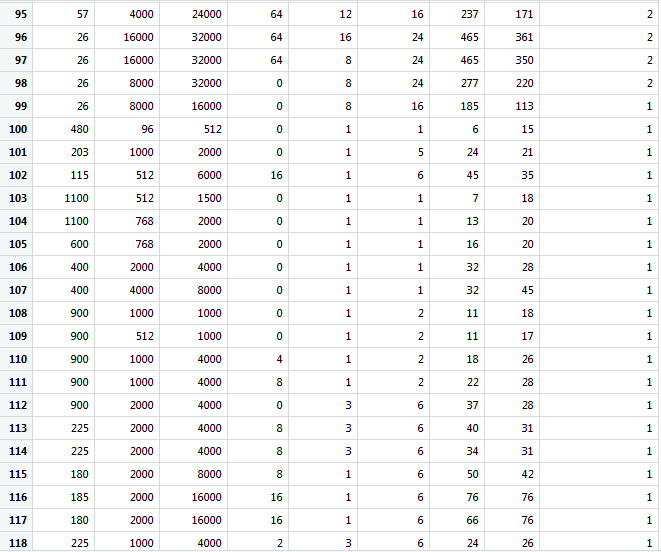
* Hasil data mining

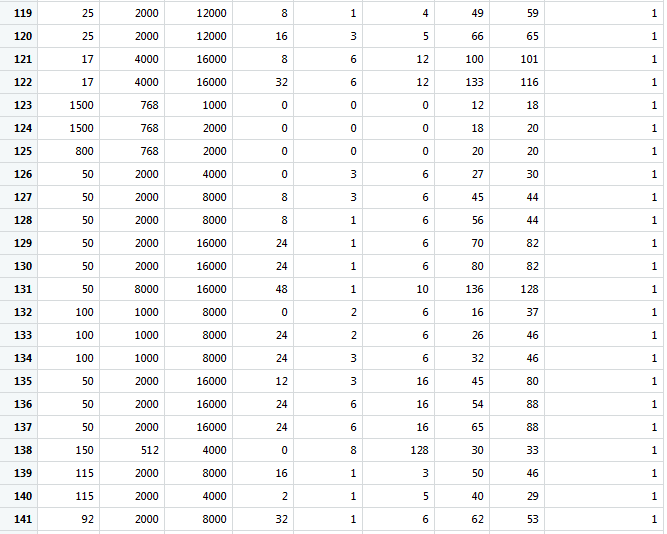


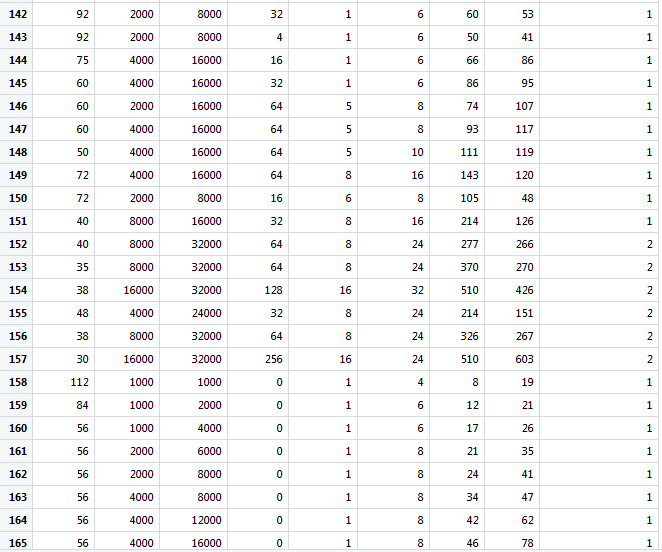


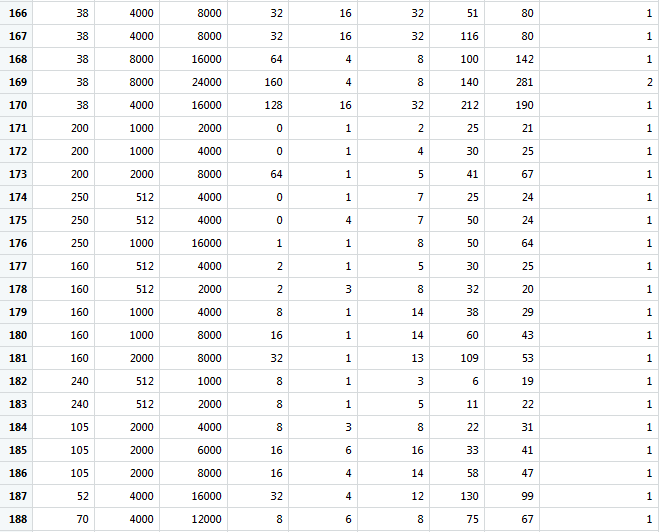


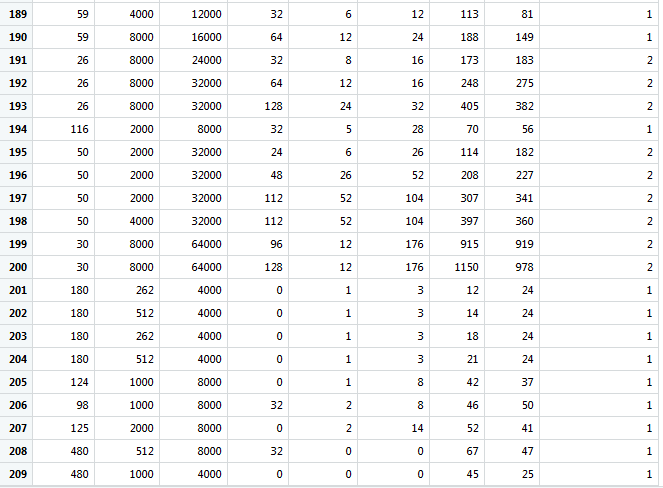












* Hasil analisa data mining

Hasil analisa data mining menunjukkan bahwa ada 2 cluster yaitu cluster pertama yang terdiri dari 178 CPU dan cluster kedua terdiri dari 38 CPU. Dari hasil analisa, pembagian cluster didapatkan berdasarkan hasil rata rata dari setiap atribut, jika rata rata masing masing CPU kurang dari 3500 maka akan masuk kedalam cluster pertama, namun jika rata rata masing masing CPU diatas 3500 maka akan masuk cluster kedua

Sehingga dapat disimpulkan bahwa :

Cluster pertama merupakan CPU komputer dengan kinerja yang cukup rendah ,Sedangkan cluster kedua merupakan CPU komputer dengan kinerja yang cukup tinggi

* Referensi

<https://medium.com/@17611063/analisis-cluster-menggunakan-k-means-clustering-rstudio-5834aee9b066>

<https://learning-if.polibatam.ac.id/course/view.php?id=43>

<https://archive.ics.uci.edu/ml/datasets/Computer+Hardware>

* Daftar pustaka

<https://medium.com/@yogasatrias98/implementasi-k-means-clustering-53e48782c9b8>

<https://medium.com/@taniasalsabila.ts/k-means-clustering-using-rstudio-d6132830f6b6>

<https://rpubs.com/nadhifanhf/k-means-clustering>

<https://learning-if.polibatam.ac.id/course/view.php?id=43>

<https://medium.com/@17611063/analisis-cluster-menggunakan-k-means-clustering-rstudio-5834aee9b066>