

BAB IV

IMPLEMENTASI DAN PEMBAHASAN

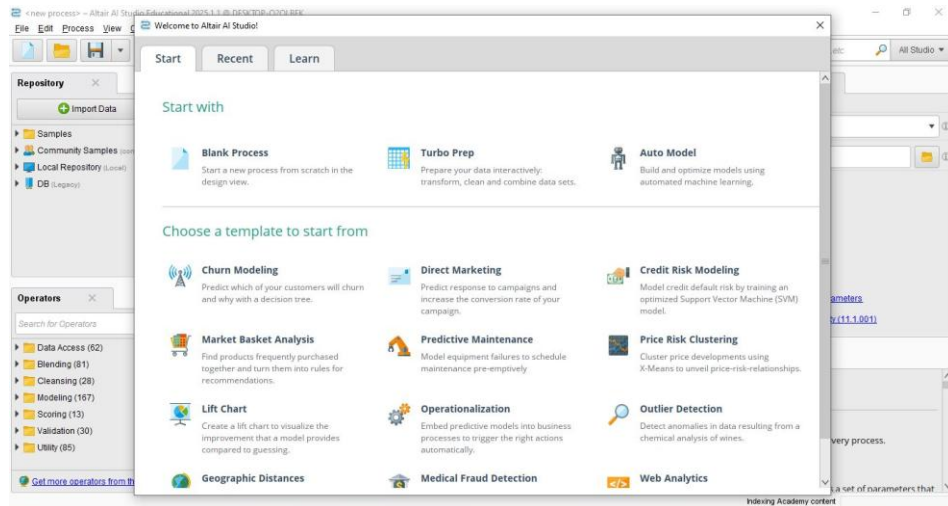
4.1. Hasil Implementasi RadpidMiner

Implementasi Fase implementasi dan pengujian termasuk dalam fase implementasi sistem. Aplikasi akan digunakan di lingkungan nyata selama fase ini. Langkah terakhir dalam penggunaan metode pengelompokan untuk mengelompokkan data sesuai dengan proses sebelumnya adalah implementasi. Agar proses pengelompokan data dapat diselesaikan dengan langkah-langkah pengelompokan dan pengujian data yang tepat, pendekatan pengelompokan diterapkan dalam aplikasi *RapidMiner*.

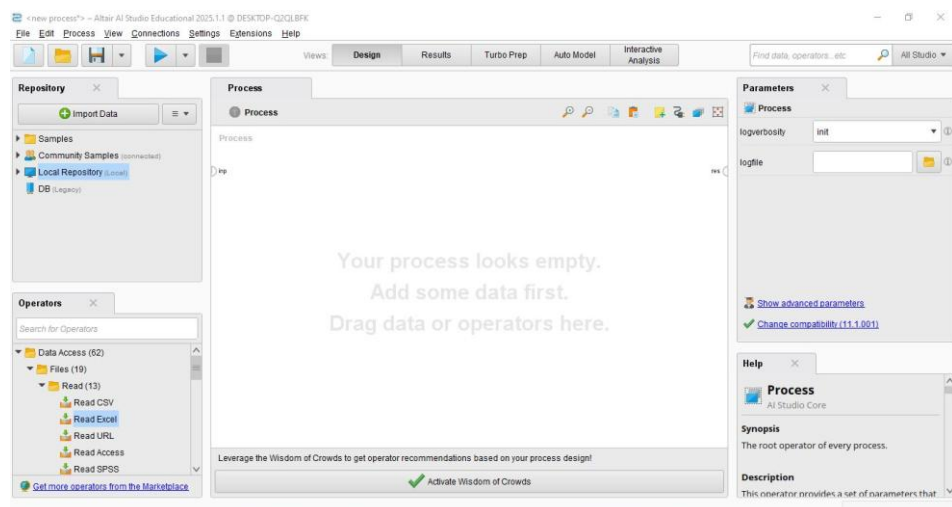


GAMBAR 4.1 Tampilan Awal *RapidMiner*

Setelah itu muncul tampilan awal yaitu membuat lembar kerja baru atau membuka file yang sudah ada. Klik blank proses untuk memulai lembar kerja baru.

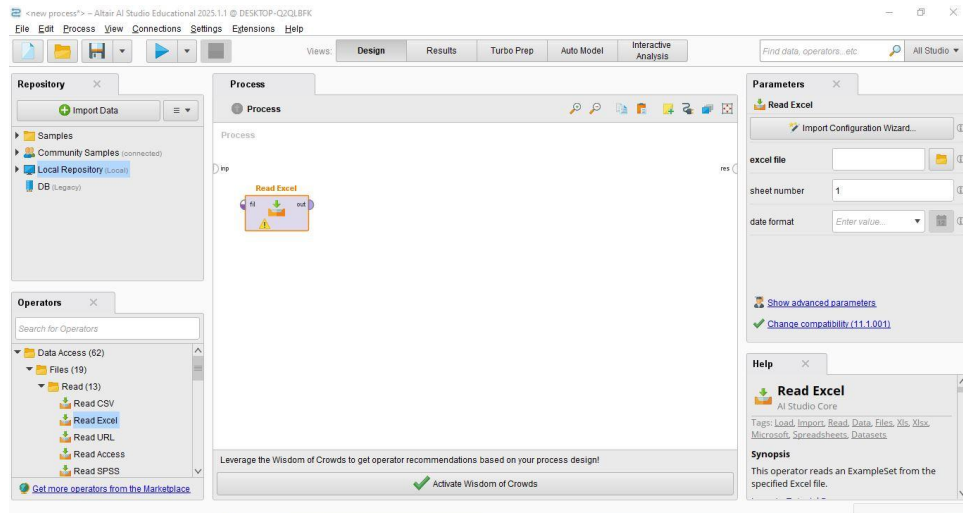


Gambar 4.2 Memilih Lembaran Kerja Baru



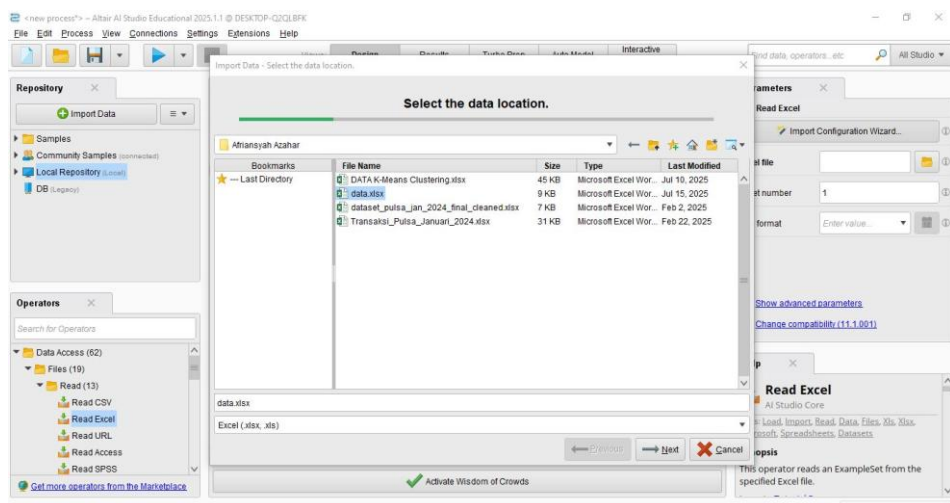
Gambar 4.3 Halaman Lembaran Kerja Baru *RapidMiner*

Selanjutnya proses yang dilakukan ialah pengimpuan data ke dalam lembar kerja. Maka membutuhkan operator *read excel* kemudian seperti gambar 4.4 lalu klik *import configuration wizard* pada parameters operator *read excel*.



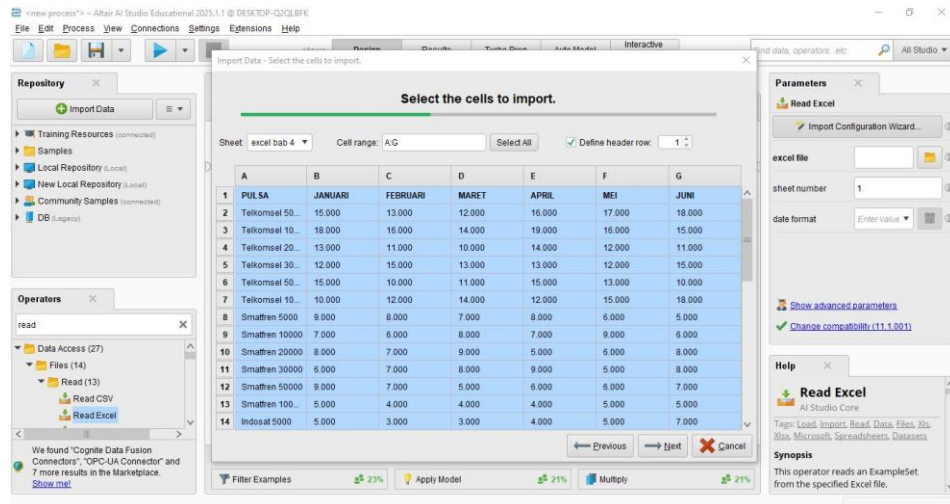
Gambar 4.4 Import Configuration Wizard

Setelah mengetik *import configuration wizard* maka akan muncul for data yang akan dipilih untuk diimport pada lembar kerja lalu klik *next*.



Gambar 4.5 Alur Proses Import Data Tahap 1

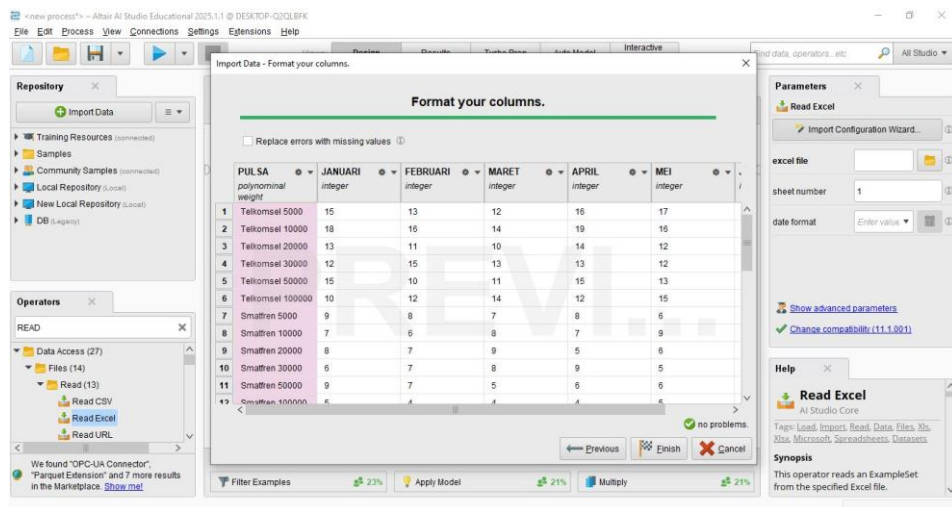
Pada tahap 1 memilih lokasi file data yang akan digunakan pilih lokasi penyimpanan data, pilih data yang akan digunakan kemudian klik *next*. Maka muncul tahap 2 yaitu *select the cells to import* seperti gambar 4.6 lalu klik *next*.



Gambar 4.6 Alur Import Data Tahap 2

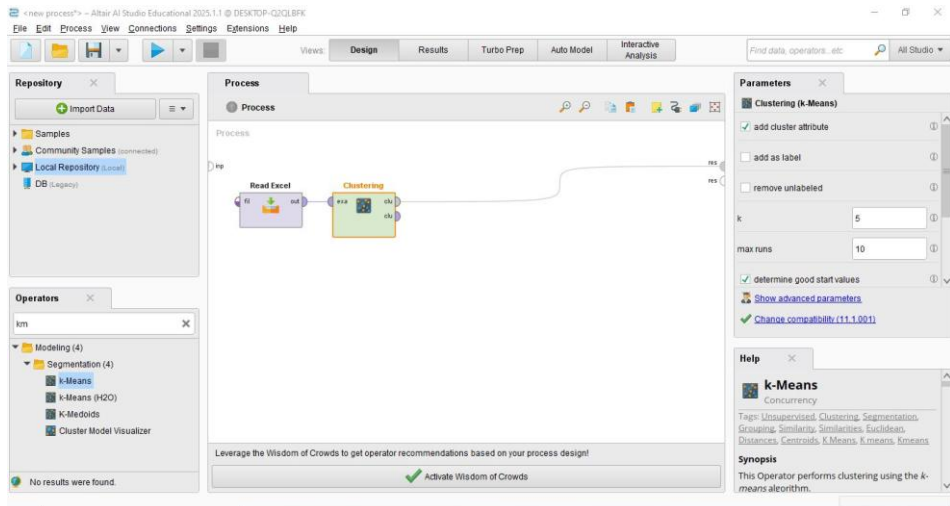
Pada tahap ke 3 tahap ini merupakan tahap penentuan tipe data dan atribut.

Pada atribut pulsa kita rubah tipe datanya menjadi *wight*, Kemudian Klik *next*.



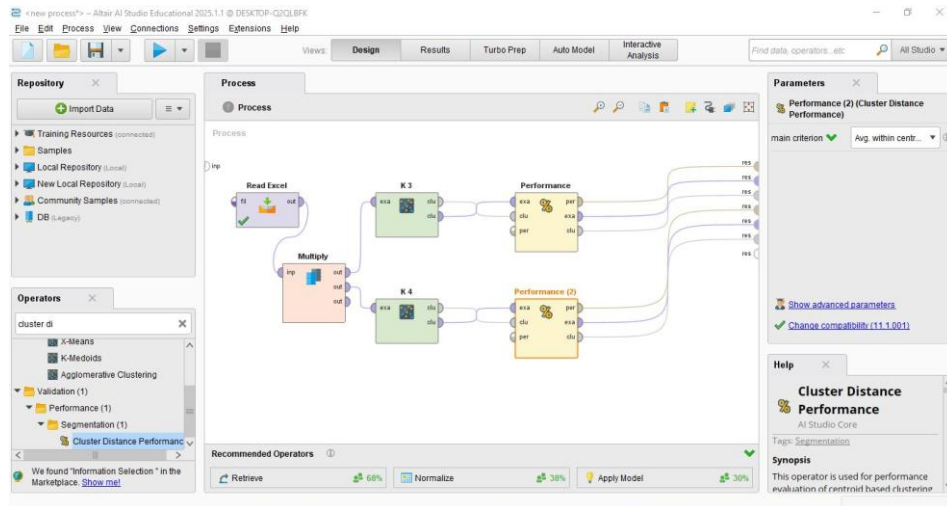
Gambar 4.7 Alur Import Data Tahap 3

Setelah tahapan *import* data selesai selanjutnya menghubungkan operator *read excel* dengan operator *k means*, *read excel* di drag dan drop kemudian hubungkan operator *clustering* dengan result.



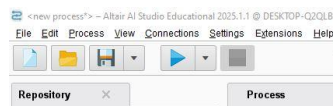
Gambar 4.8 Input Operator *Clustering*

RapidMiner akan membandingkan dua grup untuk menguji akurasi proses pengelompokan. $k = 3$ digunakan oleh grup pertama, dan $k = 4$ oleh grup kedua. Operator perkalian menggandakan jumlah operator pengelompokan pada platform, memungkinkan Anda menambahkan lebih dari satu operator pengelompokan di *RapidMiner*. Operator jarak klaster, darg, dan dop kemudian digunakan untuk mengevaluasi kinerja hasil pengelompokan. Selanjutnya, konfigurasi pengaturan operator ini; pengelompokan yang lebih baik dan lebih terorganisir ditunjukkan dengan nilai DBI yang lebih rendah.



Gambar 4.9 Susunan Operator *Read Excel*, *Mltiply*, *K-Means* dan *Perfomance*

Setelah di *drag* dan *drop* kemudian hubungkan operator *clustering* dengan operator *performance* dan hubungkan dan result. Kemudian klik tombol run untuk mengeksekusi dta dan menampilkan hasil.



Gambar 4.10 Ikon Tombol *Run*

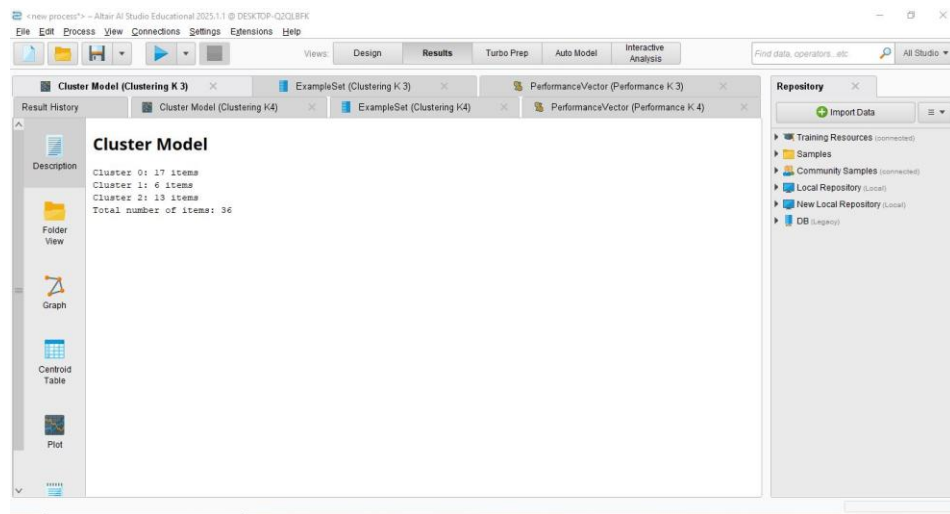
Maka akan muncul hasil *performance vector* dari kedua *clustering* yang dijalankan, sehingga dapat dilihat bahwasanya nilai DBI yang lebih rendah adalah nilai dari *clustering* dengan jumlah *cluster* sebanyak 3 degan nilai 0.674 dan *cluster* sebanyak 4 dengan nilai 0.861 maka yang akan di gunakan adalah clustering dengan jumlah klaster sebanyak 3. Selisih *Centroid Perfomancevector* K3 -17,916 centroid distance_cluster_0: -12.519, centroid distance_cluster_1: -31.694, centroid distance _cluster 2: -18.615. Seisih *Centroid Perfomancevector* K4 -15.676, centroid

distance_cluster_0: -18.615, centroid distance_cluster_1: -12.519, centroid distance_cluster_2: -10.250, centroid distance_cluster_3: -22.250.



Gambar 4.11 Evaluasi Hasil *Clustering* Menggunakan DBI

Dengan demikian, Model Klaster K 3 menghasilkan tiga kluster: kluster 0 memiliki total 17 item data, kluster 1 memiliki 6 item, dan kluster 2 memiliki total 13 item, sehingga totalnya 36 item data.



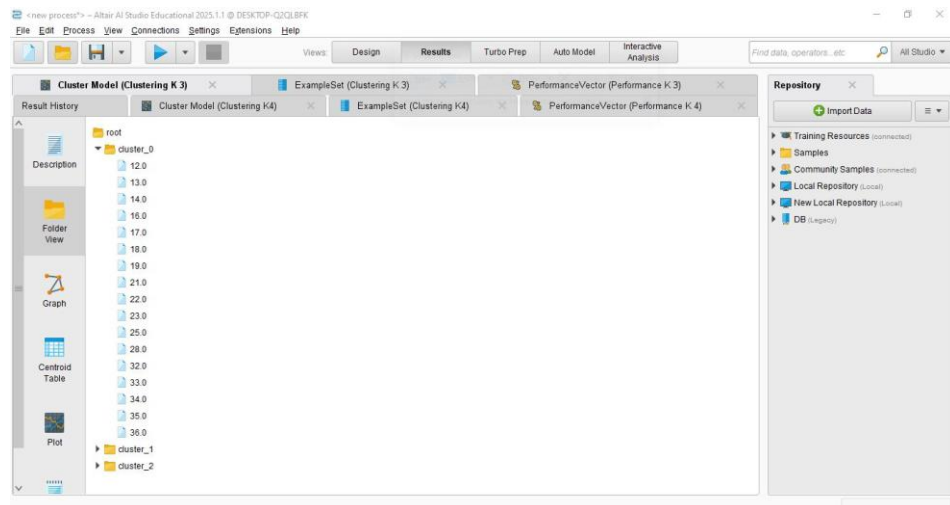
Gambar 4.12 Hasil *Clustering* Data K 3

Data yang disajikan merupakan hasil rekapitulasi penjualan pulsa dari berbagai operator dan nominal selama enam bulan, yaitu Januari hingga Juni. Setiap

baris data mewakili satu jenis produk pulsa dengan mencatat jumlah transaksi penjualan per bulan.

Hasil clustering K 3, dari anggota cluster 0 yang berisi yaitu :

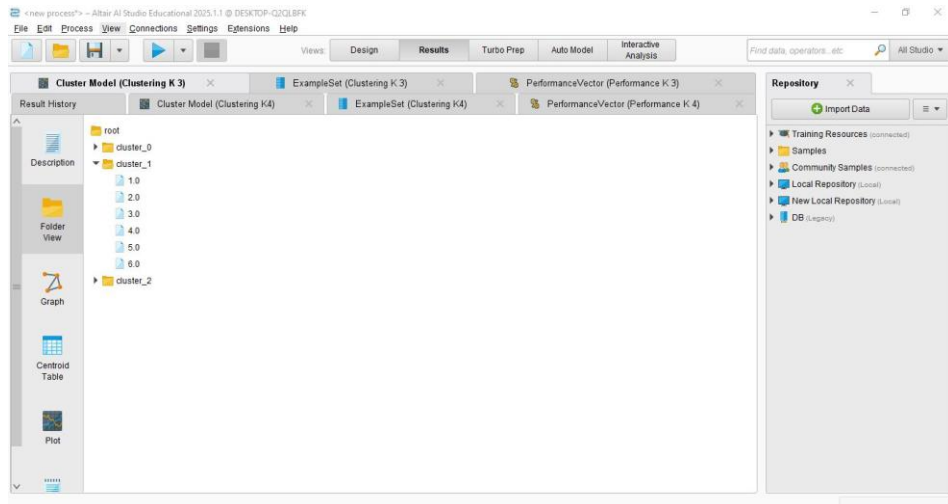
No 12 pulsa Smatfren, no 13 Indosat, no 14 Indosat, no 16 Indosat, no 17 Indosat, no 18 Indosat, no 19 XL, no 21 XL, no 22 XL, no 23 XL, no 25 Axis, no 28 Axis, no 32 Tri, no 33 Tri, no 34 Tri, no 35 Tri, dan terakhir no 36 Tri.



Gambar 4.13 Hasil Cluster 0

Hasil anggota *cluster* 1 yang berisi yaitu :

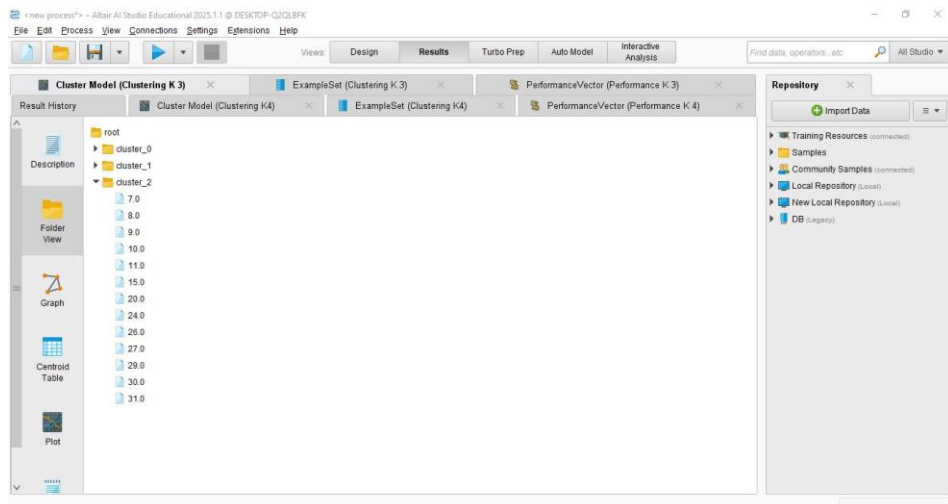
NO 1 pulsa Telkomsel, no 2 Telkomsel, no 3 Telkomsel, no 4 Telkomsel, no 5 Telkomsel dan terakhir no 6 telkomsel.



Gambar 4.14 Hasil *Cluster 1*

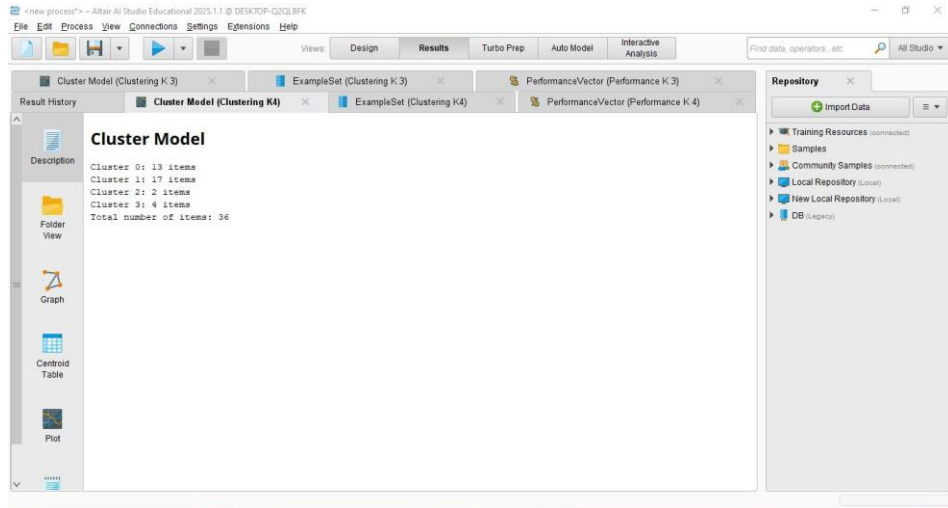
Hasil anggota *cluster 2* yang berisi yaitu :

No 7 pulsa Smatfren, no 8 Smatfren, no 9 Smatfren, no 10 Smatfren, no 11 Smatfren, no 15 Indosat no 20 XL, no 24 XL, no 26 Axis no 27 Axis, no 29 Axis, no 30 Axis dan terakhir no 31 Tri.



Gambar 4.15 Hasil *Cluster 2*

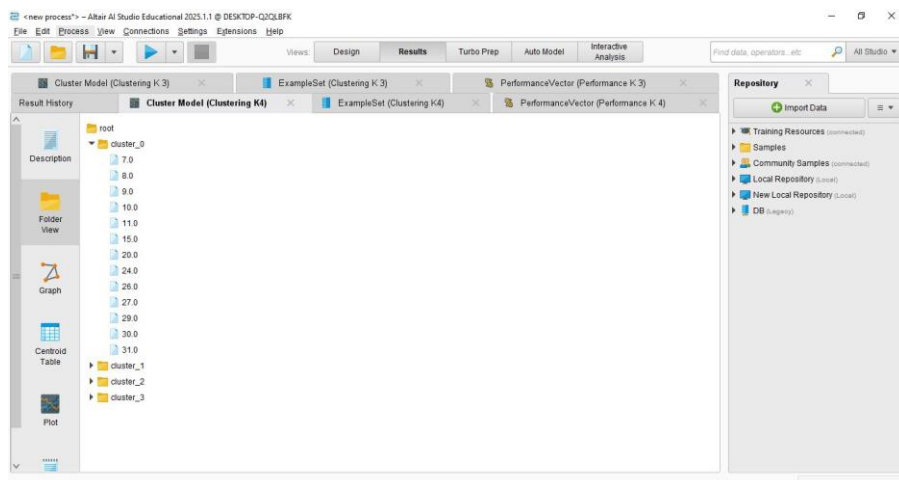
Selain itu, ada empat klaster sebagai hasil dari Model Klaster K4: klaster 0 memiliki total 13 item data, klaster 1 memiliki total 17 item, klaster 2 memiliki total 2 item, dan klaster 3 memiliki total 4 hal, sehingga totalnya 36 item data.



Gambar 4.16 Hasil Clustering Data K 4

Hasil *clustering* K 4, dari anggota *cluster* 0 yang berisi yaitu :

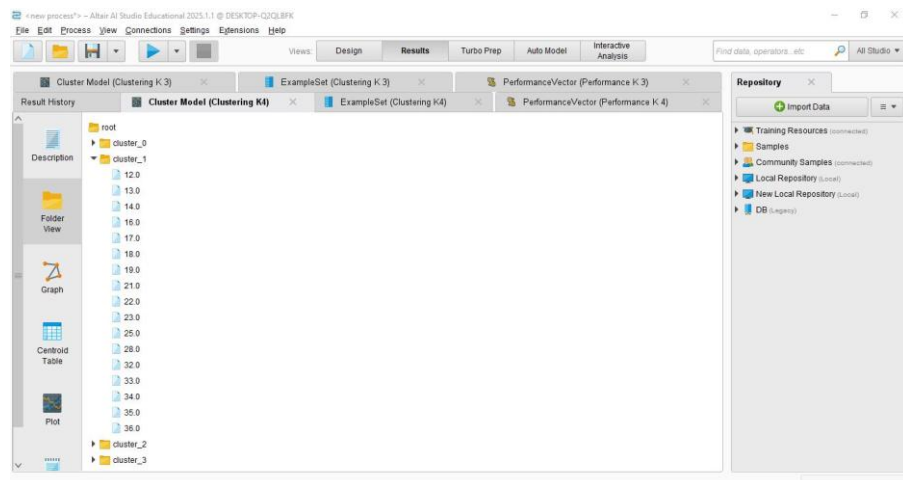
No 7 pulsa Smatfren, no 8 Smatfren, no 9 Smatfren, no 10 Smatfren, no 11 Smatfren, no 15 Indosat no 20 XL, no 24 XL, no 26 Axis no 27 Axis, no 29 Axis, no 30 Axis dan terakhir no 31 Tri.



Gambar 4.17 Hasil Cluster 0

Hasil *clustering* K 4, dari anggota *cluster* 1 yang berisi yaitu :

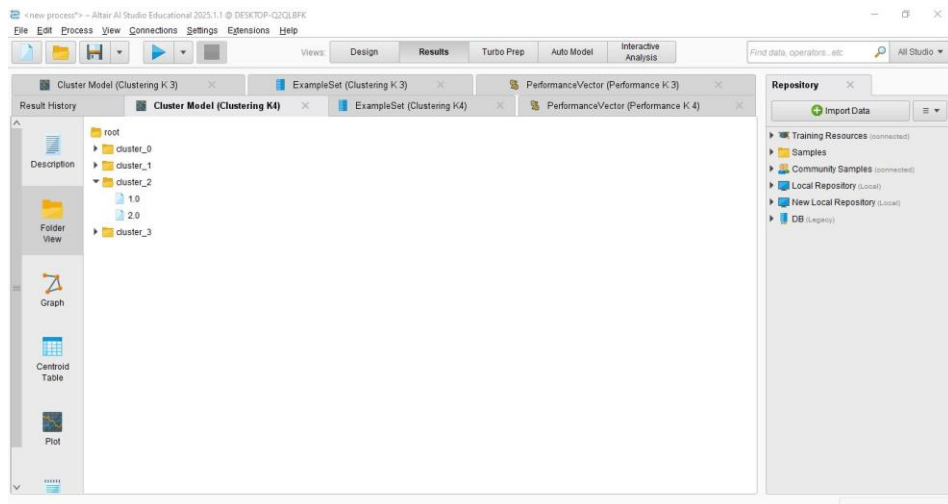
No 12 pulsa Smatfren, no 13 Indosat, no 14 Indosat, no 16 Indosat, no 17 Indosat, no 18 Indosat, no 19 XL, no 21 XL, no 22 XL, no 23 XL, no 25 Axis, no 28 Axis, no 32 Tri, no 33 Tri, no 34 Tri, no 35 Tri, dan terakhir no 36 Tri.



Gambar 4.18 Hasil Cluster 1

Hasil *clustering* K 4, dari anggota *cluster* 2 yang berisi yaitu :

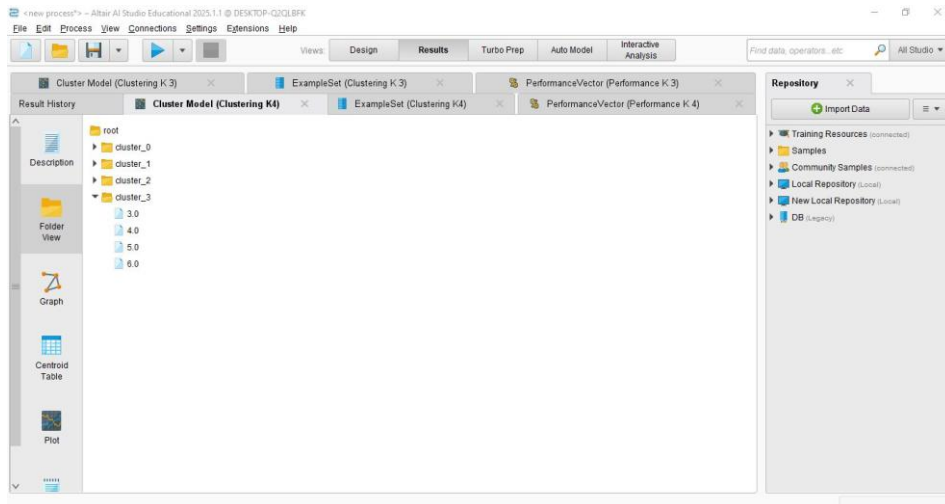
NO 1 pulsa Telkomsel dan no 2 Telkomsel.



Gambar 4.19 Hasil Cluster 2

Hasil *clustering* K 4, dari anggota *cluster* 2 yang berisi yaitu :

No 3 Telkomsel, no 4 Telkomsel, no 5 Telkomsel dan terakhir no 6 telkomsel.



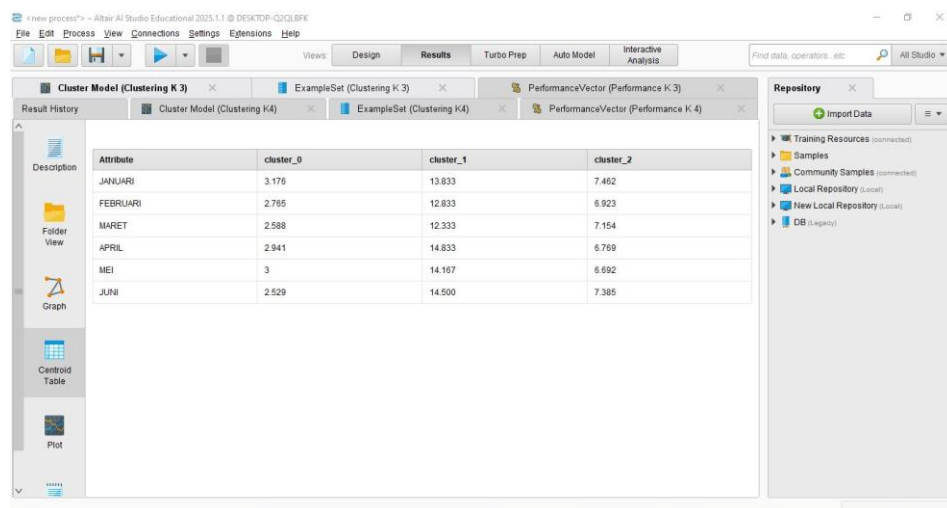
Gambar 4.20 Hasil Cluster 3

Dapat dilihat dari data pada gambar 4.21 menggambarkan persebaran kluster dalam bentuk *scatter bubble* yang di hasilkan dari proses pengelompokkan penjualan pulsa. *Scatter* bewarna orange melambangkan *cluster 0* *scater* bewarna hijau melambangkan *cluster 2* dan *scatter* bewarna biru melambangkan *cluster 1*. Pada *cluster 0* di kelompokkan sebagai frekuensi pembelian yang sedang, *cluster 1* sebagai frekuensi pembelian perbulan yang tinggi dan *cluster 2* sebagai frekuensi pembelian yang rendah.



Gambar 4.21 Visualisasi Pemodelan Clustering K 3

Selanjutnya hasil rata-rata *centroid* dari tiap *cluster* k 3 di bulan januari cluster 0 sebanyak 3.178, cluster 1 sebanyak 13.833, dan cluster 2 sebanyak 7.462, Selanjutnya di bulan februari cluster 0 sebanyak 2.765, cluster 1 sebanyak 12.833, dan cluster 2 sebanyak 6.923, Selanjutnya di bulan maret cluster 0 sebanyak 2.588, cluster 1 sebanyak 12.333, dan cluster 2 sebanyak 7.154, Selanjutnya di bulan april cluster 0 sebanyak 2.941, cluster 1 sebanyak 14.833, dan cluster 2 sebanyak 6.769, Selanjutnya di bulan mei cluster 0 sebanyak 3, cluster 1 sebanyak 14.167, dan cluster 2 sebanyak 5.802, dan terakhir di bulan juni cluster 0 sebanyak 2.529, cluster 1 sebanyak 14.500, dan cluster 2 sebanyak 7.385.



Attribute	cluster_0	cluster_1	cluster_2
JANUARI	3.178	13.833	7.462
FEBRUARI	2.765	12.833	6.923
MARET	2.588	12.333	7.154
APRIL	2.941	14.833	6.769
MEI	3	14.167	5.802
JUNI	2.529	14.500	7.385

Gambar 4.22 Hasil Rata-Rata *Centroid* Dari Tiap *Cluster* K 3

4.2. Pembahasan

Berdasarkan hasil implementasi metode *K-Means Clustering* yang telah dilakukan menggunakan aplikasi *RapidMiner*, diperoleh dua skenario pengelompokan data penjualan produk pulsa *digital*, yaitu dengan jumlah klaster $k = 3$ dan $k = 4$. Kedua hasil ini dianalisis untuk menentukan mana yang memberikan performa terbaik dalam mengelompokkan data berdasarkan frekuensi dan jumlah penjualan pulsa.

4.3. Perbandingan Hasil Pehitungan Manual *RapidMiner*

Terakhir memastikan keakuratan hasil klasterisasi maka melakukan perbandingan antara perhitungan manual menggunakan Microsoft Excel dan perhitungan otomatis menggunakan *RapidMiner*. Dari hasil perhitungan dari kedua metode ini menunjukkan adanya konsistensi dalam pengelompokkan penjualan pulsa dalam tiga *cluster* utama yaitu pulsa dengan minat tinggi, sedang dan rendah.

Perbandingan dalam perhitungan manual (*Microsoft Excel*):

1. Perhitungan dilakukan dengan metode *Eucliden Distance* untuk menentukan jarak setiap pusat *cluster*.
2. Setelah beberapa iterasi ,nilai *centroid* diperbarui hingga mencapai konvergensi.
3. Excel menghasilkan tiga kelompok dengan hasil yang sebanding dengan *RapidMiner*. Meskipun membutuhkan waktu lebih lama untuk perhitungan manual.

Perbandingan dalam perhitungan dengan *RapidMiner*:

1. *RapidMiner* mengotomatiskan proses klasterisasi dengan hasil yang lebih cepat dan akurat dibandingkan metode manual.
2. *RapidMiner* juga menghasilkan visualisasi klaster yang mempermudah interpretasi .
3. *RapidMiner* juga menghasilkan visualisasi klaster yang mempermudah interpretasi.