

## Klasterisasi Penerima Bantuan Fasilitas Sekolah di MI Datuk Singaraja Menggunakan Metode K-Means

Akhmad Khanif Zyen<sup>1\*</sup>, Gentur Wahyu Nyipto Wibowo<sup>2</sup>, Nur Aeni Widiastuti<sup>3</sup>, Buang Budi Wahono<sup>4</sup>

Fakultas Sains Dan Teknologi, Universitas Islam Nadhatul Ulama Jepara

e-mail: [khanif.zyen@unisnu.ac.id](mailto:khanif.zyen@unisnu.ac.id), [gentur23@yahoo.com](mailto:gentur23@yahoo.com), [nuraeniwidiastuti@unisnu.ac.id](mailto:nuraeniwidiastuti@unisnu.ac.id),  
[budihono78@gmail.com](mailto:budihono78@gmail.com)

### Abstrak

Pendidikan adalah kunci pembentukan sumber daya manusia yang berkualitas. Namun pada kenyataannya masih terdapat berbagai macam persoalan yang ada di dunia pendidikan di negeri ini. Adanya bantuan fasilitas dari sekolah adalah akibat dari persoalan tersebut, Namun timbul berbagai pertanyaan yang muncul terkait pemberian bantuan fasilitas sekolah ini, misal apakah sudah sesuai sasaran dan kriteria dalam pemberian bantuan tersebut. pada penelitian ini menggunakan metode *Clustering* Algoritma K-Means. Data yang digunakan sebanyak 516 data siswa tahun 2021 sampai 2022 dengan 4 atribut, yaitu Raport, Status Siswa Yatim/Tidak, Pendapatan Orangtua, Kedisiplinan. Hasil yang diperoleh yaitu 3 *Cluster* dimana *Cluster* 1 terdapat 180 siswa dengan kebanyakan siswa berprestasi dengan nilai raport sangat baik dengan keadaan ekonomi berkecukupan. Pada *Cluster* 2 terdapat 13 siswa kebanyakan siswa dengan keadaan orangtua yang tidak lengkap. Pada *Cluster* 3 terdapat 323 siswa kebanyakan siswa dengan nilai kedisiplinan yang sangat layak dengan keadaan keluarga dan ekonomi yang memang seharusnya diprioritaskan menerima bantuan. Di harapkan dengan terbentuknya klasterisasi calon penerima bantuan fasilitas sekolah MI Datuk Singaraja Kerso, maka pihak sekolah dapat menggunakan informasi tersebut sebagai referensi untuk membentuk kelompok prioritas penerima bantuan yang lebih baik agar sesuai sasaran dan kriteria.

Kata Kunci: *Data Mining, Clustering, K-Means, Bantuan*, MI Datuk Singaraja Kerso

### Abstract

*Education is the key to the formation of quality human resources. But in reality there are still various kinds of problems that exist in the world of education in this country. The existence of facility assistance from schools is a result of these problems. However, various questions arise regarding the provision of this school facility assistance, for example whether it is in accordance with the targets and criteria in providing the assistance. in this study using the K-Means Clustering Algorithm method. The data used are 516 student data from 2021 to 2022 with 4 attributes, namely report cards, Orphan/No Student Status, Parental Income, Discipline. The results obtained are 3 Clusters where in Cluster 1 there are 180 students with most students excel with very good report cards with affluent economic conditions. In Cluster 2 there are 13 students, mostly students with incomplete parents. In Cluster 3 there are 323 students, mostly students with very decent discipline values with family and economic conditions which should be prioritized to receive assistance. It is hoped that with the formation of a clustering of candidates for the assistance of MI Datuk Singaraja Kerso school facilities, the school can use this information as a reference to form a better priority group of beneficiaries to match the targets and criteria.*

Keywords: *Data Mining, Clustering, K-Means, Assistance, Islamic Elementary School Datuk Singaraja Kerso.*

## 1. PENDAHULUAN

Pendidikan merupakan bagian integral dan tak terpisahkan dari perkembangan, kemajuan dan pembangunan suatu bangsa. Melalui pendidikan yang memadai serta pemerataan kualitas pendidikan akan membuat perubahan besar pada kemajuan suatu bangsa.[1]. Pendidikan dasar adalah pendidikan yang memberikan pengetahuan dan keterampilan, menumbuhkan sikap dasar yang diperlukan dalam masyarakat, serta mempersiapkan peserta didik untuk mengikuti pendidikan menengah, untuk memberikan bekal dasar yang diperlukan untuk hidup dalam masyarakat berupa pengembangan sikap, pengetahuan dan keterampilan dasar.[2]Beasiswa adalah pemberian bantuan berupa keuangan yang diberikan kepada perorangan, pelajar atau mahasiswa yang digunakan demi keberlangsungan pendidikan yang ditempuh. Menurut Murniasih (2009) beasiswa diartikan sebagai bentuk penghargaan yang diberikan kepada individu agar dapat melanjutkan pendidikan ke jenjang yang lebih tinggi. Umumnya pemberian beasiswa diberikan berdasarkan besarnya gaji orang tua atau kepada para peserta didik yang memiliki nilai akademiknya baik di sekolah [3]

MI Datuk Singaraja adalah sebuah sekolah pendidikan dasar yang berlokasi di Jln Bugel-Jepara Km. 2 Desa Kerso Kecamatan Kedung Kabupaten Jepara, kepala sekolah MI Datuk Singaraja pada tahun 2021 adalah bapak Amin, S.Pd.I, pada tahun pelajaran 2021 sampai dengan 2022 MI Datuk Singaraja memiliki jumlah guru sebanyak 17 guru pengajar dan peserta didik sebanyak 516 meliputi kelas 1 sampai dengan kelas 6. MI Datuk Singaraja Dikelolah oleh Yayasan Datuk Singaraja yang Berdiri pada tanggal 1 Muharrom 1407 H dan bertepatan dengan tanggal 06 September 1986 M. yayasan datuk singaraja mengelola RA –MI-SMP-SMK, Majelis Ta'lim dan Pondok Pesantren. Untuk masuk menjadi warga Datuk Singaraja calon peserta didik hanya perlu di daftarkan oleh orang tua, wali atau dirinya sendiri dan mengikuti prosedur selanjutnya.

Bantuan fasilitas sekolah adalah program dari pihak sekolah untuk membantu meringankan beban siswa dengan memberikan beberapa fasilitas yang di berikan dari pihak sekolah kepada siswa siswi yang membutuhkan, sebanyak 516 siswa akan diseleksi. Bantuan fasilitas Sekolah ini sangat membantu siswa memenuhi kebutuhan di dalam kegiatan belajar maupun dalam melengkapi siswa itu sendiri ini sangat mendukung dalam program pemerintahan yang harus belajar duabelas tahun. Untuk mengetahui pola penentuan siswa penerima bantuan dapat dilakukan dengan cara memanfaatkan data siswa dan data penerima bantuan fasilitas sekolah. Karena jumlah peserta didik di MI Datuk Singaraja cukup banyak yaitu 516 siswa pada TA 2021 sampai dengan 2022, sehingga dibutuhkan *Clustering* dalam pengelompokan data untuk menentukan calon penerima bantuan bagi peserta didik dengan menggunakan metode *K-Means*. Untuk mengurangi ketidak tepat an sasaran dalam pemberian bantuan tersebut dan untuk menentukan peserta didik mana yang dapat menerima bantuan sesuai dengan kriteria yang telah di tentukan.

Masalah yang sering muncul karena ketidak tepatan penerima bantuan ini adalah masalah sosial yaitu ketika seorang anak yang seharusnya mendapatkan bantuan tetapi tidak mendapatkan secara tepat mengakibatkan kecemburuan orang tua terhadap anak yang mendapatkan nya. Sehingga menimbulkan pembicaraan pembicaraan yang tidak diinginkan dari warga maupun wali siswa terhadap sekolah maupun dewan guru yang bersangkutan.

### Rumusan Masalah

Perlu adanya metode yang dapat mengklasterisasi calon siswa penerima bantuan untuk mengurangi ketidaktepatan sasaran dalam pemberian bantuan fasilitas sekolah di MI Datuk Singaraja Kerso Jepara.

## 2. TINJAUAN PUSTAKA

Pada landasan teori ini berisikan penjelasan tentang teori-teori yang dijadikan sebagai dasar dalam melakukan penelitian ini.

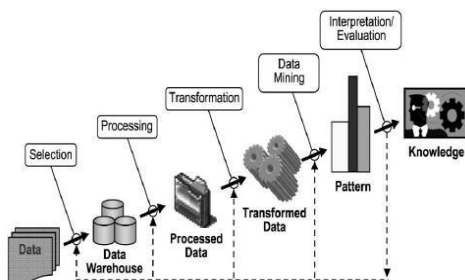
## 2.1. Data Mining

Data mining merupakan suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari suatu data. Hasil dari pengolahan data dengan menggunakan metode data mining dapat digunakan untuk pengambilan keputusan di masa depan. Data mining kegiatan pengolahan data dengan skala yang besar, sehingga data mining memiliki peranan yang penting dalam berbagai bidang seperti bidang industri, keuangan, cuaca, ilmu dan teknologi.[4]

Sebenarnya data mining merupakan suatu langkah dalam knowledge discovery in databases (KDD). Knowledge discovery sebagai suatu proses terdiri atas data cleaning, data integration, data selection, data transformation, pattern evaluation knowledge presentation.[5]

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakrit dari berbagai database besar.[6]

Istilah *data mining* dan *knowledge discovery in database* (KDD) sering digunakan untuk menjelaskan proses



penggalan informasi yang tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, berkaitan antara satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*[7].

Proses KDD secara garis besar dapat dijelaskan sebagai berikut:

**Gambar 2. 1 KDD (Knowledge Discovery in Database)**

## 2.2 Tahapan Data Mining

variabel lain di masa depan. Salah satu teknik

*Data Mining* yang merupakan serangkaian proses, terbagi dalam beberapa tahap yang harus dilalui

### 1. Data Selection

Memilih kata dari sekumpulan data operasional perlu dilakukan sebelum tahapan penggalian informasi dalam KDD dimulai.

### 2. Preprocessing

Sebelum proses *data mining* dijalankan, diperlukan proses *cleaning* yang bertujuan untuk membersihkan data yang bersifat duplikasi, data yang tidak konsisten, ataupun data yang salah cetak (tipografi).

### 3. Transformation

Merupakan kegiatan pengkodean pada data yang telah dipilih, sehingga data tersebut dapat digunakan untuk proses *data mining*.

4. *Data mining* Adalah proses mencari pola atau informasi yang menarik dalam data yang telah dipilih dan diproses pada tahap sebelumnya dengan teknik atau metode tertentu.

4. *Interpretation/Evaluation* Pola informasi yang telah didapat di tahap *data mining* kemudian akan ditampilkan kedalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini juga merupakan tahapan untuk memeriksa apakah pola atau informasi yang didapat bertentangan dengan fakta ataupun hipotesa yang ada sebelumnya atau tidak[9].

## 2.3 model data mining

*Data mining* secara garis besar dikelompokkan menjadi 2 kategori utama, yaitu [10]

1. *Descriptive mining* merupakan sebuah proses menemukan karakteristik penting dari data didalam sebuah *database*. Teknik *data mining* yang termasuk *descriptive mining* ialah *asosiasi*, *Clustering*, dan *sequential mining*.
2. *Predictive* merupakan proses menemukan pola dari suatu data menggunakan beberapa

dalam kategori ini adalah *klasifikasi*.

Menurut Larose (2005) berdasarkan tugas yang dapat dilakukan, *Data mining* dibagi menjadi 6 kelompok[8]

**Data Mining**

**2.3.1 Deskripsi**

Deskripsi dilakukan dengan tujuan untuk menggambarkan pola serta kecenderungan yang terdapat didalam data. Deskripsi dari pola serta kecenderungan ini sering memberikan kemungkinan sebuah penjelasan untuk suatu pola atau kecenderungan.

**2.3.2 Klasifikasi**

Dalam klasifikasi, ada variabel yang menentukan tujuan atau hasil. Misalnya, pengelompokan penghasilan yang dapat dibedakan dalam tiga kategori, yaitu penghasilan tinggi, penghasilan sedang, dan penghasilan rendah.

**2.3.3 Estimasi**

Dalam klasifikasi, ada variabel yang menentukan tujuan atau hasil. Misalnya, pengelompokan penghasilan yang dapat dibedakan dalam tiga kategori, yaitu penghasilan tinggi, penghasilan sedang, dan penghasilan rendah.

**2.3.4 Prediksi**

Prediksi hampir sama dengan klasifikasi dan estimasi, perbedaannya terletak pada nilai dari hasil prediksi akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi juga dapat digunakan untuk prediksi.

**2.3.5 Clustering**

*Clustering* merupakan proses mengumpulkan, mengamati, atau mengamati data untuk membentuk suatu kelas objek yang mempunyai satu atau lebih persamaan.

**2.3.6 Asosiasi**

Tugas asosiasi dalam data mining ialah menemukan sebuah atribut yang muncul dalam satu waktu. Dalam dunia bisnis, lebih umum disebut sebagai analisis keranjang belanja.

**2.3 Clustering**

*Clustering* adalah proses pengelompokan benda serupa ke dalam kelompok yang berbeda, atau lebih tepatnya partisi dari sebuah data set kedalam subset, sehingga data dalam setiap subset memiliki arti yang bermanfaat[11]

*Clustering* juga bisa dikatakan suatu proses dimana mengelompokan dan

membagi pola data menjadi beberapa jumlah data set sehingga akan membentuk pola yang serupa dan dikelompokkan pada *Cluster* yang sama dan memisahkan diri dengan membentuk[12]

*Clustering* sendiri disebut juga dengan unsupervised classification karena bertujuan untuk mempelajari dan memperhatikan. *Cluster analysis* adalah proses membagi satu set objek data menjadi subset. Karena setiap subset adalah sebuah *Cluster*, maka objek-objek dalam sebuah *Cluster* sama dan berbeda dengan objek-objek di *Cluster* lain. Partisi dilakukan sesuai dengan metode *Clustering*, tidak secara manual. Sehingga *Clustering* ini sangat berguna dan dapat menemukan *Cluster* yang tidak diketahui pada data Anda [13]

Metode *Clustering* terbagi menjadi 2, yaitu *hierarchical Clustering* dan *partitional Clustering* [14]

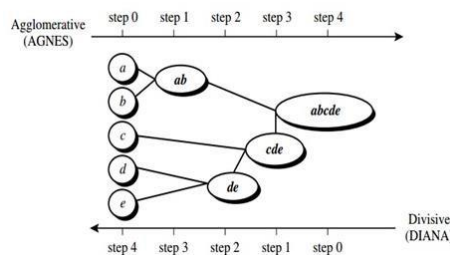
**1. hierarchical Clustering**

Pada *hierarchical Clustering* sekelompok data dikelompokkan melalui suatu bagan yang berupa hirarki, yang mana terdapat penggabungan dari 2 grup yang terdekat disetiap iterasinya

Terdapat 4 langkah dalam *hierarchical Clustering*, yaitu :

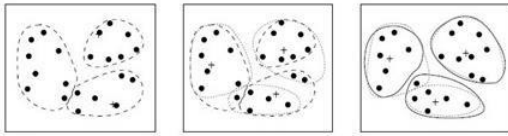
*Clustering*, yaitu :

- a. Identifikasi *item* yang memiliki jarak terdekat.
- b. *Item* tersebut digabungkan kedalam 1 *Cluster*.
- c. Penghitungan jarak antar *Cluster*.
- d. Ulangi langkah sampai semua terhubung



**Gambar 2.2 KKD**

**2. partitional Clustering**



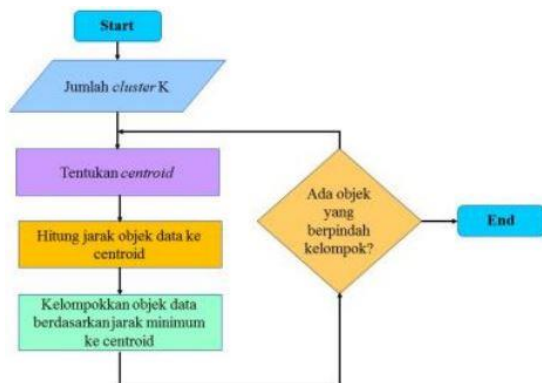
Gambar 2.3 Partitional Clustering

Pada *partitional Clustering* data dikelompokkan kedalam beberapa *Cluster*. Pada metode ini setiap *Cluster* (kelompok) memiliki titik pusat (*centroid*) serta secara umum metode ini bertujuan meminimumkan jarak (*dissimilarity*) dari seluruh data ke masing-masing *centroid Cluster*. [15]

**2.4 Algoritma K-Means**

Algoritma *K-Means Clustering* merupakan teknik *Cluster* berbasis jarak yang berusaha mempartisi data kedalam beberapa *Cluster*. Metode ini mempartisi data kedalam *Cluster* menurut karakteristik yang dimiliki setiap data, setiap data yang memiliki karakteristik Algorithm *K-Means Clustering* merupakan teknik *Cluster* berbasis jarak yang berusaha mempartisi data kedalam beberapa *Cluster*. Metode ini mempartisi data kedalam *Cluster* menurut karakteristik yang dimiliki setiap data, setiap data yang memiliki karakteristik sama dikelompokkan kedalam satu *Cluster* yang sama begitu juga dengan data yang mempunyai karakteristik berbeda dikelompokkan kedalam *Cluster* lain [16]

Berikut adalah diagram alir dari algoritma K-Means



Gambar 4 Flowcart k-means [17]

Menurut daniel dan eko, langkah-langkah algoritma *K-Means* adalah sebagai berikut. [18]:

1. Pilih secara acak nilai *k* tersebut sebagai pusat *Cluster*

2. Jarak antara data dan pusat *Cluster* dihitung dengan menggunakan rumus Euclidian Distance. Untuk mengitung jarak semua data ke setiap titik pusat *Cluster* dapat menggunakan teori Euclidean Distance dengan rumus sebagai berikut:

**Rumus Euclidean Distance**

$$D(i,j) = \sqrt{(x_i - x_{avg})^2 + (Y_i - x_{avg})^2 + (z_{ki} - z_{kj})^2}$$

Dimana :

- $D(i,j)$  = Jarak data ke *i* ke pusat *Cluster j*
- $X_i$  : data pertama (diambil dari atribut pertama)
- $X_{avg}$  : data pusat *Cluster l* centroid untuk atribut pertama
- $Y_i$  : data kedua diambil dari atribut kedua
- $Y_{avg}$  : titik pusat *Cluster/* centroid untuk atribut kedua
- $Z_i$  : data ketiga diambil dari atribut ketiga
- $Z_{avg}$  : titik pusat *Cluster /* centroid untuk atribut ketiga

**2.5 Pengujian Data**

Pada iterasi terakhir yang biasa dikenal dengan rasio, dilakukan pengujian untuk menentukan kriteria penilaian bagus atau tidaknya hasil dari perhitungan *Clustering* algoritma *K-means*. Metode yang digunakan adalah *Between-Class Variation*(BCV) dan *Within Class Variation*(WCV)

BCV merupakan rata-rata centroid, sedangkan WCV adalah nilai keseluruhan dari jarak minimum yang telah dijumlahkan

,Rumus perhitungan sebagai berikut:

**Rumus Between-Class Variation (BCV)**

$$BCV = \frac{1}{Nk} \sum_{i=1}^k d(m_i, m_i)$$

Dimana:

- $k$  = Jumlah *cluster*
- $m_i$  = Jumlah anggota dari *cluster* ke *i*
- $i$  = Nama yang mewakili *cluster* yang terbentuk
- $m_i$  = Jumlah anggota dari *cluster* ke *i*

**Rumus Within-Class Variation (WCV)**

$$WCV = \sum_{j=i}^m \sum_{p \in c_i} d(p, m_i)^2$$

Dimana:

- $p \in c_i$  = Jumlah semua data
- $k$  = Jumlah *cluster*
- $p$  = *Cluster* jarak terdekat
- $m_i$  = Jumlah anggota dari *cluster* ke *i*

Untuk menghitung rasio dapat menggunakan rumus persamaan :

$$\text{Rasio} = \frac{BCV}{WCV}$$

Apabila nilai rasio yang didapat semakin kecil maka semakin bagus tingkat hasil dari akurasi *Cluster*

Berikut ini adalah kriteria hasil ukuran rasio:

**Tabel 1. Kriteria Pengukuran Rasio**

Nilai Rasio	Kriteria
≤ 0,25	Sangat Layak
0,25-0,50	Layak
0,50-0,75	Kurang Layak
0,75-1,00	Tidak Layak

### 3. METODE PENELITIAN

#### 3.1 Identifikasi Masalah

Tahapan identifikasi masalah adalah tahap awal pada penelitian ini, pada tahapan ini dilakukan analisis masalah yang akan diselesaikan, studi kasus dilaksanakan di MI Datuk Singaraja, Metode yang digunakan dalam penelitian tugas akhir ini yaitu penerapan metode data mining *K-Means Clustering* pada data siswa siswi MI Datuk Singaraja Kerso Kedung Jepara tahun ajar 2021/2022 untuk dianalisis dan dikelompokkan sesuai dengan prioritas bantuan yang diterima, dengan parameter nilai rata-rata raport, penghasilan orang tua, yatim atau tidak dan kedisiplinan. Proses pengolahan data tersebut dilakukan dengan menggunakan aplikasi *Microsoft Excel* dan *RapidMiner*. Kemudian hasil dari pengolahan data akan dikelompokkan sesuai dengan tingkat kemiripan data.

#### 3.2 Pengumpulan Data

Di dalam penelitian ini peneliti menggunakan data siswa siswi MI Datuk singaraja kelas 1 sampai dengan 6 angkatan 2021 sampai dengan 2022 selama satu semester. Pengambilan data di kantor TU MI Datuk Singaraja dengan ijin kepala TU dan Kepala sekolah MI Datuk Singaraja, setelah mendapatkan ijin dari pihak MI Datuk Singaraja, data segera diambil dari pihak TU dengan cara mengkopi file berupa *file excel* file yang di dapat sejumlah 2 file yaitu file absensi siswa siswi angkatan 2021 sampai dengan 2022 dan file nilai raport selama 1 semester. Selanjutnya untuk melengkapi data dengan atribut yang sudah disesuaikan, peneliti meminta data kepada TU di MI Datuk Singaraja

##### 3.2.1 Observasi

Observasi adalah kegiatan pengamatan pada sebuah objek secara langsung dan detail untuk mendapatkan informasi terkait objek tersebut.

##### 3.2.2 Wawancara

Wawancara atau interview adalah kegiatan tanya jawab secara lisan untuk mendapatkan informasi. Bentuk informasi yang didapat dinyatakan dalam bentuk tulisan rekaman audio atau audio visual.

##### 3.2.3 Study Literatur

Study literatur atau pustaka adalah proses mencari referensi atau informasi yang berkaitan dengan kebutuhan penelitian, yaitu informasi tentang pemberian bantuan

#### 3.3 Pengolahan Data

Dalam penelitian ini dibutuhkan perangkat keras (Hardware) dan perangkat lunak (Software). Untuk lebih jelasnya dijelaskan berikut ini:

##### 3.3.1 Perangkat Keras

Perangkat keras (Hardware) yang digunakan dalam penelitian ini yaitu menggunakan 1 buah laptop yang digunakan selama penelitian. Spesifikasi laptop sebagai berikut: ACER ASPIRE 5, Intel CORE i3-6006U, CPU 2.0 GHz, Ram 4gb, SSD Storage 240gb, Windows 10 Education 64 bit.

##### 3.3.2 Perangkat Lunak

Peneliti menggunakan beberapa perangkat lunak atau *Software* untuk mendukung penelitian ini. Perangkat lunak yang digunakan saat melaksanakan penelitian ini, yaitu:

###### 1. Sistem Operasi

Peneliti menggunakan laptop dengan sistem operasi *Windows 10 Education 64-bit* dalam penelitian ini.

###### 2. Microsoft Word

Pada penelitian ini *Software Microsoft Word* digunakan untuk membuat laporan penelitian. Versi dari *Microsoft Word* yang digunakan adalah versi 2019.

###### 3. Microsoft Excel

Dalam penelitian ini *Software Microsoft Excel* digunakan untuk pengolahan data siswa MI Datuk Singaraja. Versi dari *Microsoft Excel* yang digunakan adalah versi 2019.

###### 4. RapidMiner 9.10.2

Dalam penelitian ini juga digunakan *Software RapidMiner 9.10.2* untuk melakukan proses perhitungan *Data Mining* metode *Clustering* menggunakan algoritma *K-means* terhadap data siswa MI Datuk Singaraja



**3.4 Tahapan Metode**

Peneliti menggunakan salah satu metode pada *Data Mining* yaitu metode *Clustering* algoritma *K-means*. Di dalam metode ini, terdapat beberapa tahapan yang harus dilakukan, tahapan-tahapan tersebut adalah:

**3.4.1 Pre Processing Data**

*Pre-processing* data adalah tahap dimana data yang sudah didapatkan, di pilah dan di pisahkan agar mendapatkan data yang dibutuhkan di proses selanjutnya. Tahapan proses dalam *pre-processing* data sebagai berikut:

1. Data Reduction  
Data Reduction adalah proses untuk mengurangi dimensi, atribut atau data yang tidak dibutuhkan dalam sebuah file data.
2. Data Cleaning  
Data Cleaning merupakan suatu proses analisa mengenai kualitas dari data dengan mengubah. Bisa juga pengelola mengoreksi ataupun menghapus . Data yang dibersihkan tersebut adalah data yang salah, rusak, tidak akurat, tidak lengkap dan salah format.
3. Data Transformation  
Proses transformasi adalah tahap untuk mengubah data atribut yang selain angka kedalam nilai angka agar dapat diolah menggunakan algoritma *K-Means Clustering*.

**Tabel 3. Tabel bobot status siswa**

NO	VARIABEL	FREKUENSI	BOBOT
1	Tidak	12	1
2	Yatim	182	2

**Tabel 4 Tabel Data Bobot data Pendapatan Orangtua**

NO	VARIABEL	FREKUENSI	BOBOT
1	Gaji >3000,000	57	1
2	Gaji >2000,0000	122	2
3	Gaji <1000,000	15	3

**Tabel 5 Tabel Data Bobot data Kedisiplinan**

NO	VARIABEL	FREKUENSI	BOBOT
1	Cukup	161	1
2	Baik	33	2

**Tabel 6 Tabel Bobot Nilai**

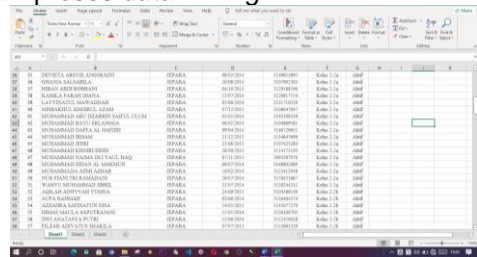
NO	NILAI	BOBOT
1	<60-74	1
2	75-84	2
3	85-100	3

4. Data Integration  
Data Integration adalah proses penggabungan data dari beberapa file sumber.

**4. HASIL DAN PEMBAHASAN**

**4.1 Pre Processing Data**

Tahapan atau proses untuk melakukan pengolahan data mentah yang diperoleh, supaya didapatkan data yang sesuai dengan kebutuhan analisa dan siap untuk masuk ke dalam proses *data Mining*

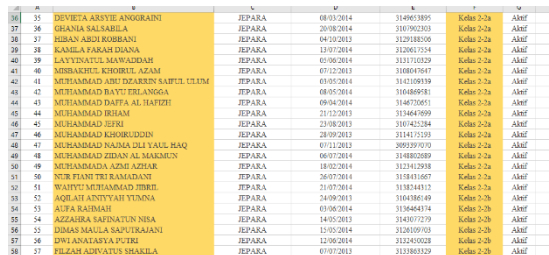


**Gambar 9 Data siswa siswi MI Datuk Singaraja**

Tahapan dari *pre-processing* data seperti berikut ini:

**4.1.1 Data Reduction**

Semua data dari database tidak serta merta diambil dan di proses, hanya data yang diperlukan untuk kebutuhan analisis saja yang digunakan. Dalam penelitian ini data yang digunakan dari data mentah adalah hanya data nama siswa dan kelas siswa saja. Maka dari itu yang diambil adalah kolom nama dan kelas siswa MI Datuk Singaraja tahun pelajaran 2021 sampai dengan 2022.



**Gambar 10 Data Reduction**

maka data yang tidak direduksi adalah Nama Siswa Dan Kelas. Selain itu data yang direduksi adalah Nomor, Tempat Lahir, Tanggal Lahir, NISN(Nomor Induk Siswa Nasional) dan Status.

**4.1.2 Data Cleaning**

Pada data siswa siswi MI Datuk Singaraja tahun pelajaran 2021 sampai dengan 2022 yang berjumlah 516 siswa dan sudah melalui *data reduction*, peneliti sudah melakukan pemeriksaan setiap kolom dan tidak menemukan data yang kosong dan inkonsisten ataupun data dengan penulisan yang salah.

**4.1.3 Data Transformasi**

Pada *file* data siswa ditambahkan kolom yang sesuai dengan atribut yang akan digunakan dalam penelitian ini

A	B	C	D	E	F	G
NO	Kelas	NAMA	RAPORT	Status Siswa Yatim/Tidak	Pendapatan Orangtua	Kedisiplinan
1	Kelas 6-0b	Muhammad Nur Fadilah Syahid				
2	Kelas 6-0a	Talia Zaidatur Zalfa				
3	Kelas 6-0b	Muhammad Fatih Roziqin				
4	Kelas 6-0b	Muhammad Rafi Adhya				
5	Kelas 3-3a	Lutfiah Anifah May Nita				
6	Kelas 3-0a	Risa Ramadhani Safri				
7	Kelas 3-0a	Wafiq Awadhya				
8	Kelas 2-2a	NUR IDA FITRI WAHYUNI				
9	Kelas 1-1	Muhammad Amri Adhar				
10	Kelas 4-0b	Vebby Noor Hidayah				
11	Kelas 3-0a	Intan Nur Ami				
12	Kelas 6-0b	Muhammad Kevin Oktama				
13	Kelas 1-1	Apreliani Putri				
14	Kelas 6	Muhammad Agus Salimman				
15	Kelas 2-2a	WI KHANA WAHDATUL MU				
16	Kelas 3-0a	Devietta Arsyie Angraeni				
17	Kelas 2-2a	AHMAD BAIHAQI				
18	Kelas 3-3a	Rista Oktaviani				
19	Kelas 4-0a	Anita Fairus Khalsa				
20	Kelas 1-0a	MUHAMMAD KHORUDDIN				

**Gambar 11 Penyesuaian Atribut**

Setelah file sudah siap dan telah ditambahkan beberapa atribut yang digunakan dalam penelitian ini, yaitu Nilai Raport, Status Siswa Yatim/Tidak, Pendapatan Orangtua. Hal ini dilakukan sesuai dengan kebutuhan analisa. selanjutnya peneliti meminta data sesuai atribut diatas kepada TU atau operator MI Datuk Singaraja Kerse Jepara.

A	B	C	D	E	F	G
data no	Kelas	NAMA	RAPORT	Status Siswa Yatim/Tidak	Pendapatan Orangtua	Kedisiplinan
1	Kelas 6-0b	Muhammad Nur Fadilah Syahid	80	tidak	>3000,000	CUKUP
2	Kelas 6-0a	Talia Zaidatur Zalfa	81	tidak	<1000,000	BAIK
3	Kelas 6-0b	Muhammad Fatih Roziqin	84	tidak	>2000,000	BAIK
4	Kelas 6-0b	Muhammad Rafi Adhya	77	tidak	>2000,000	BAIK
5	Kelas 3-3a	Lutfiah Anifah May Nita	85	tidak	>3000,000	BAIK
6	Kelas 3-0a	Risa Ramadhani Safri	79	tidak	>2000,000	BAIK
7	Kelas 3-0a	Wafiq Awadhya	80	tidak	>2000,000	BAIK
8	Kelas 2-2a	NUR IDA FITRI WAHYUNI	84	yatim	<1000,000	BAIK
9	Kelas 1-1	Muhammad Amri Adhar	72	tidak	>2000,000	CUKUP
10	Kelas 4-0b	Vebby Noor Hidayah	73	tidak	>2000,000	CUKUP
11	Kelas 3-0a	Intan Nur Ami	80	tidak	>3000,000	BAIK
12	Kelas 6-0b	Muhammad Kevin Oktama	83	tidak	>3000,000	BAIK
13	Kelas 1-1	Apreliani Putri	70	tidak	>2000,000	BAIK
14	Kelas 6	Muhammad Agus Salimman	79	tidak	>2000,000	BAIK
15	Kelas 2-2a	DEWI KHANA WAHDATUL N	87	tidak	>3000,000	BAIK
16	Kelas 2-2a	AHMAD BAIHAQI	78	tidak	>3000,000	BAIK
17	Kelas 3-0a	Devietta Arsyie Angraeni	74	tidak	>2000,000	BAIK
18	Kelas 2-2a	Rista Oktaviani	80	tidak	>2000,000	BAIK
19	Kelas 4-0a	Anita Fairus Khalsa	80	tidak	>2000,000	BAIK
20	Kelas 2-2a	MUHAMMAD KHORUDDIN	88	tidak	>2000,000	BAIK
21	Kelas 6-0b	Muhammad Rizki Abillah	78	tidak	>2000,000	BAIK

**Gambar 12 Data siswa dan Atribut**

Kemudian data siswa sudah lengkap, maka data tersebut sudah siap untuk masuk ke proses selanjutnya, yaitu proses *Data Transformation*.

*Data Transformation* adalah suatu proses untuk mengubah data yang sudah diperoleh untuk mendapatkan data yang sesuai dengan kebutuhan analisa.

Langkah yang dilakukan peneliti dalam proses transformasi data adalah menulis data/ nilai dari atribut yang digunakan kemudian setiap nilai di inisiasi ke dalam bentuk angka diinisialisasi berdasarkan *value* tertinggi sampai *value* terendah. Seperti pada tabel 3,4,5 dan 6. Setelah data siswa melewati proses *data reduction*, *Data Cleaning*, dan *Data Transformation*, maka diperoleh data yang siap untuk dihitung seperti berikut:

A	B	C	D	E	F	G
NO	Kelas	NAMA	RAPORT	Status Siswa Yatim/Tidak	Pendapatan Orangtua	Kedisiplinan
1	Kelas 6-0b	Muhammad Nur Fadilah Syahid	2	1	1	1
2	Kelas 6-0a	Talia Zaidatur Zalfa	2	1	3	2
3	Kelas 6-0b	Muhammad Fatih Roziqin	2	1	2	2
4	Kelas 6-0b	Muhammad Rafi Adhya	2	1	2	2
5	Kelas 3-3a	Lutfiah Anifah May Nita	3	1	1	2
6	Kelas 3-0a	Risa Ramadhani Safri	2	1	2	2
7	Kelas 3-0a	Wafiq Awadhya	2	2	3	2
8	Kelas 2-2a	NUR IDA FITRI WAHYUNI	1	1	2	1
9	Kelas 1-1	Muhammad Amri Adhar	1	1	2	1
10	Kelas 4-0b	Vebby Noor Hidayah	1	1	2	1
11	Kelas 3-0a	Intan Nur Ami	2	1	1	2
12	Kelas 6-0b	Muhammad Kevin Oktama	2	1	1	2
13	Kelas 1-1	Apreliani Putri	1	1	2	2
14	Kelas 6	Muhammad Agus Salimman	2	1	2	2
15	Kelas 2-2a	WI KHANA WAHDATUL MU	3	1	1	2
16	Kelas 3-0a	Devietta Arsyie Angraeni	2	1	1	2
17	Kelas 2-2a	AHMAD BAIHAQI	1	1	2	2
18	Kelas 3-3a	Rista Oktaviani	2	1	2	2
19	Kelas 4-0a	Anita Fairus Khalsa	2	1	2	2

**Gambar 12 Dataset siap dihitung**

**4.2 Penerapan Algoritma K-Means**

Setelah data mentah yang telah diterima dari pihak sekolah telah melewati proses *pre-processing data* dan akhirnya di dapat *dataset* yang sudah siap untuk masuk ke tahap proses selanjutnya adalah penerapan algoritma K-Means dalam perhitungannya digunakan dua aplikasi yaitu aplikasi *Microsoft Excel* dan *RapidMiner*.

**4.3 Penerapan Algoritma K-means dengan Microsoft Excel 2019**

Pada penerapan algoritma *k-means* dengan menggunakan aplikasi *Microsoft Excel 2019* berikut tahapannya:

**1. Tentukan Jumlah Cluster**

Pada penelitian ini, peneliti memutuskan untuk membentuk *cluster* dengan jumlah 3 *cluster*.

**2. Menentukan Titik Pusat Cluster**

Pada metode *clustering* algoritma *k-means* ada titik pusat *cluster* atau biasa dikenal dengan *centroid*. *Centroid* adalah suatu nilai yang digunakan sebagai pengurang dalam perhitungan jarak antara data dengan setiap *cluster*. Pada penelitian ini, metode yang digunakan untuk menentukan nilai titik pusat *cluster* yaitu metode acak. Data yang terpilih ditempatkan pada *cluster* satu sampai dengan



cluster tiga. sehingga nilai dari setiap titik pusat cluster seperti berikut ini:

**Tabel 10. Titik Pusat Cluster Awal**

Cluster	NAMA	Raport	Status	Pendapatan	Kedisiplinan
1	Puja Naura Bidrinza	2	2	1	1
2	Muhamad Khoirudin	3	1	2	2
3	Wilda Izzaturrohmah	1	2	3	1

**1. Menghitung Jarak Data Ke Setiap Cluster**

langkah selanjutnya adalah menghitung jarak data ke setiap Cluster

Dengan menggunakan aplikasi Microsoft Excel peneliti mengadaptasi rumus euclidean distance untuk melakukan perhitungan jarak seperti berikut:

$$= \text{SQRT}(((\text{data pertama status siswa} - \text{centroid status siswa})^2) + ((\text{data pertama kedisiplinan} - \text{centroid kedisiplinan})^2) + ((\text{data pendapatan orang tua} - \text{centroid pendapatan orang tua})^2) + ((\text{data pertama raport} - \text{centroid raport})^2))$$

**Gambar 13 Perhitungan Jarak Data Ke Setiap Cluster**

**4.3 Mengalokasikan Data Ke Dalam Cluster**

Apabila jarak data memiliki jarak terdekat ke suatu Cluster, maka data tersebut masuk dalam Cluster tersebut untuk mengalokasikan data ke Cluster tertentu, rumus yang digunakan seperti berikut:

$$= \text{IF}(U2=\text{MIN}(U2;V2;W2);1;\text{IF}(V2=\text{MIN}(U2;V2;W2);2;3))$$

Dimana:

= IF yaitu fungsi yang biasa digunakan untuk menentukan suatu nilai dengan kondisi tertentu.

$$=(U2=\text{MIN}(U2;V2;W2);1;$$

(U2 adalah column dari Cluster 1

Min MIN(U2;V2;W2);1 adalah lebih kecil dari Cluster 1, 2, 3 maka masuk ke Cluster 1

$$;=\text{IF}(V2=\text{MIN}(U2;V2;W2);2;3))$$

(V2= adalah column dari Cluster 2

MIN(U2;V2;W2);2;3)) adalah memiliki arti lebih kecil dari Cluster 2, 2, 3 maka masuk ke Cluster 2 dan selain itu masuk ke dalam Cluster 3

seperti pada gambar berikut:

**Gambar 14. Mengalokasikan Data ke Setiap Cluster**

**5. Menentukan Titik Pusat Cluster Baru**

Cara untuk menentukan titik pusat cluster baru ini menggunakan nilai rata rata sesuai dengan data yang telah dikelompokkan ke dalam masing masing cluster.

**Tabel 11. Titik Pusat Cluster Baru 1**

Titik Pusat Cluster Baru	Raport	Status Siswa	Pendapatan Orangtua	Kedisiplinan
Cluster 1	313,00	178,00	208,00	265,00

Cluster 1 yang berjumlah 165 data

**Tabel 12. Titik Pusat Cluster Baru 2**

Titik Pusat Cluster Baru	Raport	Status Siswa	Pendapatan Orangtua	Kedisiplinan
Cluster 2	652,00	294,00	575,00	583,00

Pada Cluster 2 yang berjumlah 294 data

**Tabel 13. Titik Pusat Cluster Baru 3**

Titik Pusat Cluster Baru	Raport	Status Siswa	Pendapatan Orangtua	Kedisiplinan
Cluster 3	1000,00	1000,00	1000,00	1000,00

Cluster 3	77,00	79,00	141,00	97,00
-----------	-------	-------	--------	-------

Pada Cluster 3 yang berjumlah 59 data

**6. Verifikasi Titik Pusat Cluster**

Proses ini merupakan suatu proses untuk melakukan verifikasi terhadap titik pusat Cluster baru dengan titik pusat Cluster lama. Apabila ditemukan perbedaan dari nilai centroid yang dimiliki dari kedua titik pusat Cluster tersebut, maka proses K-Means akan tetap berlanjut dan dimulai lagi dari proses ke tiga yaitu (proses menghitung jarak data ke setiap Cluster) dengan menggunakan titik pusat Cluster yang baru. Namun apabila nilai centroid dari kedua titik pusat Cluster tersebut sama, maka proses K-Means dihentikan.

Pada penelitian ini perhitungan dilakukan sebanyak 5 kali sampai proses k-means

**Gambar 15 verifikasi titik pusat cluster pada literasi pertama**

**Gambar 16 Verifikasi titik pusat cluster pada literasi terakhir**

Bahwa hasil dari perhitungan literasi ke 4 dan ke 5 memiliki hasil yang sama. Maka proses K-Means Clustering sudah selesai.

Kemudian peneliti melakukan penghitungan jumlah data pada setiap kriteria pada masing atribut yang digunakan, dengan menggunakan rumus Microsoft Excel untuk mempermudah proses perhitungan berikut rumus yang digunakan:

=COUNTIF(range;"syarat")

- Dimana COUNTIF digunakan untuk menghitung jumlah data dengan kondisi syarat tertentu

- Range yaitu area colum atau row data yang akan masuk jangkauan hitungan
- Syarat adalah kondisi tertentu yang harus diterapkan dalam perhitungan data tersebut

=COUNTIF(S3:S518;"2")

Maka data yang dihitung adalah data dengan kriteria Layak atau memiliki nilai 2 pada atribut status siswa, maka data yang akan dijumlah adalah data yang memiliki nilai atribut 2 saja, seperti pada gambar berikut:

**Gambar 17. Perhitungan jumlah data di setiap kriteria**

Setelah proses perhitungan dengan menggunakan algoritma K-Means menggunakan Microsoft Excel sudah terselesaikan dan telah didapatkan hasil dari perhitungannya, maka langkah selanjutnya adalah menampilkan hasil dari perhitungan tersebut.

**Tabel 14. Hasil Perhitungan Cluster 1**

Cluster 1				
Kriteria	Atribut			
	RAPORT	Status Siswa (Yatim/Piatu/Tidak)	Pendapatan Orangtua	Kedisiplinan
Sangat Layak	64	0	0	152
Layak	98	0	37	0
Tidak Layak	18	180	143	28
Sangat Tidak Layak	0	0	0	0
<b>Total</b>	<b>180</b>	<b>180</b>	<b>180</b>	<b>180</b>

**Sebanyak 180 Siswa yang masuk dalam Cluster 1**

Cluster. Informasi yang paling terlihat pada Cluster 1 yaitu kriteria sangat layak berdasarkan atribut nilai raport sebanyak 64 siswa dan layak sebanyak 98 siswa dan kedisiplinan sebanyak 152 siswa dengan kriteria sangat layak.

**Tabel 15. Hasil Perhitungan Cluster 2**

Cluster 2	
Kriteria	Atribut

	Raport	Status Siswa (Yatim/Piatu/Tidak)	Pendapatan Orang tua	Kedisiplinan
Sangat Layak	0	13	0	10
Layak	13	0	8	0
Tidak Layak	0	0	5	3
Sangat Tidak Layak	0	0	0	0
<b>Total</b>	<b>13</b>	<b>13</b>	<b>13</b>	<b>13</b>

Sebanyak 13 Siswa yang masuk dalam Cluster 2

Informasi yang dapat didapat pada Cluster 2 yaitu kriteria layak berdasarkan atribut nilai raport sebanyak 13 siswa dan dari atribut status siswa sangat layak sebanyak 13 siswa. Berdasarkan hal diatas dapat dilihat bahwa banyakan siswa yang terdapat dalam Cluster 2 memiliki kriteria sangat layak dan layak berdasarkan atribut status dan kedisiplinan dan sisanya berada kriteria yang tidak layak sesuai dengan tabel diatas.

**Tabel 16. Hasil Perhitungan Cluster 3**

Cluster 3				
Kriteria	Atribut			
	Raport	Status Siswa (Yatim/Piatu/Tidak)	Pendapatan Orang tua	Kedisiplinan
Sangat Layak	5	22	40	267
Layak	277	0	283	0
Tidak Layak	41	301	0	56
Sangat Tidak Layak	0	0	0	0
<b>Total</b>	<b>323</b>	<b>323</b>	<b>323</b>	<b>323</b>

Sebanyak 323 Siswa yang masuk dalam Cluster 3

Informasi yang paling terlihat pada Cluster 3 yaitu kriteria sangat layak berdasarkan atribut nilai raport sebanyak 5 siswa dan layak sebanyak 277 siswa, berdasarkan atribut status siswa dengan kriteria sangat layak sebanyak 22 siswa, sebanyak 40 siswa dengan kriteria sangat layak dari atribut berdasarkan pendapatan orang tua, dan 267 siswa sangat layak berdasarkan atribut kedisiplinan. Sisanya berada pada kriteria yang tidak layak sesuai dengan tabel diatas.

**4.4 Evaluasi dengan metode BCV Dan WCV**

Dalam penelitian terhadap data siswa ini, dilakukan evaluasi untuk mengetahui baik tidaknya hasil perhitungan yang diperoleh. Semakin kecil nilai error yang didapat maka semakin baik hasil perhitungan yang diperoleh. Berikut ini adalah proses evaluasi tersebut.

**1. Menentukan Literasi ke berapa untuk Dihitung**

literasi yang digunakan adalah literasi terakhir karena iterasi terakhir mempunyai kualitas centroid yang lebih baik dari iterasi sebelumnya.

**Tabel 17. Pusat Centroid Baru**

Titik Pusat	Raport	Status Siswa Yatim/Piatu/Tidak	Pendapatan Orangtua	Kedisiplinan
Cluster 1	2,26	1,00	1,21	1,84
Cluster 2	2,00	2,00	1,62	1,77
Cluster 3	1,89	1,07	2,12	1,83

Selanjutnya lakukan perhitungan dengan persamaan *Between-Class Variation (BCV)* untuk nilai centroid.

$$C1 = \sqrt{(2,26-1,00)^2+(1,00-2,00)^2+(1,21-1,32)^2+(1,84-1,77)^2}$$

$$C2 = \sqrt{(2,26-1,89)^2+(1,00-1,07)^2+(1,21-2,12)^2+(1,84-1,83)^2}$$

$$C3 = \sqrt{(2,00-1,89)^2+(2,00-1,07)^2+(1,62-2,12)^2+(1,77-1,83)^2} = 3,17$$

**2. Menentukan Jarak Minimum Centroid**

Pada proses ini menggunakan jarak minimum pusat centroid yang didapat pada literasi terakhir,

dengan menggunakan persamaan *Within-Class Variation (WCV)*. Berikut ini adalah perhitungannya:

$$WCV = 0,906^2 + 0,903^2 + 0,250^2 + 0,250^2 + 0,788^2 + 0,250^2 + 0,250^2 + 1,296^2 + 1,222^2 + 1,222^2 + 0,363^2 + 0,363^2 + 0,916^2 + 0,250^2 + 0,250^2 + 0,363^2 + 0,846^2 = 272,495$$

Sehingga hasil dari perhitugan WCV = 272,495

**3. Menghitung Perbandingan BCV dan WCV**

Pada tahap terakhir adalah menghitung

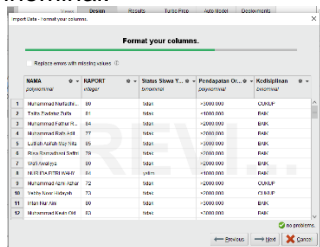
perbandingan nilai BCV dan WCV sehingga menghasilkan nilai eror hitung dengan persamaan rasio seperti terlihat pada hasil dibawah ini:

$$\text{Rasio} = \frac{272,495}{3,17} = 0,0116459$$

Untuk menentukan bagus tidaknya nilai rasio dari yang didapat maka harus memperhatikan kriteria pengukuran rasio, dapat dilihat di tabel rasio atau tabel 1 Hasil pengujian menggunakan perbandingan Between Class Variation mendapatkan nilai rasio yang tidak tinggi yaitu 0,0116459 dan artinya rasio mendapatkan kualitas yang sangat layak atau sangat baik.

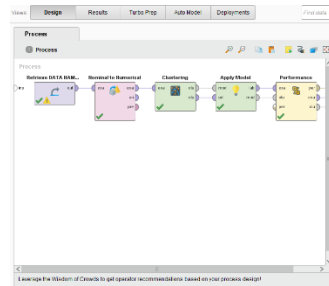
**4.4 Penerapan Algoritma K-Means dengan Aplikasi RapidMiner**

Penerapan Algoritma *K-Means Clustering* selanjutnya menggunakan aplikasi *RapidMiner*. Dengan menggunakan aplikasi *RapidMiner* ini perlu dipersiapkan dataset yang akan diimport ke dalam aplikasi *RapidMiner*. Dataset tersebut disiapkan dengan menggunakan aplikasi *Microsoft Excel 2019*. Selanjutnya import dataset, dan tentukan indikator atribut yang terdapat didalamnya dengan menggunakan atribut yang telah disiapkan, dan atribut tersebut lebih dari dua klasifikasi, maka atribut tersebut bertipe polynominal jika atribut tersebut memiliki dua klasifikasi maka bertipe binominal.



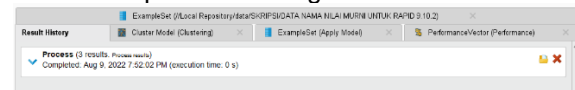
**Gambar 18. Proses Import Dataset Pada Aplikasi RapidMiner**

Langkah selanjutnya adalah masuk ke desain, berikut langkah desain dari metode *K-Means Clustering* seperti gambar di bawah ini:



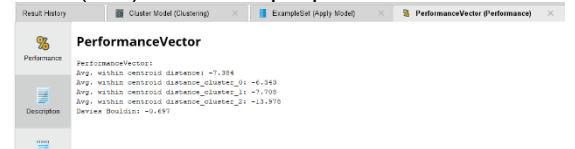
**Gambar 1. Desain Clustering di Aplikasi RapidMiner**

Proses dimulai dengan retrieve data siswa siswi MI Datuk Singaraja, dataset disimpan dengan nama NAMA NILAI MURNI UNTUK RAPID 9.10.2. Selanjutnya DATASET DISAMBUNGAN ke nominal to numerical sebelum dilakukan pengujian *K-Means*. Kemudian di sambung ke apply mode, setelah itu disambungkan ke performance dan kemudian klik RUN. Kemudian klik Result History maka diperoleh informasi tentang lama waktu proses algoritma *K-Means*

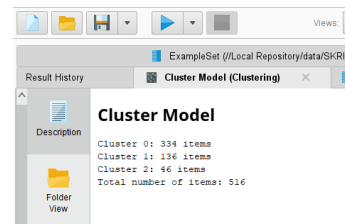


**Gambar 19. Result Histori**

Berikut adalah hasil dari nilai Davies Bouldin Index(DBI) dan deskripsi performance



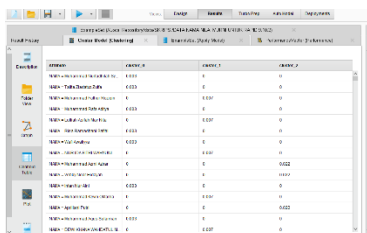
**Gambar 20. Performance dan DBI**



**Gambar 21. Hasil Cluster model**

Dalam penelitian ini peneliti menggunakan 3 Cluster, dimulai dari Cluster 0 dan yang terakhir adalah Cluster 2. Sesuai dengan pendefinisian nilai k pada Cluster 0 adalah 334 items, Cluster 1 yang berjumlah 136 items dan Cluster 2 yang berjumlah 46 items dengan total dataset 516 items.

Dengan menggunakan aplikasi *RapidMiner*, berikut adalah hasil dari perhitungan jarak Cluster dengan centroid



**Gambar 22. Hasil Perhitungan Jarak Cluster Terhadap Centroid Di Aplikasi RapidMiner**



**Gambar 23. Hasil Set Diagram Scatter**

Berdasarkan hasil penelitian yang dilakukan terhadap data siswa siswi MI Datuk Singaraja dari tahun 2021 sampai dengan 2022, menggunakan aplikasi *Microsoft Excel* 2019 dan *RapidMiner*. Berikut ini adalah perbandingan hasil perhitungan yang telah diperoleh.

**Tabel 18. Tabel Perbandingan Hasil Perhitungan K-Means Clustering**

Perbandingan	Microsoft Excel			RapidMiner		
Banyaknya literasi	5					
Jumlah Cluster	1	2	3	1	2	3
	180	13	323	334	136	46
Jumlah centroid	3			3		
Jumlah nilai rasio	34,9 %	2,50 %	62,6 0%	64,8 %	26,3 %	8,9 %

**5. KESIMPULAN**

Berdasarkan penelitian yang dilakukan oleh peneliti, kesimpulan yang di dapat yaitu: Pada penelitian ini diperoleh 3 *cluster*.

Pada Cluster 1 terdapat 180 siswa, 64 siswa sangat baik, 98 siswa mendapatkan kriteria baik dan 18 siswa mendapatkan nilai cukup pada atribut nilai raport. Dalam atribut Status siswa yatim/tidak tidak ada siswa atau nol. Pada atribut pendapatan orang tua 37

siswa mendapatkan kriteria layak dan 143 siswa cukup layak.

Maka dapat disimpulkan pada *Cluster* 1 kebanyakan siswa berprestasi dengan nilai raport sangat baik dengan keadaan ekonomi berkecukupan, dan termasuk dalam *Cluster* yang tidak di prioritaskan menerima bantuan.

Pada *Cluster* 2 terdapat 13 siswa, semuanya mendapatkan hasil yang layak pada kriteria nilai raport, pada atribut status siswa yatim/tidak terdapat 13 siswa mendapatkan kriteria sangat layak.

Maka dapat disimpulkan pada *Cluster* 2 termasuk dalam *Cluster* yang cukup di prioritaskan.

Pada *Cluster* 3 terdapat 323 siswa dengan hasil yang variatif, pada atribut raport, 5 siswa masuk dalam kriteria sangat layak, 277 layak, 41 cukup layak, pada atribut status siswa yatim/ tidak terdapat 22 siswa yang mendapatkan kriteria sangat layak Maka dapat disimpulkan pada *Cluster* ke 3 termasuk dalam *Cluster* yang di prioritaskan menerima bantuan.

Dengan terbentuknya *Cluster* penerima bantuan fasilitas sekolah di MI Datuk Singaraja Kerso, Hasil yang di dapat dari penelitian tersebut dapat dijadikan masukan kepada pihak sekolah agar dapat lebih objektif lagi dalam memprioritas penerima bantuan fasilitas sekolah dengan menggunakan 3 kriteria yang telah dilakukan.

**6. REFERENSI**

[1] S. Vaporizki, "Manajemen Kurikulum Berbasis Industri Kreatif Pada Kompetensi Keahlian Kriya Kreatif Logam dan Perhiasan SMKN 12 Surabaya," *J. Din. Manaj. Pendidik.*, vol. 3, no. 2, p. 62, 2019, doi: 10.26740/jdmp.v3n2.p62-71.

[2] M. A. Apriliani, A. Maksum, P. A. Wardhani, S. Yuniar, and S. Setyowati, "Pengembangan media pembelajaran PPKn SD berbasis Powtoon untuk mengembangkan karakter tanggung jawab," *J. Ilm. Pendidik. Dasar*, vol. 8, no. 2, p. 129, 2021, doi: 10.30659/pendas.8.2.129-145.

[3] D. P. Sari, "Penerapan Metode SAW (Simple Additive Weighting) Dalam Sistem Pendukung Keputusan Untuk Menentukan Penerima Beasiswa Pada SMK N 2 Payakumbuh," vol. 5, no. 2, pp. 207–216, 2018.

- [4] F. Marisa, "Educational Data Mining (Konsep dan Penerapan)," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 91–93, 2013.
- [5] M. Ayub, "Proses Data Mining dalam Sistem Pembelajaran Berbantuan Komputer," pp. 21–30.
- [6] B. S. Praja, P. D. Kusuma, and C. Setianingsih, "Penerapan Metode K-Means Clustering Dalam Pengelompokan Data Penumpang Dan Kapal Angkutan Laut Di Indonesia," *e-Proceeding Eng.*, vol. 06, no. 1, p. 1442, 2019.
- [7] A. K-means and W. Kabupaten, "Penentuan Alih Fungsi Lahan Marginal Menjadi Lahan Pangan Berbasis," 2011.
- [8] Widodo PP, Handayanto RT, Herlawati. 2013. *Penerapan Data Mining Dengan Matlab*. Bandung: Rekayasa Sains.
- [9] Maimon, O. dan Last, M.2000. "Knowledge Discovery and Data Mining, The Info-Fuzzy Network (IFN) Methodology" Dordrecht Kluwer Academic.
- [10] Prasetyowati, Erwin, 2017. *Data Mining Pengelompokan Data untuk Informasi dan Evaluasi*. Pamekasan: Duta Media Publishing
- [11] Larose DT. 2005. *Discovering Knowledge in Data: An Introduction to Data*.
- [12] N. Putu, E. Merliana, and A. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means," pp. 978–979.
- [13] Han J, Kamber M, Pei J. 2012. *Data Mining Concepts and Techniques Third Edition*. Waltham: Elsevier Inc.
- [14] Tan PN, Stenbach M, Kumar V. 2006. *Introduction to Data Mining*. Boston: Pearson Education
- [15] A. Agrawal and H. Gupta, "Global K-Means (GKM) Clustering Algorithm: A Survey," *Int. J. Comput. Appl.*, vol. 79, no. 2, pp. 20–24, 2013, doi: 10.5120/13713-1472.
- [16] B. S. Praja, P. D. Kusuma, and C. Setianingsih, "Penerapan Metode K-Means Clustering Dalam Pengelompokan Data Penumpang Dan Kapal Angkutan Laut Di Indonesia," *e-Proceeding Eng.*, vol. 06, no. 1, p. 1442, 2019.
- [17] N. Damanik And M. Sigiro, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," Pp. 33–43.
- [18] Hariyanto, Muhammad, Dan Rizky Tahara Shita, 2018. Clustering Pada Data Mining Untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means Dan Metode Perhitungan Jarak Euclidean Distance. SKANIKA, 1 (1), Pp. 117-122