



**PENERAPAN ALGORITMA *K-MEANS* UNTUK MENKLAUSTER  
DATA NILAI BERPRESTASI PADA SMP S**

**LAPORAN TUGAS DATA MINING**

**DISUSUN OLEH:**

<b>20.01.013.064</b>	<b>ABIGAIL PERKASA</b>
<b>20.01.013.067</b>	<b>SAHARA HASIBUAN</b>
<b>20.01.013.041</b>	<b>SHAKIRA AZZAHRA HADI P.</b>
<b>20.01.013.013</b>	<b>NUR ISLAMIA</b>

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS REKAYASA SISTEM**

**UNIVERSITAS TEKNOLOGI SUMBAWA**

**2023**

## ABSTRAK

Data siswa setiap tahunnya terus bertambah dan menghasilkan data yang berlimpah sehingga terjadi penumpukan data. Data yang berlimpah perlu dilakukan pengolahan data untuk menggali informasi yang terdapat didalam data tersebut. Tujuan penelitian ini untuk mengkluster data siswa melalui proses data mining dengan menggunakan algoritma *K-Means* untuk pembentukan *cluster*. Atribut data digunakan adalah nama siswa dan nilai dari beberapa mata pelajaran. Data yang digunakan adalah data siswa dengan sampel data 106 items dan sumber data berasal dari website. *Cluster* siswa yang terbentuk ada 3 yaitu *cluster* 0 memiliki 27 anggota, *cluster* 1 memiliki 45 anggota, dan *cluster* 2 memiliki 34c anggota dari total 106 dataset yang di uji. Hasil dari penelitian ini di gunakan sebagai dasar untuk menentukan siswa yang berprestasi yang ada di data ini berdasarkan hasil *clustering*.

*Keywords: K-Means, Clustering, Data Mining, Data, Cluster*

## **KATA PENGANTAR**

Segala puji hanya milik Allah SWT, shalawat serta salam selalu tercurahkan kepada Rasulullah SAW. Berkat limpahan rahmat-Nya, penulis mampu menyelesaikan laporan ini guna memenuhi tugas mata kuliah Data Mining. Cukup banyak hambatan yang penulis hadapi dalam penyusunan laporan ini. Namun penulis menyadari bahwa kelancaran dalam penyusunan materi ini tidak lain berkat bantuan, dorongan dan bimbingan orang tua serta ibu Ekastini, M.Kom selaku dosen mata kuliah Data Mining sehingga kendala-kendala yang penulis hadapi dapat teratasi.

Laporan disusun agar pembaca dapat memperluas wawasan mengenai penelitian menggunakan metode pendekatan K-Means yang penulis sajikan berdasarkan pengamatan dari berbagai sumber informasi, referensi dan contoh soal. Makalah ini penulis susun dengan berbagai rintangan baik itu yang datang dari diri penulis maupun yang datang dari luar. Namun dengan penuh kesabaran dan terutama pertolongan dari Allah akhirnya laporan ini dapat terselesaikan.

Semoga makalah ini dapat memberikan wawasan yang lebih luas dan menjadi sumbangan pemikiran kepada pembaca khususnya para mahasiswa Universitas Teknologi Sumbawa. Penulis sadar bahwa laporan ini masih banyak kekurangan dan jauh dari sempurna. Untuk itu, kepada dosen pengampu, penulis meminta masukannya demi perbaikan pembuatan laporan penulis di masa yang akan datang dan mengharapkan kritik dan saran dari para pembaca.

Sumbawa Besar, 17 Maret 2023

Penyusun

## DAFTAR ISI

COVER .....	i
ABSTRAK .....	ii
KATA PENGANTAR .....	iii
DAFTAR ISI.....	iv
DAFTAR TABEL.....	vi
DAFTAR GAMBAR .....	vii
BAB I PENDAHULUAN .....	1
1.1    Latar Belakang .....	1
1.2    Rumusan Masalah .....	1
1.3    Batasan Masalah.....	1
1.4    Manfaat dan Tujuan.....	2
1.5    Sistematika Penulisan.....	2
BAB II LANDASAN TEORI .....	4
2.1    Penelitian Terdahulu.....	4
2.2    Landasan Teori .....	6
BAB 3 METODOLOGI PENELITIAN.....	16
3.1    Objek Penelitian .....	16
3.2    Waktu dan Tempat Penelitian .....	16
3.3    Jenis dan Sumber Data .....	16
BAB IV HASIL DAN PEMBAHASAN .....	18
4.1    Proses <i>Clustering</i> .....	18
4.2    Pengujian <i>Rapidminer</i> .....	23
4.3    Pembahasan Hasil <i>Clustering</i> K-Means.....	30

BAB V PENUTUP.....	31
5.1    Kesimpulan.....	31
5.2    Saran .....	31
DAFTAR PUSTAKA .....	32

## DAFTAR TABEL

Tabel 1. Penelitian Terdahulu .....	4
Tabel 2. Data <i>Clustering</i> .....	18

## DAFTAR GAMBAR

Gambar 1. Penerapan Data Mining .....	7
Gambar 2. Metode Data Mining .....	8
Gambar 3. <i>Design Process</i> .....	23
Gambar 4. <i>Example Set Result</i> .....	24
Gambar 5. <i>Cluster Model</i> .....	24
Gambar 6. <i>Centroid Table</i> .....	25
Gambar 7. Anggota <i>Cluster 0</i> .....	26
Gambar 8. Anggota <i>Cluster 1</i> .....	26
Gambar 9. Anggota <i>Cluster 2</i> .....	27
Gambar 12. <i>Example Set Statistic</i> .....	27
Gambar 13. <i>Graph Result</i> .....	28
Gambar 14. <i>Plot Result</i> .....	29
Gambar 15. <i>Chart Clustering</i> .....	29
Gambar 16. Hasil Rata-Rata <i>Clustering</i> .....	30

## **BAB I PENDAHULUAN**

### **1.1 Latar Belakang**

Pada zaman sekarang ini, perkembangan pengolahan data elektronika telah menjadi kebutuhan yang sangat utama. Perkembangan pesat dalam teknologi informasi yang menjadikan semua informasi dapat disimpan dalam jaringan komputer telah membuat munculnya sistem basis data yang sangat besar. Dalam hitungan detik, data-data dalam berbagai basis data akan senantiasa terbarukan, dikarenakan adanya *update* maupun perubahan data baru. Permasalahan yang kemudian muncul adalah bagaimana mengetahui informasi yang terdapat dalam basis data yang sangat besar.

Dalam hal ini siswa yang berkompeten dalam bidang akademik berupa pengetahuan dan keterampilan. Salah satu cara untuk mengukur kompetensi akademik adalah dengan nilai mata pelajaran. Permendikbud nomor 66 tahun 2013 mengenai Standar Penilaian Pendidikan menyebutkan bahwa penilaian adalah proses pengumpulan dan pengolahan informasi untuk mengukur capaian hasil belajar peserta didik, dalam hal ini adalah siswa. Sedangkan untuk mengetahui tingkat kompeten akademik, diperlukan suatu prosedur penilaian yakni tes atau ujian.

### **1.2 Rumusan Masalah**

Berdasarkan latar belakang masalah di atas, maka rumusan masalah yang ada adalah bagaimana cara melakukan prediksi tingkat prestasi murid yang rendah dan tinggi dengan menggunakan metode *K-Means*?

### **1.3 Batasan Masalah**

Agar pembahasan dalam penelitian ini dapat lebih jelas dan terarah maka penulis memberikan batasan terhadap permasalahan yang akan diteliti yang berfokus pada:



1. Hanya membahas tentang prediksi tingkat prestasi murid yang rendah dan tinggi.
2. Data yang diambil hanya data tentang nilai siswa.
3. Data yang diolah dalam penelitian ini menggunakan metode K-Means.
4. Tools yang digunakan dalam penerapan data mining ini adalah *Rapid Miner*.

#### **1.4 Manfaat dan Tujuan**

Adapun manfaat dan tujuan dari penelitian adalah sebagai berikut :

1. Mempermudah pengelompokan data dalam mengevaluasi dan menganalisis tingkat prestasi belajar murid yang rendah dan tinggi pada SMP S.

#### **1.5 Sistematika Penulisan**

Sistematika penulisan tugas akhir ini disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan tugas akhir ini adalah sebagai berikut:

### **BAB I. PENDAHULUAN**

Bab ini berisi latar belakang masalah, identifikasi masalah, maksud dan tujuan yang ingin dicapai, batasan masalah, metodologi penelitian yang diterapkan dalam memperoleh dan mengumpulkan data serta sistematika penulisan.

### **BAB II. LANDASAN TEORI**

Menjelaskan tentang kajian pustaka serta teori yang melandasi penelitian algoritma *clustering* untuk menentukan siswa berprestasi.

### **BAB III. METODE PENELITIAN**

Menjelaskan tentang metode penelitian dari pengumpulan data eksperimen dengan menguji data yang ada menggunakan algoritma *clustering* yang memprediksi siswa berprestasi.

#### BAB IV. HASIL DAN PEMBAHASAN

Menjelaskan dan menampilkan hasil prediksi dengan menggunakan algoritma *clustering*.

#### BAB V. PENUTUP

Berisi kesimpulan dari implementasi dan uji coba yang dilakukan. Selain itu berisi pula saran yang diharapkan dapat menjadi masukan untuk pengembangan di masa datang.

## BAB II LANDASAN TEORI

### 2.1 Penelitian Terdahulu

Tabel 1. Penelitian Terdahulu

No	Penulis	Judul Jurnal	Pembahasan
1	Berlian Juliarta Martin Putra, Dwi Ariani Finda Yuniarti (2020)	Analisis Hasil Belajar Mahasiswa dengan <i>Clustering</i> Menggunakan Metode <i>K-Means</i>	1. Penelitian menggunakan metode <i>K-Means</i> dengan tools <i>Waikato Environment</i> <i>for Knowledge</i> <i>Analysis</i> (WEKA). 2. Pengelompokkan hasil belajar mahasiswa menggunakan variabel IPK.
2	Shefia Natalia Br Sembiring, Hendryan Winata, Sri Kusnasari (2022)	Pengelompokan Prestasi Siswa Menggunakan Algoritma <i>K-Means</i>	1. Pengelompokkan data menggunakan teknik <i>clustering</i> . 2. Menggunakan algoritma <i>K-Means</i> . 3. Pengelompokkan data menggunakan nilai <i>centroid</i> dan mencari nilai terdekat 4. Hasil akhir yaitu pembuatan aplikasi berbasis desktop
3	Fajar Nur Rohmat	Implementasi	- Menggunakan metode

	Fauzan Jaya Aziz, Budi Darma Setiawan, Issa Arwani (2018)	Algoritma <i>K-Means</i> untuk Klustering Kinerja Akademik Mahasiswa	<i>K-Means clustering.</i> - Data yang digunakan adalah data mahasiswa dan data dasar.
4	Juniar Hutagalung, Yopi Hendro Syahputra, Zohana Pertiwi Tanjung (2022)	Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma <i>K-Means</i> <i>Clustering</i>	- Sampel berasal dari nilai raport siswa. - Pengelompokkan data menggunakan nilai <i>centroid</i> dan mencari nilai terdekat.
5	Ari Sulistiyawati, Eko Supriyanto (2018)	Implementasi Algoritma <i>K-Means</i> <i>Clustering</i> dalam Penentuan Siswa Kelas Unggulan	- Dilakukan dengan metode pengembangan SDLC dengan model sekuensial linier atau alur hidup klasik. - Hasil penelitian berupa aplikasi pengelolaan data penelitian yang terpusat dalam menyajikan laporan. - Data yang digunakan adalah nilai raport .

## **2.2 Landasan Teori**

### **2.2.1 Data Mining**

Perkembangan teknologi informasi telah memberikan kontribusi pada cepatnya pertumbuhan jumlah data yang dikumpulkan dan disimpan dalam basis data berukuran besar (*big data*). *Big data* adalah istilah yang menggambarkan volume data yang besar, baik data yang terstruktur maupun data yang tidak terstruktur. *Big data* memiliki potensi tinggi untuk mengumpulkan wawasan kunci dari informasi bisnis. *Big data* dapat dianalisis untuk wawasan yang mengarah pada pengambilan keputusan dan strategi bisnis yang lebih baik.

Sebuah metode atau teknik diperlukan untuk dapat merubah data tersebut menjadi sebuah informasi berharga atau pengetahuan yang bermanfaat untuk mendukung pengambilan keputusan. Suatu teknologi yang dapat digunakan untuk mewujudkannya adalah data mining. Belakangan ini data mining telah diimplementasikan kedalam berbagai bidang, diantaranya dalam bidang bisnis atau perdagangan, bidang pendidikan, dan telekomunikasi.

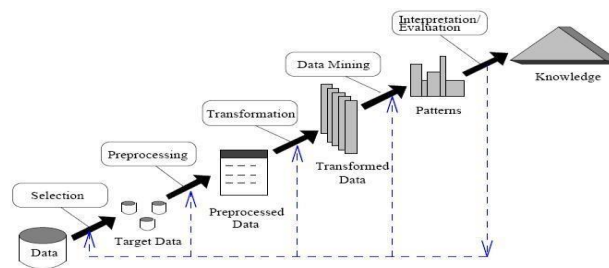
Menurut Stanton (2013:173) data mining adalah bidang penelitian dan praktik yang berfokus pada penemuan pola-pola baru dalam data yang mengacu pada penggunaan algoritma dan komputer untuk menemukan pola baru dan menarik dalam data.

Menurut Pramudiono dalam Baskoro, dkk (2013:42) data mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaanya.

Menurut Suyatno (2017:2) data mining ditujukan untuk mengekstrak pengetahuan dari sekumpulan data sehingga didapatkan struktur yang dapat dimengerti manusia serta meliputi basis data dan manajemen data, prapemrosesan data, pertimbangan model dan inferensi, ukuran ketertarikan,

pertimbangan kompleksitas, pascapemrosesan terhadap struktur yang ditemukan, visualisasi dan online updating.

Sebagai teknologi umum, data mining dapat diterapkan ke semua jenis data selama data bermakna untuk aplikasi target. Bentuk data paling dasar untuk penambangan aplikasi adalah *database*, data *warehouse* dan data transaksional. Data mining juga dapat diterapkan ke bentuk data lain (misalnya, aliran data, data urutan/ urutan, grafik atau data jaringan, data spasial, data teks, data multimedia).



Gambar 1. Penerapan Data Mining

#### 2.2.1.1 Fungsi Data Mining

Secara umum, kegunaan data mining terbagi menjadi dua yaitu deskriptif dan prediktif. Deskriptif memiliki arti untuk mencari pola-pola yang dapat dipahami manusia yang menjelaskan karakteristik data sedangkan prediktif digunakan untuk membentuk sebuah model pengetahuan guna melakukan prediksi. Berdasarkan fungsionalitasnya, tugas-tugas data mining bisa dikelompokkan menjadi enam kelompok yaitu :



Gambar 2. Metode Data Mining

Adapun penjelasan rinci dari enam kelompok tersebut sebagai berikut:

1. Klasifikasi (*classification*)

Proses generalisasi struktur yang diketahui untuk diaplikasikan pada data-data baru.

2. Klasterisasi (*clustering*)

Mengelompokkan data yang belum diketahui label kelasnya ke dalam sejumlah kelompok tertentu sesuai dengan ukuran kemiripannya.

3. Regresi (*regression*)

Menemukan suatu fungsi yang memodelkan data dengan kesalahan prediksi seminimal mungkin.

4. Deteksi anomali (*anomaly detection*)

Mengidentifikasi data yang tidak umum, berupa *outlier* (pencilan), perubahan atau deviasi yang mungkin sangat penting dan perlu investigasi lebih lanjut.

5. Pemodelan ketergantungan (*Dependency modeling*)

Mencari relasi antar tabel.

## 6. Perangkuman (*summarization*)

Menyediakan representasi data yang lebih sederhana, meliputi visualisasi dan pembuatan laporan.

### 2.2.1.2 Teknik Pembelajaran Data Mining

Teknik yang digunakan dalam data mining erat kaitannya dengan penemuan dan pembelajaran yang terbagi dalam tiga metode utama pembelajaran yaitu :

#### 1. *Supervised learning*

Teknik yang melibatkan fase pelatihan dimana data pelatihan historis yang karakterkarakternya dipetakan ke hasil-hasil yang telah diketahui dan diolah dalam algoritma data mining. Proses ini melatih algoritma untuk mengenali variabel-variabel dan nilai-nilai kunci yang nantinya akan digunakan sebagai dasar dalam membuat perkiraan-perkiraan ketika diberikan data baru.

#### 2. *Unsupervised learning*

Teknik pembelajaran yang tidak melibatkan fase pelatihan seperti *supervised learning* yakni bergantung pada penggunaan algoritma yang mendeteksi semua pola yang muncul dari kriteria penting yang spesifik dalam data masukan. Pendekatan ini mengarah pada pembuatan banyak aturan yang mengkarakteristikan penemuan *associations*, *clusters* dan *segment* yang kemudian dianalisis untuk menemukan hal-hal yang penting.

#### 3. *Reinforcement learning*

Teknik yang memiliki penerapan-penerapan yang terus dioptimalkan dari waktu ke waktu dan memiliki kontrol adaptif. Menyerupai kehidupan nyata yaitu seperti “*on job training*” dimana seorang pekerja diberikan sekumpulan tugas yang membutuhkan



keputusan-keputusan. *Reinforcement learning* sangat tepat digunakan untuk menyelesaikan masalah-masalah sulit yang bergantung pada waktu.

### **2.2.1.3 Proses Data Mining**

Data mining biasanya terdiri dari empat proses (Stanton 2013:173) :

1. Persiapan data

Melibatkan memastikan bahwa data diatur dengan cara yang benar, bahwa bidang data yang hilang terisi, bahwa data yang tidak akurat berada dan diperbaiki atau dihapus, dan data tersebut "didaur ulang" seperlunya.

2. Analisis data eksploratori

Proses eksplorasi juga melibatkan mencari keluar nilai-nilai yang tepat untuk parameter kunci.

3. Pengembangan model

Yaitu menguji pilihan penambangan data yang paling sesuai teknik. Tergantung pada struktur dataset dan memilih yang paling menjanjikan di dalamnya sebagai sains.

4. Interpretasi hasil.

Berfokus untuk memahami dari apa algoritma data mining telah dihasilkan yang merupakan langkah penting dari perspektif pengguna data, karena ini adalah tempat kesimpulan yang dapat ditindaklanjuti terbentuk.

Beberapa tahun terakhir data tumbuh menjadi semakin heterogen dan kompleks dengan volume yang meningkat cepat secara eksponensial. Selain itu, beberapa faktor pendorong kemajuan yang berlanjut dalam bidang data mining ialah:

- a Pertumbuhan yang cepat dalam pertumbuhan data.
- b Penyimpanan data dalam data warehouse, sehingga seluruh perusahaan memiliki akses ke dalam database yang handal.
- c Adanya peningkatan akses data melalui navigasi web dan internet.
- d Perkembangan teknologi perangkat lunak untuk data mining (ketersediaan teknologi).

- e Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

#### **2.2.1.4 Pengelompokan Teknik Data Mining**

Menurut Baskoro,dkk (2013:43) data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

1. *Classification*

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan.

2. *Association*

Digunakan untuk mengenali kelakuan dari kejadian-kejadian khusus atau proses dimana hubungan asosiasi muncul pada setiap kejadian.

3. *Clustering*

Digunakan untuk menganalisis pengelompokan berbeda terhadap data, mirip dengan klasifikasi, namun pengelompokan belum didefinisikan sebelum dijalankannya *tool* data mining.

4. *Forecasting*

Teknik *forecasting* sebagai input kemudian akan mengambil sederetan angka yang menunjukkan nilai yang berjalan seiring waktu dan kemudian Teknik forecasting ini akan menghubungkan nilai masa depan dengan menggunakan bermacam-macam teknik *machine learning* dan teknik statistik yang berhubungan dengan musim, *trend*, dan *noise* pada data.

5. *Prediction*

*Prediction* (prediksi), untuk memperkirakan nilai masa mendatang, misalnya memprediksi stok barang satu tahun ke depan.

## 2.2.2 *K-Means*

### 2.2.2.1 Pengertian *K-Means*

*K-means* merupakan algoritma klasterisasi yang paling tua dan paling banyak digunakan dalam berbagai aplikasi kecil hingga menengah karena kemudahan implementasinya. Menurut Suyanto (2017:262) Algoritma *k-means* bekerja dengan empat langkah, yaitu :

1. Himpunan data yang akan diklasterisasi, dipilih sejumlah  $k$  objek secara acak sebagai *centroid* awal.
2. Setiap objek yang bukan *centroid* dimasukkan ke *cluster* terdekat berdasarkan ukuran jarak tertentu.
3. Setiap *centroid* diperbarui berdasarkan rata-rata dari objek yang ada di dalam setiap *cluster*.
4. Langkah kedua dan ketiga dilakukan secara diulang-ulang (diiterasi) sampai semua *centroid* stabil dalam arti semua *centroid* yang dihasilkan dalam iterasi saat ini sama dengan semua *centroid* yang dihasilkan pada iterasi sebelumnya.

Berikut ini adalah langkah-langkah algoritma *k-means*:

1. Penentuan cluster awal

Dalam menentukan  $n$  buah pusat *cluster* awal dilakukan pembangkitan bilangan random yang merepresentasikan urutan data input. Pusat awal *cluster* didapatkan dari data sendiri bukan dengan menentukan titik baru, yaitu dengan random pusat awal dari data.

2. Perhitungan jarak dengan pusat *cluster*

Untuk mengukur jarak antar data dengan pusat dengan *cluster* digunakan *euclidian distance*, algoritma perhitungan jarak data dengan pusat *cluster*:

- a Pilih nilai data dan nilai pusat *cluster*
- b Hitung *euclidian distance* data dengan tiap pusat *cluster*

$$d(x_i, u_j) = \sqrt{(x_i - u_j)^2}$$

Penjelasan:

$x_i$  : Data kriteria

$u_j$  : Centroid pada *cluster* ke  $j$

### 3. Pengelompokan data

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat *cluster* terdekat. Adapun cara pengelompokan data tersebut adalah:

- a Pilih nilai jarak tiap pusat *cluster* dengan data.
- b Cari nilai jarak terkecil.
- c Kelompokkan data dengan pusat *cluster* yang memiliki jarak terkecil.

### 4. Penentuan pusat *cluster* baru

Untuk mendapatkan pusat *cluster* baru bisa dihitung dari rata-rata nilai anggota *cluster* dan pusat *cluster*. Pusat *cluster* yang baru digunakan untuk melakukan iterasi selanjutnya, jika hasil yang didapatkan belum konvergen. Proses iterasi akan berhenti jika telah memenuhi maksimum iterasi yang dimasukkan oleh user atau hasil yang dicapai sudah konvergen (pusat *cluster* baru sama dengan pusat *cluster* lama).

Algoritma penentuan pusat *cluster*:

- a Cari jumlah anggota tiap *cluster*
- b Hitung pusat baru dengan rumus

$$u_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in S_j} x_j \dots \dots (2)$$

Penjelasan :

$u_j(t+1)$  : Centroid baru pada iterasi ke 1

Nsj : Banyak data pada *cluster* sj

#### 2.2.2.2 Keuntungan dan Kekurangan K-Menas

Sebagai fungsi penambangan data, analisis *cluster* dapat digunakan sebagai alat yang berdiri sendiri untuk memperoleh wawasan ke dalam distribusi data. Adapun keuntungan lain dari metode ini (Han,dkk , 2012:445) antara lain :

1. *K-means* juga disebut segmentasi data di beberapa aplikasi karena pengelompokan mempartisi set data besar ke dalam grup sesuai dengan kemiripannya.
2. *K-means* bisa juga digunakan untuk deteksi *outlier* (nilai yang "jauh" dari mana pun *cluster*).
3. *K-means* mempartisi sekumpulan objek data (atau pengamatan) ke dalam himpunan bagian, sehingga banyak digunakan dalam banyak aplikasi seperti intelijen bisnis, pengenalan pola gambar, pencarian web, biologi, dan keamanan.

Selain itu, metode *clustering* memiliki beberapa kekurangan (Suyanto, 2017:262) antara lain:

1. *K-means* tidak dapat menjamin konvergen pada optimum global.
2. *K-means* sering terjebak pada optimum lokal, dimana centroid akhir yang dihasilkan tidak benar-benar menjadi pusat *cluster* yang sesungguhnya.
3. Keluaran dari *k-means* bergantung pada *centroid* awal yang ditentukan secara acak.

#### 2.2.3 Clustering

Menurut Suyanto (2017:260) *Clustering* adalah proses pengelompokan satu set objek data (into multiple groups) atau *cluster* sehingga benda-benda dalam suatu kelompok memiliki kesamaan yang tinggi, tetapi sangat berbeda dengan objek di kelompok lain.

Menurut Han,dkk (2012:445) *clustering* adalah proses mempartisi sekumpulan objek data (pengamatan) kedalam himpunan bagian yang dapat digunakan untuk mengatur hasil pencarian ke dalam kelompok dan menyajikan hasil dengan cara yang ringkas dan mudah diakses. *Clustering* banyak digunakan dalam berbagai bidang dengan beragam aplikasi yang sangat penting diantaranya riset pasar, sistem rekomendasi, sistem keamanan dan mesin pencarian.

#### **2.2.4 Rapidminer**

Menurut Baskoro,dkk (2013:8) *Rapidminer* merupakan perangkat lunak yang bersifat terbuka (open source). *Rapidminer* adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. *Rapidminer* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *Rapidminer* memiliki beberapa sifat sebagai berikut:

1. Ditulis dengan bahasa pemrograman *java* sehingga dapat dijalankan di berbagai sistem operasi.
2. Proses penemuan pengetahuan dimodelkan sebagai operator *trees*.
3. Representasi XML internal untuk memastikan format standar pertukaran data.
4. Bahasa *scripting* memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen.
5. Konsep *multi-layer* untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
6. Memiliki GUI, command line mode, dan *java* API yang dapat dipanggil dari program lain

## **BAB 3 METODOLOGI PENELITIAN**

### **3.1 Objek Penelitian**

Objek penelitian adalah suatu tempat yang akan diselidiki dalam kegiatan penelitian untuk menelusuri masalah dan menerapkan hasil dari penelitian tersebut. Objek penelitian ini berupa data siswa dari SMP S serta nilai-nilai dari beberapa mata pelajaran.

### **3.2 Waktu dan Tempat Penelitian**

Penelitian ini dilakukan mulai dari tanggal 30 Maret 2023 hingga selesainya pembuatan laporan ini. Sedangkan tempat penelitian dilakukan di SMP S dan pembuatan laporan dilakukan di kos salah satu penulis.

### **3.3 Jenis dan Sumber Data**

#### **3.3.1 Jenis Data**

Dalam penelitian ini penulis menggunakan jenis data kuantitatif yang dijadikan sebagai pendukung dalam penyelesaian tugas ini. Definisi dan Jenis dari data yang di ambil oleh penulis dari objek penelitian yaitu menggunakan Data kuantitatif. Data kuantitatif adalah data dari hasil penelitian yang bersifat terstruktur atau berpola sehingga ragam data yang diperoleh dari sumber riset lebih mudah dibaca oleh peneliti.

#### **3.3.2 Sumber Data**

Sumber data yang digunakan penulis dalam mendukung penelitian untuk menyelesaikan tugas akhir ini yaitu data primer dan data sekunder. Adapun pengertian dan contoh dari data yang diambil penulis pada objek penelitian adalah:

##### **1. Data Primer**

Data primer adalah jenis data yang dikumpulkan secara langsung dari sumber utamanya seperti melalui wawancara, survei, dataset statistik, dan sebagainya. Dalam pengumpulan data primer dalam

penelitian ini menggunakan metode dataset statistik yang dimana penggunaan dataset statistik ini merupakan penggunaan data yang sudah tersedia.

## 2. Data Sekunder

Data sekunder adalah data pendukung yang sumbernya didapat dari sumber yang telah ada atau peneliti sebagai tangan kedua. Data sekunder dapat diperoleh dari berbagai sumber seperti laporan, jurnal, dan lainnya. Data sekunder yang digunakan dalam penelitian ini adalah data yang berhubungan dengan data sebelumnya.



## BAB IV HASIL DAN PEMBAHASAN

### 4.1 Kriteria Data

Data yang digunakan dalam pengujian *cluster* menggunakan Rapid Miner ini berjumlah 106 data training

Tabel 2. Data *Clustering*

NIS	Nama	Fisika	Biologi	Matematika	Kimia	Bahasa Inggris
2012173	ABDUL JABAR	76	68	89	76	65
2012174	ADITIA ALFARISI	95	96	66	83	66
2012175	ADITYA PRAMUDITA	96	89	89	73	65
2012176	AINUN ZARIYAH	85	74	67	74	69
2012177	ALIEF NOVALYANSYAH	88	78	80	76	79
2012178	ALIEFVIA REZQA	70	79	82	70	84
2012179	ALMA SHARIKA SOFYANTI	75	87	93	95	71
2012180	ALVIEN ALFARIZI SANTOSO	72	75	91	86	76
2012181	AMARULAH ABDUL HAMID	72	90	97	87	75
2012182	ARI HERMANSYAH	95	88	74	65	94
2012183	AYU DWI CAHYANI	83	66	90	86	90
2012184	DIAN NATULHIKMAH	69	72	87	80	80
2012185	DIAN SUNARSIH	67	80	91	80	72
2012186	DINI WIDIYA OKTAPIANI	78	96	96	97	88
2012187	DWI SAKINA	65	69	87	79	82
2012188	EKA SUPARTINI	91	93	65	74	76
2012189	FAHRIYAN	85	71	95	86	87

	HARIS					
2012190	FITRI DIANSARI	88	71	96	76	78
2012191	FITRIANI	83	70	77	67	79
2012192	HANISYA RAHMI NOVIA SUMBAWATI	73	83	78	92	85
2012193	IIN PUTRI ANDANI	85	67	97	96	85
2012194	IKHSAN SAPUTRA	89	93	69	75	82
2012195	ILA MULYANI	84	73	81	80	92
2012196	INDA SARI	73	88	70	75	80
2012197	INDAS KHOFIFAH	65	90	77	93	78
2012198	JULIANA FEBRIANI	71	80	83	76	65
2012199	KHAERUL ANNAM	87	65	82	96	71
2012200	LALA APRILIA SALSABILA	65	93	85	96	93
2012201	LELY RAHMAWATI	96	90	94	95	75
2012202	M REZA JULIANTO	67	72	69	72	87
2012203	MINI SEPTIKA	89	85	95	78	90
2012204	MOHAMMED SESAY	97	93	90	88	80
2012205	MUHAMAD IKSAN	81	73	81	82	96
2012206	MUHAMMAD IQBAL FUSTHAHULLA H	86	65	71	83	94
2012207	MUHAMMAD RAMDANI	87	65	73	79	66
2012208	MUHAMMAD WAHYU MIFTAH JUANDI	96	71	86	71	68
2012209	NADIA	96	80	68	69	90
2012210	NADIA SAFITRI	67	73	91	91	84
2012211	NANDA LARA SAFITRI	79	80	81	92	85

2012212	NAWAB WINARDI	81	65	65	68	85
2012213	NENGSIH LESTARI	68	71	91	95	90
2012214	NUNUNG AFRIANI	93	82	91	70	65
2012215	NURILA	93	77	96	95	77
2012216	NURWAHIDA AFLIA	92	94	69	90	97
2012217	PRASTIA JULIA UTAMI	90	91	77	83	93
2012218	RAMONA DWI PUTRI	96	67	96	76	67
2012219	RATU WIDIA	90	89	96	93	67
2012220	RENI OKTAPIANI	88	73	90	97	75
2012221	RIRIN TRIANI	69	79	65	65	96
2012222	SARTIKA OKTARIDA	78	81	85	95	65
2012223	SATRIA HERNAWAN	74	91	79	70	65
2012224	SATRIA MOKTAR	89	75	71	82	93
2012225	SEPTIANA EKA PRATIWI/putri	73	70	79	90	71
2012226	SITI AMNAL ASKIYA	80	82	95	66	97
2012227	SITI RAHMAWATI	87	76	75	93	85
2012228	SONI APRIAWAN	66	75	76	79	69
2012229	SRI RAHAYU	94	79	74	89	97
2012230	SRI WULANDARI	79	88	77	78	87
2012231	SURYA YUDIANTO	71	88	66	91	90
2012232	TIARA FITRI RAMDANI	97	76	76	78	93
2012233	TRIANA RIZTA MUHARROZI	86	88	80	79	76
2012234	VIDYANA WULANDARI	91	72	84	89	74
2012235	VINA FEBRIANA	85	80	92	74	81

2012236	WULANDARI AHDIAT	95	82	94	95	80
2012237	YOPIN INTAN SEPTIANTI	84	86	67	73	92
2012238	YULIA PUTRI ANGGINI	88	80	93	94	74
2012239	YUYUN FEBRIANTY	84	71	65	91	75
2012240	AINI FEBRIANTI	78	84	90	74	97
2012241	ALICHA SYEHAN	84	96	72	75	65
2012242	ALVINA SUMANTY	88	94	83	89	70
2012243	ANA SAPITRI	72	80	76	91	65
2012244	ANDI RAHMAT HIDAYAT	78	82	75	72	90
2012245	ANIS SELPIANI	68	88	81	77	88
2012246	ANJAS MARA TRI SYAPUTRA	94	72	83	73	88
2012247	ANUGRAH ADE CANTARY	81	85	86	89	91
2012248	ARDIANSYAH	70	92	76	87	74
2012249	ARMELIA PUSPITASARI	93	65	85	66	92
2012250	ASRI ARSITA	83	71	96	93	78
2012251	AZRIL NURAQSYA PRANA	83	68	94	85	76
2012252	CANRIKA SYAHPUTRI	80	84	84	97	93
2012253	DEBBY TRI CAHYANI	72	84	84	69	70
2012254	DEWI LESTARI	83	82	77	80	70
2012255	DISA SELPIAH	84	97	96	92	96
2012256	DWI FEBRI AMANDARI	83	73	67	85	78
2012257	DWI PUTRI	85	67	80	69	90
2012258	DWI YANA RAVIKA	88	72	82	87	86
2012259	EFA ROSIFA	91	86	79	93	85
2012260	EKA RAHMAT TULYANTI	85	70	86	75	79

2012261	EKA SAPUTRIANI	77	77	83	90	71
2012262	ELI ERMAWATI	83	87	68	72	94
2012263	ERIK SOLIKIN	88	65	87	92	67
2012264	ERNA WAHYUNI	87	86	83	72	66
2012265	FADILLAH EKA MEILANY	79	85	88	73	69
2012266	FARHAN	92	75	92	91	71
2012267	FAUZI HAMDANI	90	82	76	78	97
2012268	FITRI DAMAYANTI	94	81	95	65	91
2012269	GUSTINA REJAUNA	84	68	85	78	68
2012270	HADID SURYADIN	96	95	94	66	73
2012271	HAMDANI	90	65	80	69	80
2012272	HERNI NUR AZIZA	75	85	79	89	95
2012273	HIDAYATULLA H	71	93	71	66	82
2012274	IDRIS FEBRIANSYAH	89	82	69	89	68
2012275	IKHSAN ADITYA	93	67	76	95	73
2012276	ILHAM	71	69	82	80	77
2012277	ILHAM ANUGRAH PUTRA	68	95	91	93	87
2012278	NADYA AULIA ISNAINI	94	75	89	96	96

#### 4.2 Proses *Clustering*

Pada tahap ini akan dilakukan proses utama yaitu pengelompokan data dalam mengevaluasi dan menganalisis tingkat prestasi belajar murid yang rendah, dan tinggi pada SMP S. Berikut ini merupakan penerapan algoritma *k-means* dengan asumsi bahwa parameter *input* adalah jumlah dataset sebanyak  $n$  data dan jumlah inisialisasi *centroid*  $k = 3$  sesuai dengan penelitian. Data yang diambil

untuk penelitian berjumlah 106 untuk dijadikan contoh penerapan algoritma *k-means*. Percobaan dilakukan dengan menggunakan parameter-parameter berikut :

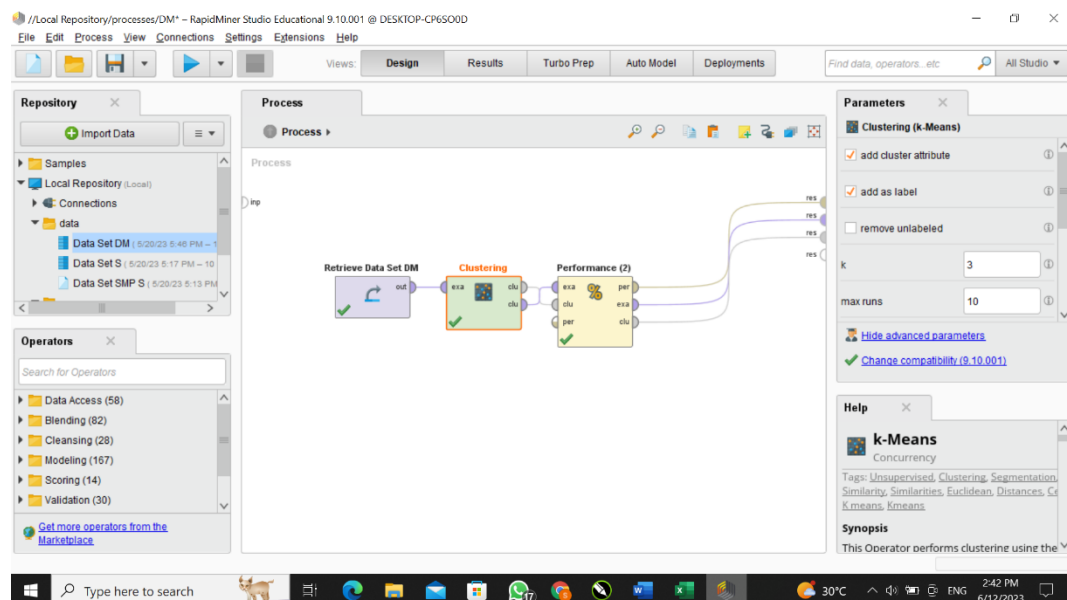
Jumlah *cluster* : 3

Jumlah data : 106

Jumlah atribut : 6

### 4.3 Pengujian *Rapidminer*

Pada penelitian ini penulis menggunakan *tool rapidminer* sebagai alat dataset. Adapun tahapan pengujian yang dilakukan yaitu sebagai berikut :



Gambar 3. *Design Process*

Pada tahapan ini dilakukan 1 proses dikarenakan data set yang digunakan sudah bersih yaitu *clustering*. Tahapan ini dilakukan operasi *clustering* sebagai algoritma yang digunakan pada penelitian ini.

Row No.	id	label	Nama	Fisika	Biologi	Matematika	Kimia	Bahasa Indonesia
1	1	cluster_1	ABDUL JABAR	79	73	81	78	82
2	2	cluster_0	ADITIA ALFA...	87	94	95	77	78
3	3	cluster_1	ADITYA PRA...	87	96	69	86	94
4	4	cluster_1	AINUN ZARIY...	95	82	79	85	88
5	5	cluster_2	ALJEF NOVAL...	70	80	82	89	86
6	6	cluster_1	ALJEFVIA RE...	77	92	65	74	93
7	7	cluster_0	ALMA SHARI...	73	74	95	76	70
8	8	cluster_2	ALVIEN ALFA...	69	95	74	78	80
9	9	cluster_0	AMARULAH A...	71	89	94	73	91
10	10	cluster_1	ARI HERMAN...	82	81	67	68	84
11	11	cluster_2	AYU DWI CA...	70	91	75	97	69
12	12	cluster_2	DIAN NATUL...	88	93	81	96	70
13	13	cluster_2	DIAN SUNAR...	71	67	90	81	73

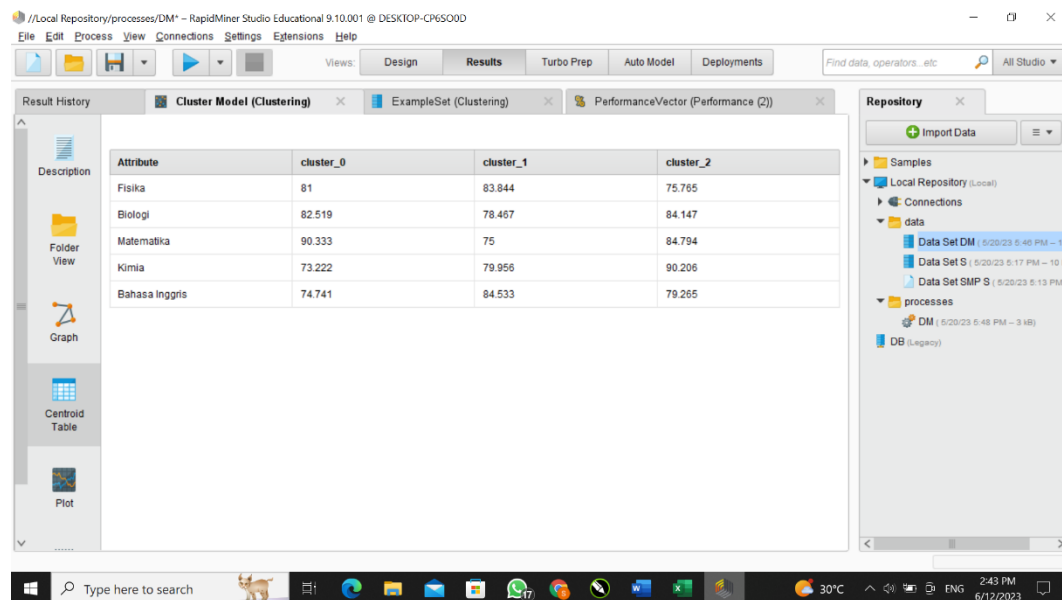
Gambar 4. *Example Set Result*

Pada tahapan ini ditampilkan hasil dari klasterisasi data. Label *cluster* terbagi menjadi 3 kelompok yaitu *cluster 0*, *cluster 1*, *cluster 2*. Pembagian ini berdasarkan abjad absen siswa.

Cluster	Number of Items
Cluster 0	27 items
Cluster 1	45 items
Cluster 2	34 items
<b>Total</b>	<b>106 items</b>

Gambar 5. *Cluster Model*

Pada tahapan ini ditampilkan hasil pembagian data terhadap tiap *cluster*. *cluster* 0 memiliki 27 anggota, *cluster* 1 memiliki 45 anggota, *cluster* 2 memiliki 34 anggota dari total 106 dataset yang di uji.

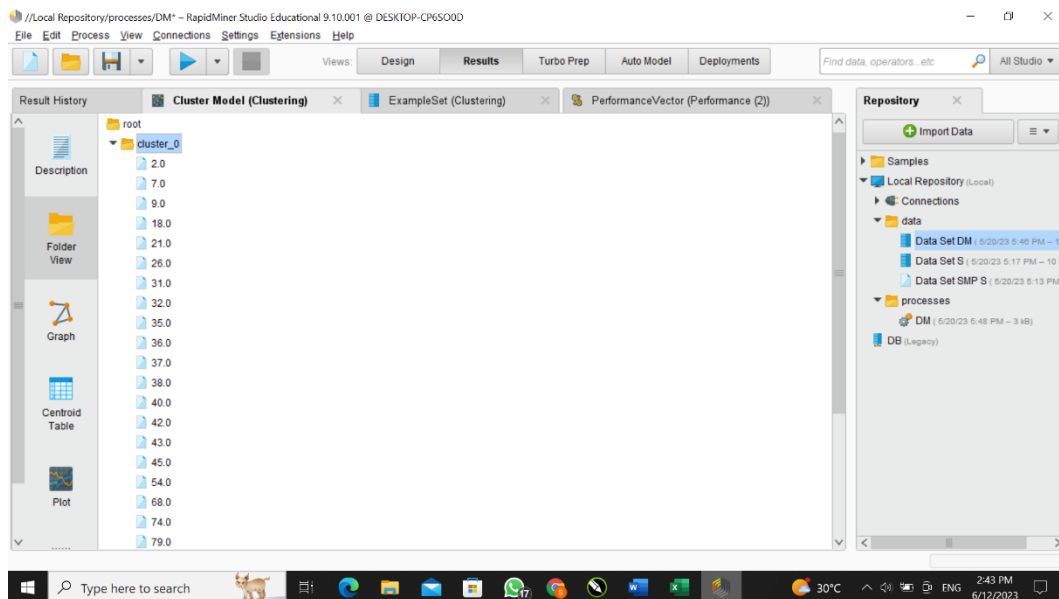


Attribute	cluster_0	cluster_1	cluster_2
Fisika	81	83.844	75.765
Biologi	82.519	78.467	84.147
Matematika	90.333	75	84.794
Kimia	73.222	79.956	90.206
Bahasa Inggris	74.741	84.533	79.265

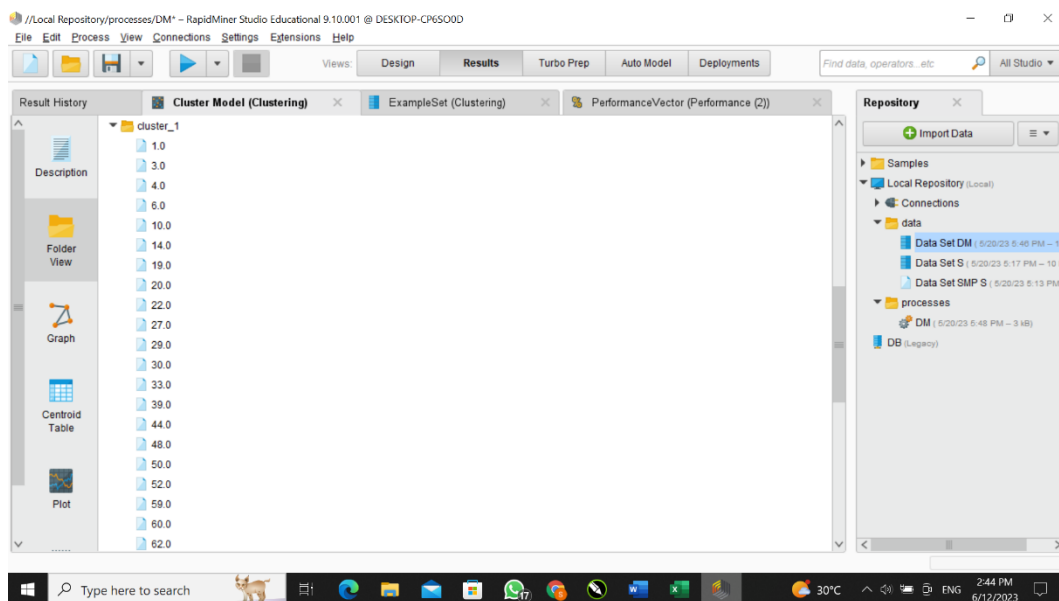
Gambar 6. *Centroid Table*

Pada tahapan ini ditampilkan nilai titik pusat pada tiap *cluster*. Nilai tersebut menjadikan acuan perhitungan pada tiap-tiap dataset dengan cara mengukur nilai yang ada dengan masing-masing titik pusat *cluster*.

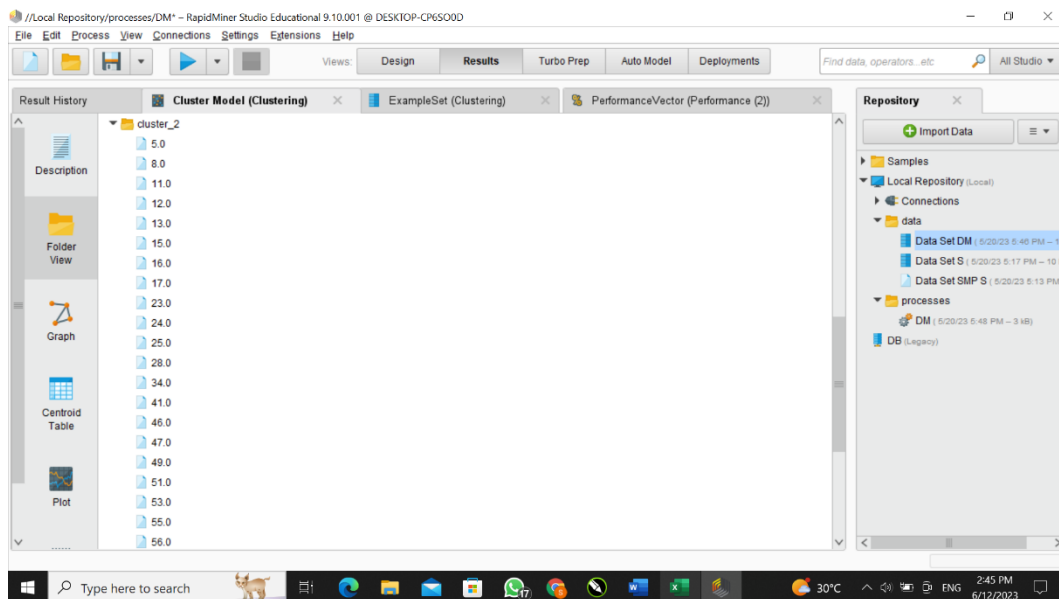




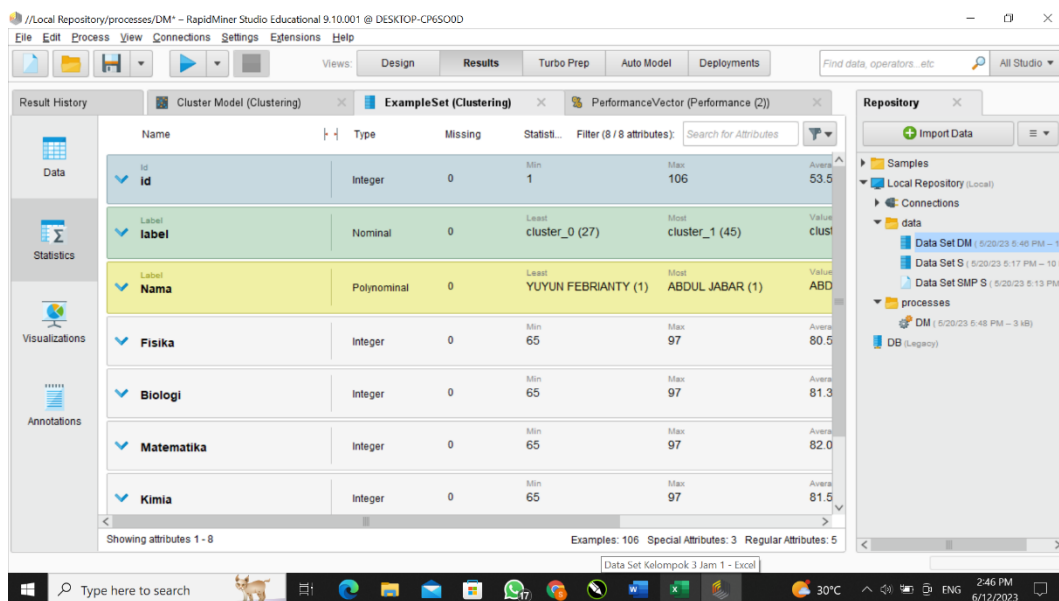
Gambar 7. Anggota *Cluster 0*



Gambar 8. Anggota *Cluster 1*



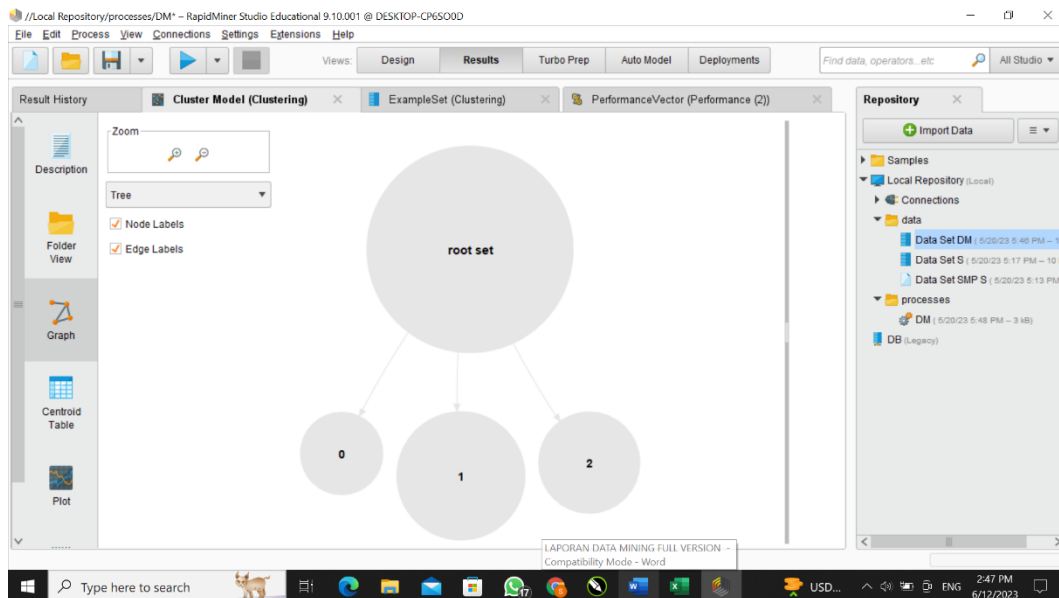
Gambar 9. Anggota *Cluster 2*



Gambar 10. *Example Set Statistic*

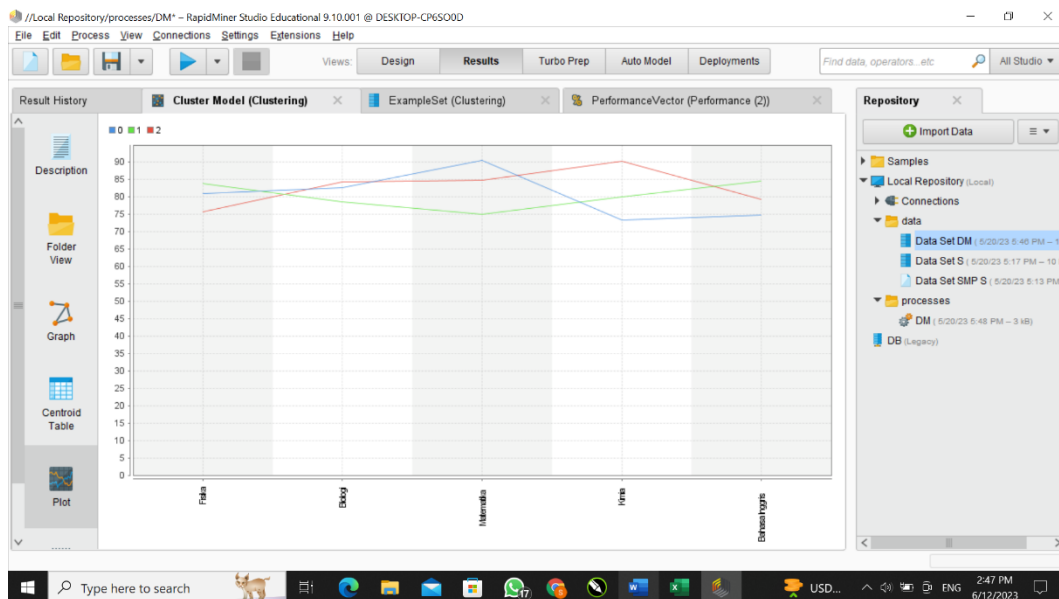
Pada tahapan ini ditampilkan hasil statistik dari data yang sudah di uji. Pada tabel Fisika, Biologi, Matematika, Kimia dan Bahasa Inggris terdapat 3 atribut yaitu *min* sebagai nilai terendah pada tabel dataset, *max* sebagai nilai

tertinggi pada tabel dataset dan *average* sebagai nilai rata-rata dari penjumlahan tabel dataset tersebut.



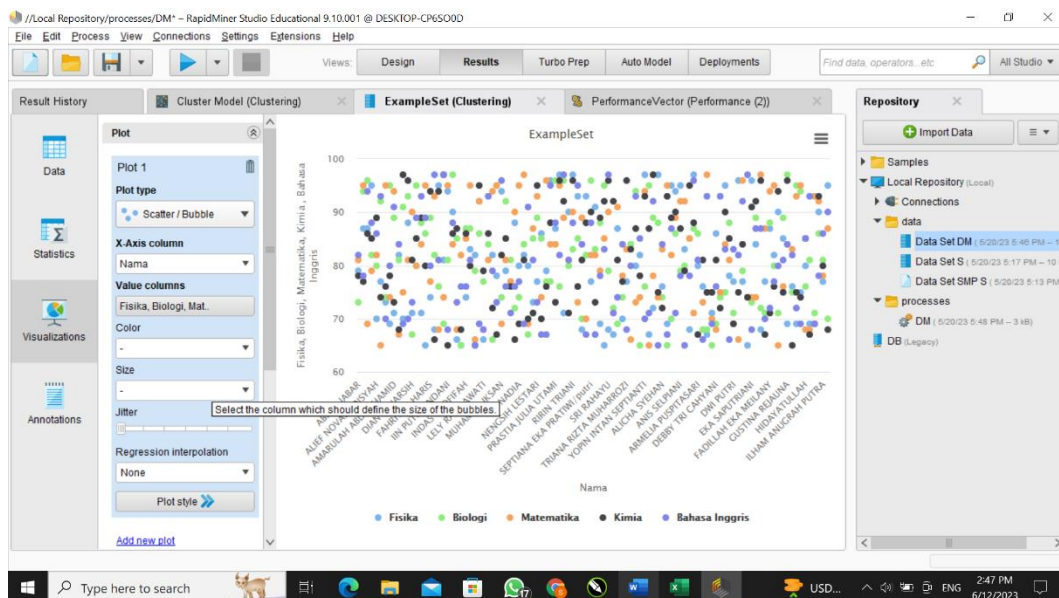
Gambar 11. *Graph Result*

Pada tahapan ini ditampilkan hasil pembagian dari lima kelompok berupa lingkaran. Ukuran tiap lingkaran mendeskripsikan jumlah banyaknya anggota tiap *cluster*.



Gambar 12. *Plot Result*

Pada tahapan ini ditampilkan hasil plot dari hasil pengujian. Pada bagian sebelah kiri menunjukkan angka nilai siswa dan bagian sebelah kanan ditampilkan garis sebagai gambaran rata-ran nilai siswa.

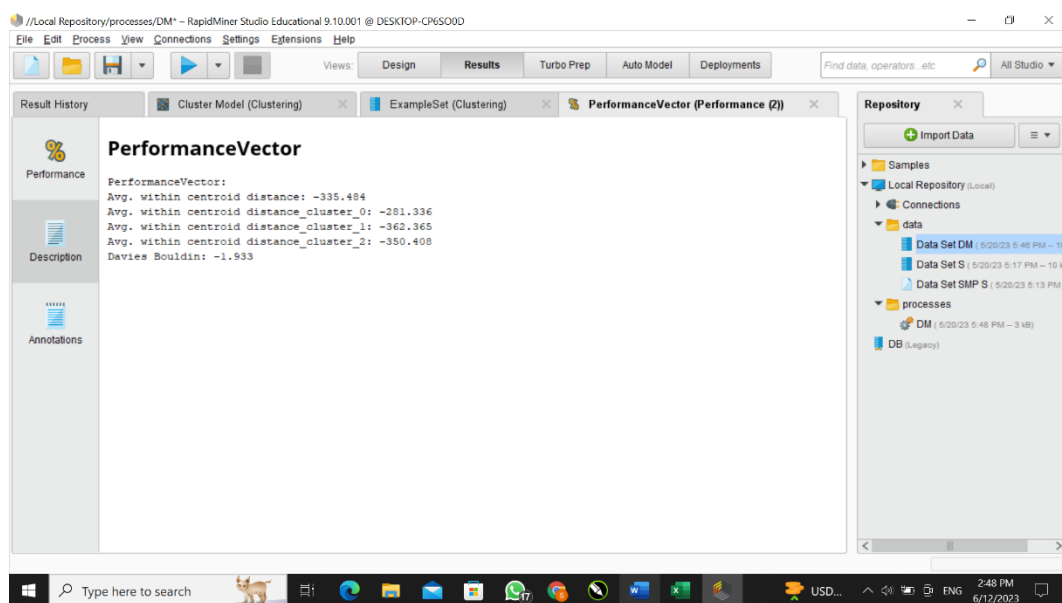


Gambar 13. *Chart Clustering*

Pada tahapan ini ditampilkan hasil pengelompokan data dalam bentuk grafik titik dengan warna. Warna biru muda mengartikan fisika, warna hijau mengartikan biologi, warna orange mengartikan matematika, warna hitam mengartikan kimia dan warna ungu mengartikan bahasa inggris.

#### 4.4 Pembahasan Hasil *Clustering* K-Means

Setelah dilakukan pengujian dengan *tool rapidminer*, maka didapatkan kesimpulan sebagai berikut :



Gambar 14. Hasil Rata-Rata *Clustering*

Jadi berdasarkan klastering yang telah dilakukan, diperoleh hasil rata-rata centroid dari masing-masing *cluster* adalah -335.484 untuk centroid pusat sementara untuk rata-rata *cluster* 0 adalah -281.336; *cluster* 1 adalah -362.356; *cluster* 2 adalah -350.408.

## **BAB V PENUTUP**

### **5.1 Kesimpulan**

Berdasarkan uraian yang telah ada pada bab – bab sebelumnya maka dapat ditarik kesimpulan sebagai berikut :

1. Dari hasil penelitian diketahui bahwa membentuk data menjadi 3 *cluster* sejalan dengan beberapa penelitian serupa karena dapat menunjukkan performa lebih baik dibandingkan dengan pengujian menggunakan *cluster* yang lain.
2. Untuk membuktikan bahwa data yang dihasilkan ada tahapan awal bernilai benar, maka perlu dilakukan perhitungan kembali berdasarkan anggota *cluster* yang telah terkumpul untuk menentukan *centroid* kedua sekaligus mengetahui perbandingan antara centroid awal dan *centroid* kedua.
3. Dari penelitian tersebut dapat dikatakan bahwa hasil setiap cluster diurutkan berdasarkan siswa yang memiliki nilai paling tinggi yaitu *cluster* 0.
4. Metode Clustering Algoritma K-Means dapat diterapkan pada pengelompokan hasil evaluasi Siswa SMP S, sehingga metode ini sangat membantu pihak akademik dalam menentukan siswa berprestasi.
5. Hasil pengelompokan data akademik siswa dapat berfungsi sebagai acuan bagi perencanaan akademik untuk memantau dan mengevaluasi perkembangan kinerja akademik setiap siswa.

### **5.2 Saran**

Mengingat masih banyaknya hal-hal yang belum dapat diimplementasikan dari penelitian ini, maka penulis mempertimbangkan sebuah saran yaitu, saat pemilihan data set usahakan untuk data atribut lebih dari 100 an agar mudah menemukan algoritma clusteringnya.

## **DAFTAR PUSTAKA**

- Aziz, F. N., Setiawan, B. D., & Arwani, I. (2018). Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa. 2243.
- Hutagalung, J., Syahputra, Y. H., & Tanjung, Z. P. (2022). Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering. 606.
- Putra, J. M., & Yuniarti, D. A. (2020). Analisis Hasil Belajar Mahasiswa dengan Clustering menggunakan Metode K-Means. 49.
- Sembiring, S. N., Kusnasari, S., & Winata, H. (2022). Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means. 31.
- Sulistiyawati, A., & Supriyanto, E. (2018). Implementasi Algoritma K-Means Clustering dalam Penentuan Siswa Kelas Unggulan. 25.