

BAB IV

HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Pengumpulan data dalam penelitian ini merupakan hasil dari proses pengumpulan data primer dan sekunder yang berisi informasi terkait karakteristik perilaku membaca siswa berdasarkan beberapa indikator utama, yaitu frekuensi membaca, durasi membaca, jenis bacaan, motivasi membaca, dan partisipasi dalam kegiatan literasi. Data ini dihimpun dengan tujuan untuk mendukung analisis tingkat minat baca pada masing-masing siswa dengan identitas tertentu. Kategori data diperoleh melalui observasi langsung di perpustakaan sekolah, penyebaran kuesioner kepada siswa, serta dokumentasi dari catatan kunjungan dan peminjaman buku di perpustakaan. Informasi yang diperoleh ini penting sebagai dasar untuk proses clusterisasi menggunakan algoritma K-Means, agar dapat mengelompokkan siswa dengan karakteristik membaca yang serupa secara sistematis dan objektif guna mendukung upaya sekolah dalam merancang strategi peningkatan minat baca yang lebih tepat sasaran.

Tabel 4. 1. Data Asli Penelitian

No.	Nama Siswa	Frekuensi Membaca	Durasi Membaca	Jenis Bacaan	Motivasi Membaca	Partisipasi
1	Abdullah Rambe	6	0.5	2	0	0
2	Abel Monautu Siregar	4	2.1	2	0	0
3	Adelia Marbun	2	2.6	0	0	1
4	Aldi Manosor Rajagukguk	4	1.8	3	0	1
5	Amanda Putra Manurung	1	2.7	2	2	2
6	Anggy Sri Mahesti Br Purba	5	2.5	3	0	1
7	Arifin Sitorus Pane	2	1.5	3	1	2

8	Arini Anastasya Marbun	2	1.9	1	1	2
9	Arsita Ayu Lestari	2	1	0	0	0
10	Arya Maulana	6	2.7	3	1	2
11	Arya Sadewa	6	2	2	2	1
12	Aulia Rahmadani Harahap	7	1.6	0	2	0
13	Berni Nduru	5	1.8	0	1	2
14	Boni Asian Panjaitan	1	2.2	3	2	1
15	Candra Putra Marbun	5	2.1	3	0	0
16	Candy Mayasari Br Silalahi	7	0.6	2	0	0
17	Celvin Lusiana Waruwu	1	1.7	3	2	1
18	Cenli Indriani Naibaho	5	2.3	1	2	2
19	Charli Billvans Sihite	6	1.5	3	0	1
20	Chokky Parlindungan Sitohang	4	1.4	3	2	1

Data pada tabel di atas merupakan sampel penelitian yang terdiri dari 20 siswa SD Negeri 01 Kotapinang dengan informasi mengenai perilaku membaca mereka. Setiap siswa direpresentasikan oleh lima variabel, yaitu frekuensi membaca, durasi membaca, jenis bacaan, motivasi membaca, dan partisipasi dalam kegiatan literasi. Seluruh variabel sudah dikonversi ke bentuk numerik agar dapat diproses secara lebih mudah dalam analisis menggunakan algoritma K-Means Clustering. Data sampel ini akan diolah dengan bantuan aplikasi RapidMiner untuk mengelompokkan siswa berdasarkan kesamaan perilaku membaca yang dimilikinya, sehingga menghasilkan informasi pola minat baca yang dapat dijadikan dasar bagi sekolah dalam merancang strategi peningkatan literasi siswa secara lebih tepat sasaran.

4.2. Preprocessing Data

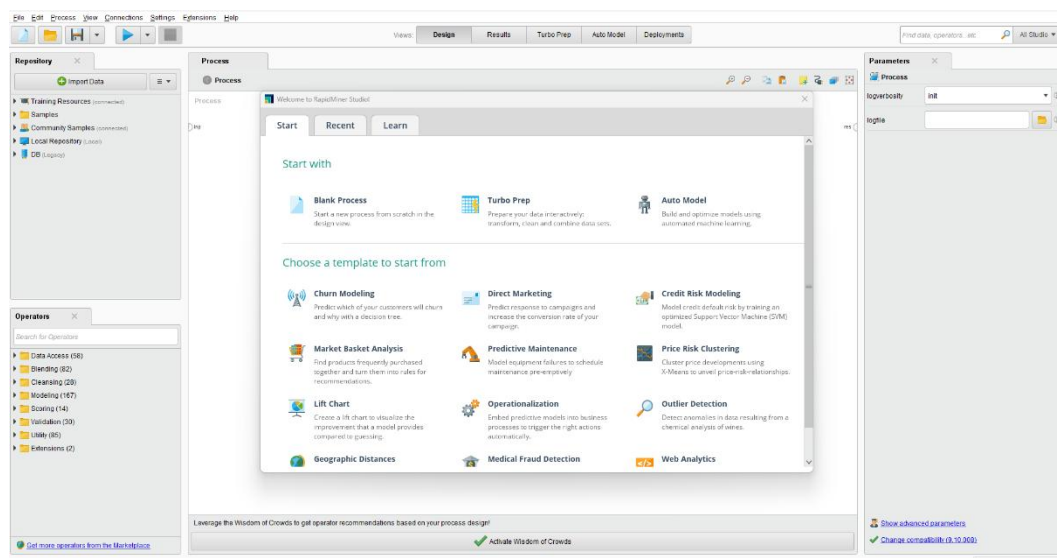
Tahapan preprocessing data dalam penelitian ini merupakan langkah penting yang bertujuan untuk mempersiapkan data agar siap dianalisis menggunakan algoritma K-Means. Data yang dikumpulkan dari siswa sebenarnya sudah tersusun

dengan baik, lengkap, dan rapi, sehingga tidak memerlukan banyak proses pembersihan atau perbaikan. Namun demikian, tetap dilakukan proses penting berupa transformasi data, yaitu mengubah variabel kategorikal seperti jenis bacaan, motivasi membaca, dan partisipasi menjadi bentuk numerikal sesuai dengan partisi yang telah ditentukan. Hal ini dilakukan karena algoritma K-Means hanya dapat mengolah atribut dalam bentuk angka. Dengan demikian, preprocessing pada penelitian ini lebih berfokus pada konversi format data agar sesuai dengan kebutuhan analisis clustering sehingga hasil pengelompokan dapat dilakukan secara akurat dan sistematis.

4.3. Implementasi Metode

4.3.1. Buka Aplikasi

Untuk memulai proses penelitian, langkah awal yang dilakukan adalah membuka aplikasi RapidMiner sebagai tools utama dalam pengolahan data. Setelah aplikasi terbuka, peneliti dapat langsung mengimpor dataset untuk kemudian diproses sesuai dengan tahapan analisis yang telah dirancang.

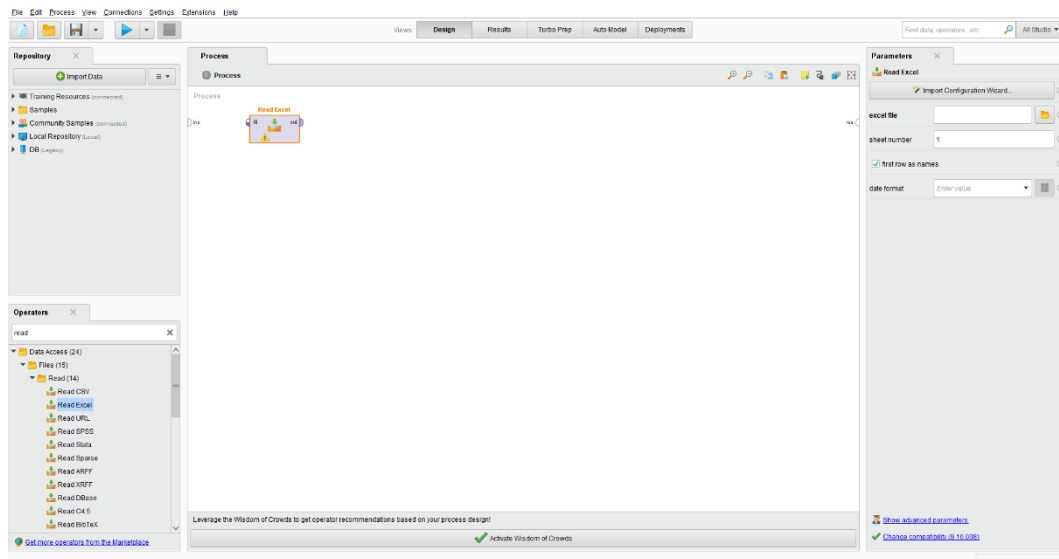


Gambar 4. 1. Tampilan Awal Aplikasi

Pada saat aplikasi RapidMiner pertama kali dibuka, pengguna akan disajikan dengan tampilan awal berupa halaman Start yang berfungsi sebagai beranda utama. Pada halaman ini terdapat beberapa bagian penting, yaitu menu Blank Document yang memungkinkan pengguna untuk memulai proyek baru dari awal, menu Recent yang menampilkan daftar dokumen atau proyek yang terakhir kali dibuka sehingga memudahkan untuk melanjutkan pekerjaan sebelumnya, serta menu Learn yang berisi materi pembelajaran, panduan, dan contoh proyek yang dapat digunakan sebagai referensi. Tampilan awal ini dirancang untuk mempermudah pengguna dalam memilih langkah awal sesuai kebutuhan, baik memulai proyek baru, membuka kembali proyek lama, maupun mempelajari lebih dalam tentang penggunaan aplikasi.

4.3.2. Tambah Operator Read Excel

Untuk melakukan proses analisis di RapidMiner, pengguna perlu menambahkan operator sesuai dengan kebutuhan, seperti operator *Read Excel*, *K-Means*, atau *Performance*. Penambahan operator ini dilakukan dengan cara menyeret (drag and drop) dari panel operator ke area proses kerja sehingga alur analisis dapat tersusun secara sistematis.

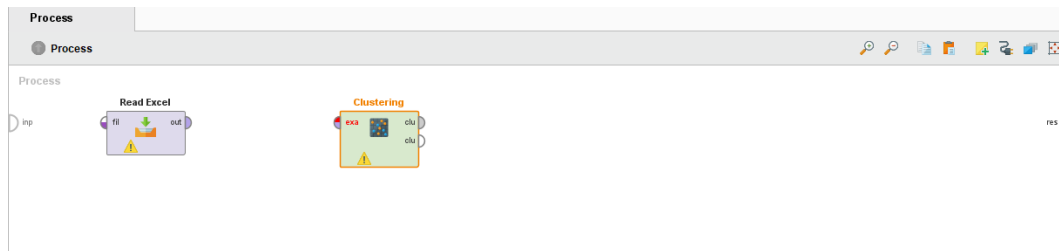


Gambar 4. 2. Operator Read Excel

Penambahan operator Read Excel pada aplikasi RapidMiner dilakukan untuk mengimpor dataset yang akan digunakan dalam penelitian ke dalam workspace analisis. Operator ini memungkinkan pengguna membaca file berformat Excel (.xls atau .xlsx) dengan mudah sehingga data dapat langsung diolah tanpa perlu konversi manual. Setelah operator ditambahkan ke area proses, pengguna dapat mengatur parameter seperti lokasi file, sheet yang digunakan, hingga opsi pembacaan baris header. Dengan adanya operator ini, data penelitian yang telah disusun sebelumnya dapat dimasukkan secara otomatis ke dalam sistem untuk diproses lebih lanjut menggunakan algoritma yang diinginkan.

4.3.3. Tambahkan Operator K-Means Clustering

Untuk memperluas fungsionalitas analisis, pengguna dapat menambahkan operator baru ke aplikasi RapidMiner sehingga memungkinkan integrasi berbagai metode pemrosesan, mulai dari preprocessing, klasifikasi, hingga evaluasi model.

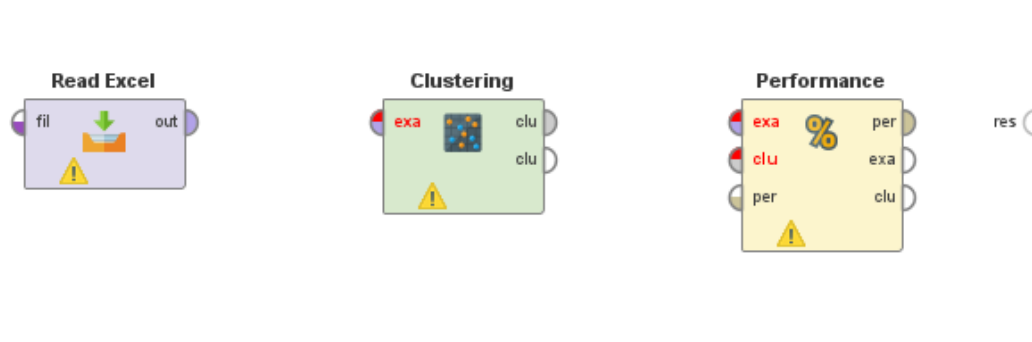


Gambar 4. 3. Penambahan Operator

Pada gambar di atas merupakan tampilan penambahan operator Read Excel dan K-Means pada aplikasi RapidMiner, di mana operator Read Excel berfungsi untuk mengimpor data dari file Excel ke dalam sistem sehingga data dapat diolah lebih lanjut, sedangkan operator K-Means digunakan untuk melakukan clustering dengan cara mengelompokkan data berdasarkan kesamaan atribut tertentu. Kombinasi kedua operator ini memudahkan pengguna untuk membaca data mentah langsung dari Excel dan kemudian mengelompokkannya menggunakan algoritma K-Means sehingga pola atau informasi tersembunyi dalam data dapat dianalisis secara lebih efektif.

4.3.4. Lengkapi Operator

Setelah semua operator yang dibutuhkan sudah ditambahkan ke area proses, langkah berikutnya adalah melengkapi alur kerja dengan menyambungkan setiap operator sesuai fungsinya. Proses penyambungan ini memastikan bahwa data dapat mengalir dengan benar dari input hingga menghasilkan output yang sesuai dengan tujuan analisis.



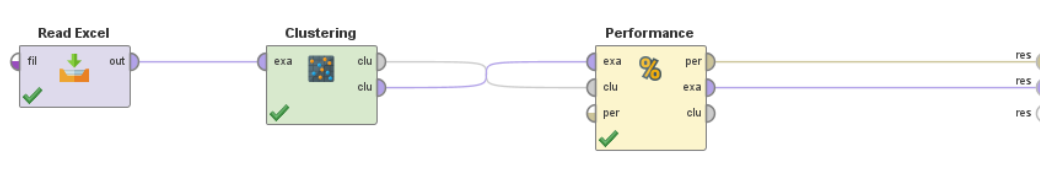
Gambar 4. 4. Lengkapi Operator

Pada gambar di atas merupakan tampilan rangkaian operator yang telah dilengkapi dengan Read Excel, Clustering (K-Means), dan Performance sebagai alur utama dalam proses analisis data. Operator Read Excel digunakan untuk membaca dan memasukkan dataset penelitian ke dalam RapidMiner, kemudian operator Clustering (K-Means) berfungsi untuk melakukan pengelompokan data berdasarkan kemiripan atribut yang dimiliki oleh setiap objek. Selanjutnya, operator Performance digunakan untuk mengevaluasi kualitas hasil cluster yang terbentuk sehingga dapat diketahui seberapa baik model K-Means dalam mengelompokkan data. Rangkaian operator ini membentuk alur kerja yang sistematis, mulai dari input data, proses analisis, hingga tahap evaluasi.

4.3.5. Hubungkan Operator

Untuk menjalankan proses analisis dengan benar, semua operator yang telah ditambahkan perlu dihubungkan sesuai alurnya, mulai dari **Read Excel** menuju **Clustering (K-Means)**, lalu dilanjutkan ke **Performance**. Penghubungan ini

memastikan bahwa data dapat mengalir secara berurutan dari tahap input, proses pengelompokan, hingga evaluasi hasil.



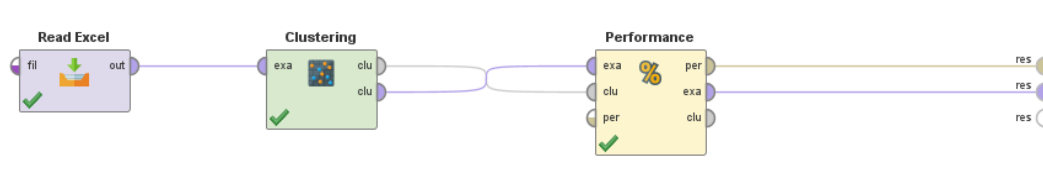
Gambar 4. 5. Hubungkan Operator

Pada gambar di atas merupakan model yang sudah lengkap dan seluruh operator telah dihubungkan sesuai alurnya sehingga membentuk proses analisis yang utuh. Model ini diawali dengan operator Read Excel sebagai input data, dilanjutkan dengan operator Clustering (K-Means) untuk melakukan pengelompokan data, dan diakhiri dengan operator Performance yang berfungsi mengevaluasi hasil cluster. Dengan tersusunnya model yang saling terhubung tersebut, sistem dapat dijalankan secara otomatis mulai dari membaca dataset, mengolah data berdasarkan algoritma K-Means, hingga menghasilkan output berupa evaluasi kinerja model. Hal ini menunjukkan bahwa rancangan analisis sudah siap digunakan untuk memperoleh hasil penelitian yang akurat.

4.3.6. Perancangan Model Cluster

Perancangan model Cluster merupakan tahap inti dalam penelitian ini yang bertujuan untuk membentuk kelompok siswa berdasarkan kemiripan karakteristik perilaku membaca menggunakan algoritma K-Means. Pada tahap ini, variabel-variabel seperti frekuensi membaca, durasi membaca, jenis bacaan, motivasi membaca, dan partisipasi dijadikan sebagai atribut utama untuk membentuk cluster yang merepresentasikan pola minat baca siswa. Proses perancangan model diawali dengan penentuan jumlah cluster yang optimal, kemudian dilanjutkan dengan

perhitungan jarak antar data dengan pusat cluster (centroid) hingga tercapai konvergensi, yaitu kondisi di mana pengelompokan tidak mengalami perubahan signifikan. Dengan merancang model cluster yang tepat, hasil pengelompokan ini diharapkan mampu mencerminkan kondisi riil minat baca siswa di lapangan serta dapat menjadi dasar yang kuat bagi pihak sekolah dalam merumuskan strategi peningkatan literasi.



Gambar 4. 6. Perancangan Model Cluster

Pada gambar di atas merupakan model proses clustering yang dirancang menggunakan aplikasi RapidMiner dengan algoritma K-Means. Proses diawali dengan operator Read Excel yang berfungsi untuk membaca dan memasukkan data penelitian yang berisi variabel frekuensi membaca, durasi membaca, jenis bacaan, motivasi membaca, dan partisipasi siswa. Selanjutnya, data tersebut diproses melalui operator Clustering dengan jumlah cluster ditentukan sebanyak $K=3$, sehingga data siswa akan dikelompokkan ke dalam tiga kelompok berdasarkan kemiripan karakteristik minat baca mereka. Tahap berikutnya adalah operator Performance, yang digunakan untuk mengevaluasi hasil clustering serta menampilkan kualitas pengelompokan yang terbentuk. Dengan alur ini, model mampu memberikan gambaran sistematis mengenai pola minat baca siswa di SD Negeri 01 Kotapinang sehingga dapat menjadi dasar dalam perumusan strategi peningkatan literasi di sekolah.

4.3.7. Hasil Cluster

Hasil *Cluster* yang telah diperoleh dari proses analisis menggunakan algoritma K-Means akan disajikan dalam bentuk tabel untuk memudahkan interpretasi. Tabel tersebut akan menampilkan masing-masing data beserta *Cluster* yang menjadi hasil akhir pengelompokan. Dalam penelitian ini, hasil yang disajikan mencakup dua skenario *Cluster*, yaitu dengan jumlah cluster $K = 3$ dan $K = 5$, sehingga memungkinkan perbandingan hasil pengelompokan berdasarkan variasi jumlah *Cluster* yang digunakan.

Tabel 4. 2. Hasil Cluster

No.	Nama Siswa	Frekuensi Membaca	Durasi Membaca	Jenis Bacaan	Motivasi Membaca	Partisipasi	Cluster
1	Abdullah Rambe	6	0.5	2	0	0	3
2	Abel Monautu Siregar	4	2.1	2	0	0	1
3	Adelia Marbun	2	2.6	0	0	1	2
4	Aldi Manosor Rajagukguk	4	1.8	3	0	1	1
5	Amanda Putra Manurung	1	2.7	2	2	2	2
6	Anggy Sri Mahesti Br Purba	5	2.5	3	0	1	1
7	Arifin Sitorus Pane	2	1.5	3	1	2	2
8	Arini Anastasya Marbun	2	1.9	1	1	2	2
9	Arsita Ayu Lestari	2	1	0	0	0	2
10	Arya Maulana	6	2.7	3	1	2	1
11	Arya Sadewa	6	2	2	2	1	3
12	Aulia Rahmadani Harahap	7	1.6	0	2	0	3
13	Berni Nduru	5	1.8	0	1	2	3
14	Boni Asian Panjaitan	1	2.2	3	2	1	2
15	Candra Putra Marbun	5	2.1	3	0	0	1
16	Candy Mayasari Br Silalahi	7	0.6	2	0	0	3
17	Celvin Lusiana Waruwu	1	1.7	3	2	1	2
18	Cenli Indriani Naibaho	5	2.3	1	2	2	3
19	Charli Billvans Sihite	6	1.5	3	0	1	1
20	Chokky Parlindungan Sitohang	4	1.4	3	2	1	1

Berdasarkan tabel hasil clustering di atas, dari total 20 data siswa, diperoleh pembagian ke dalam tiga cluster dengan jumlah yang berbeda. Cluster 1 (C1) beranggotakan 7 siswa, yaitu Abel Monautu Siregar, Aldi Manosor Rajagukguk, Anggy Sri Mahesti Br Purba, Arya Maulana, Candra Putra Marbun, Charli Billvans Sihite, dan Chokky Parlindungan Sitohang. Cluster 2 (C2) terdiri dari 7 siswa, yaitu Adelia Marbun, Amanda Putra Manurung, Arifin Sitorus Pane, Arini Anastasya Marbun, Arsita Ayu Lestari, Boni Asian Panjaitan, dan Calvin Lusiana Waruwu. Sementara itu, Cluster 3 (C3) beranggotakan 6 siswa, yaitu Abdullah Rambe, Arya Sadewa, Aulia Rahmadani Harahap, Berni Nduru, Candy Mayasari Br Silalahi, dan Cenli Indriani Naibaho.

Cluster 1 (C1) menggambarkan kelompok siswa dengan pola membaca cukup aktif, di mana frekuensi membaca mereka cenderung menengah hingga tinggi, dengan durasi membaca yang relatif bervariasi. Mereka lebih banyak memilih bacaan berupa majalah dan buku pelajaran, serta mengikuti kegiatan literasi seperti kunjungan perpustakaan atau klub buku. Karakteristik ini menunjukkan bahwa siswa dalam C1 sudah memiliki kebiasaan membaca yang cukup baik, meskipun motivasi mereka sebagian besar masih berkaitan dengan hiburan atau tuntutan tugas sekolah.

Sementara itu, Cluster 2 (C2) merepresentasikan kelompok siswa dengan frekuensi membaca rendah hingga sedang dan durasi membaca yang cenderung lebih panjang. Mereka banyak memilih bacaan digital, buku cerita, dan sebagian majalah, dengan motivasi yang lebih beragam, mulai dari hiburan hingga tugas sekolah. Partisipasi mereka juga bervariasi, meliputi kunjungan perpustakaan

maupun lomba. Adapun Cluster 3 (C3) mencerminkan kelompok siswa dengan frekuensi membaca cukup tinggi dan durasi membaca yang relatif singkat. Mereka lebih sering memilih buku pelajaran dan bacaan digital, dengan motivasi membaca yang didominasi oleh tugas maupun keinginan menambah pengetahuan. Siswa dalam kelompok ini juga cukup aktif dalam kegiatan literasi seperti klub buku dan lomba, sehingga C3 dapat dianggap sebagai cluster dengan tingkat minat baca paling konsisten.

4.3.8. Hasil Evaluasi

Hasil evaluasi menggunakan metrik Average Within Centroid Distance pada penelitian ini masih belum diperoleh. Tahap evaluasi ini sebenarnya bertujuan untuk mengukur seberapa rapat dan kompak data yang tergabung dalam setiap cluster terhadap pusat clusternya. Semakin kecil nilai yang dihasilkan, maka semakin baik kualitas pengelompokan yang terbentuk.

4.3.9. Hasil Avg. Within Centroid Distance

Average Within Centroid Distance adalah ukuran yang menunjukkan rata-rata jarak antara setiap titik data dalam suatu *Cluster* dengan pusat *Cluster* (centroid)-nya. Nilai ini digunakan untuk menilai seberapa rapat data dalam *Cluster*; semakin kecil nilainya, semakin baik kekompakan *Cluster* tersebut.

Avg. within centroid distance

Avg. within centroid distance: -2.932

Gambar 4. 7. Hasil Evaluasi

Berdasarkan hasil evaluasi yang ditunjukkan pada gambar, nilai Average Within Centroid Distance (AWCD) yang diperoleh adalah sebesar -2,932. Metrik.