**UAS DATA MAINING**

Nama : Rahmat Tegar Patriot Hari Lambang

NIM : 221101021

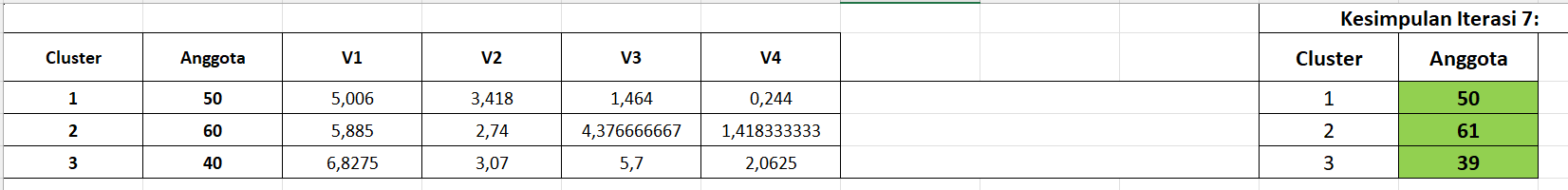
Kelas : TI3A

**A. Pengimplementasian Algoritma K-Means (Euclidean & Manhattan Distance)**

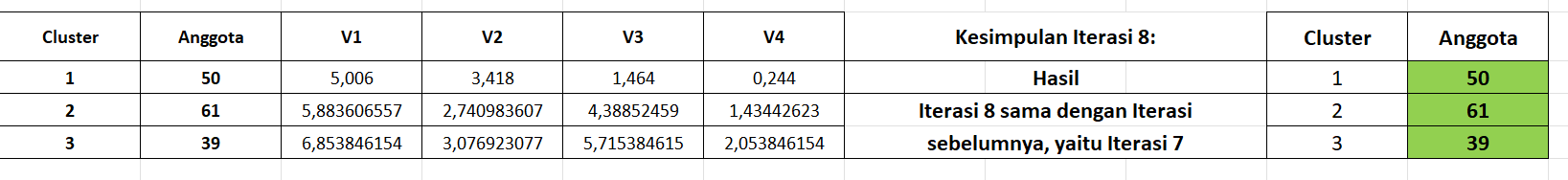
**1.**Hasil dari Clustering Euclidien Distance:

- Perhitungan **Iterasi 7 & 8** mempunyai hasil sama yaitu: **Iris-setosa(1)=50, Iris-virginica(2)=61, Iris-versicolor(3)=39**. Yang berarti hasil telah di temukan.

Iterasi 7 :



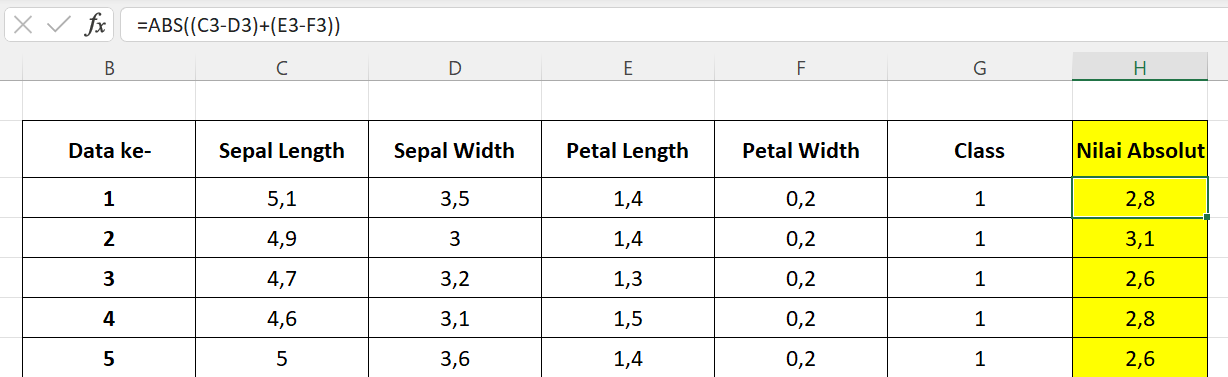
Iterasi 8 :

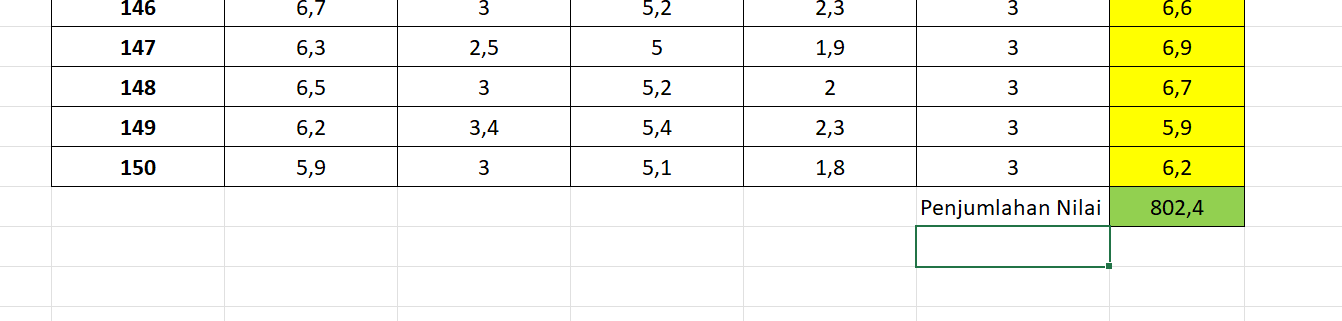


**2.**Hasil dari Clustering Manhattan Distance:

Perhitungan nilai data sebelumnya di jumlahkan dan dijadikan nilai absolut, dengan kata lain jikalau hasil adalah nilai minus(-) akan diubah menjadi nilai plus(+). Setelah itu di jumlah nilai keseluruhan.

ABS(Absolute):

****

****

**B. Perbandingan hasil clustering dari poin A.**

Euclidean Distance.

Pada analisis data yang telah diberikan sebelumnya, dengan metode Euncliden distance lebih mempertimbangkan hasil dengan perbedaan kuadrat yang mana hasil dapat dipengaruhi oleh perbedaan skala antar variable, dapat memberikan hasil yang lebih baik jika data sesuai dengan asumsi distribusi normal.

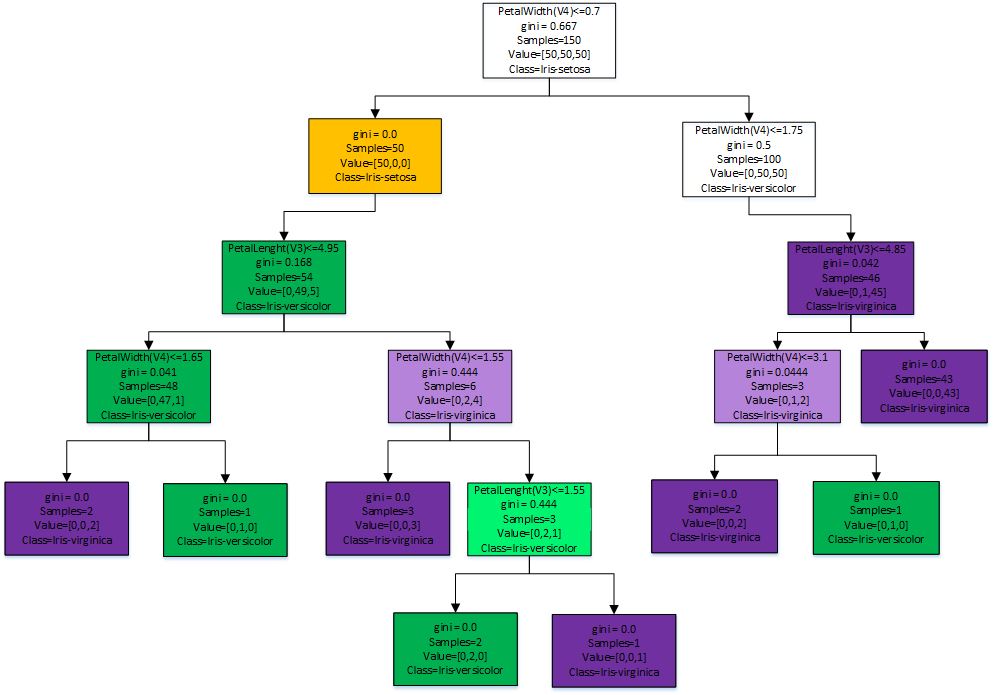
Manhattan Distance.

Pada analisis data yang telah diberikan sebelumnya, dengan metode Manhattan distance mempertimbangkan hasil menggunakan perbedaan absolut, yang tidak dapat dipengaruhi oleh perbedaan skala antar variabel, Manhattan distance dapat memberikan hasil yang lebih baik jika data memiliki distribusi yang lebih kompleks.

**C. Perbandingan nilai clustering menggunakan data asli**

Pada perbandingan data awal dengan hasil akhir cluster yang telah di dapat, terdapat perbedaan hasil yang mana pada 150 data asli = Iris-setosa:50, Irisversicolor:50, Iris-virginica:50. Sedangkan hasil data akhir Iris-setosa:50, Irisversicolor:61, Iris-virginica:39.

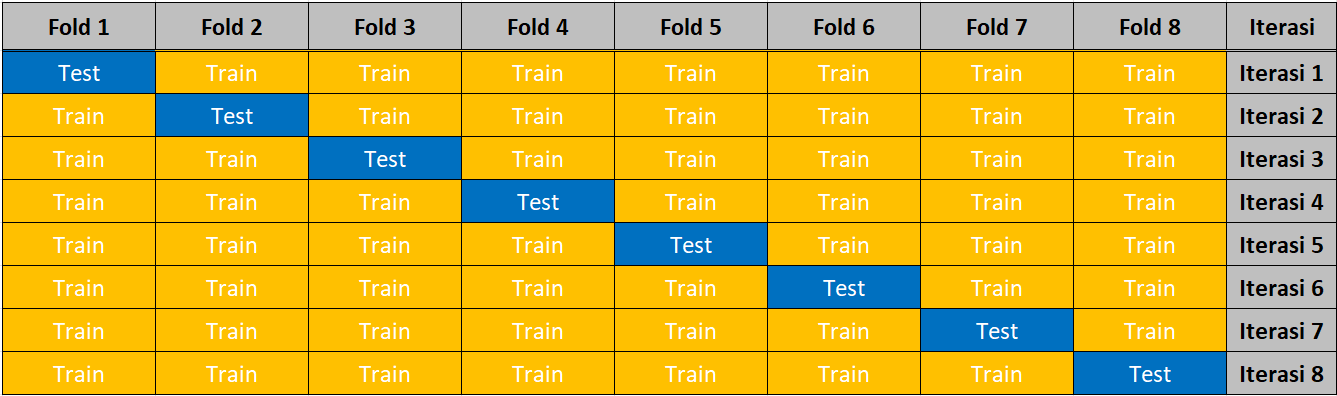
**D. Decision Tree**

****

**E.**

**1.)** Data yang dimiliki ada 150. Ibarat pakai K=5, berarti dibagi 150 data menjadi 5 lipatan, isinya masing-masing 30 data. Selanjutnya, perlu menentukan mana yang training data dan mana yang test data. Karena perbandingannya 80:20, berarti 120 data adalah training data dan 30 sisanya adalah test data. Berdasarkan ke-5 lipatan tadi, berarti akan ada 4 lipatan x 30 data = 120 training data. Dan sisanya ada 1 partisi test data berisi 30 data.

**2.)** K-Fold Cross Validation

****