

**SKRIPSI**

**PENERAPAN *CLUSTERING* PADA SISTEM  
PEMBAGIAN ROMBONGAN BELAJAR SISWA  
MENGUNAKAN ALGORITMA *K-MEANS***



**WIDHI ZAEFUAN  
NPM. 13.0504.0119**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH MAGELANG  
FEBRUARI, 2018**

**SKRIPSI**

**PENERAPAN *CLUSTERING* PADA SISTEM  
PEMBAGIAN ROMBONGAN BELAJAR SISWA  
MENGUNAKAN ALGORITMA *K-MEANS***

**Disusun Sebagai Salah Satu Syarat Memperoleh Gelar Sarjana Komputer (S. Kom)  
Program Studi Teknik Informatika Jenjang Strata Satu (S-1) Fakultas Teknik  
Universitas Muhammadiyah Magelang**



**WIDHI ZAEFUAN  
NPM. 13.0504.0119**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH MAGELANG  
FEBRUARI, 2018**

## **HALAMAN PENEGASAN**

Tugas Akhir/Skripsi ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.

Nama : WIDHI ZAEFUAN

NPM : 13.0504.0119

Magelang, 13 Februari 2018

A handwritten signature in black ink, appearing to read 'Widhi Zaefuan', enclosed within a large, stylized circular flourish.

NPM. 13.0504.0119

## SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini :

Nama : Widhi Zaefuan  
NPM : 13.0504.0119  
Program Studi : Teknik Informatika S1  
Fakultas : Teknik

Dengan ini saya menyatakan bahwa **SKRIPSI** dengan judul “**PENERAPAN *CLUSTERING* PADA SISTEM PEMBAGIAN ROMBONGAN BELAJAR SISWA MENGGUNAKAN ALGORITMA *K-MEANS***” beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disertakan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk risiko / sanksi yang berlaku.

Magelang, 13 Februari 2018  
Yang Menyatakan



**Widhi Zaefuan**  
NPM. 13.0504.0119

HALAMAN PENGESAHAN

SKRIPSI

PENERAPAN *CLUSTERING* PADA SISTEM PEMBAGIAN  
ROMBONGAN BELAJAR SISWA MENGGUNAKAN ALGORITMA  
*K-MEANS*

dipersiapkan dan disusun oleh

**WIDHI ZAEFUAN**  
**NPM. 13.0504.0119**

Telah dipertahankan di depan Dewan Penguji  
Pada tanggal 13 Februari 2018

Susunan Dewan Penguji

Pembimbing I



R. Arri Widyanto, S.Kom., MT  
NIDN. 0616127102

Penguji I



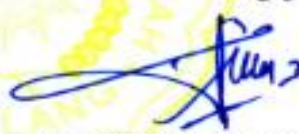
Purwono Hendradi, M.Kom  
NIDN. 0624077101

Pembimbing II



Endah Ratna A, S.Kom., M.Cs  
NIDN. 0601129001

Penguji II



Auliya Burhanuddin, S.Si., M.Kom  
NIDN. 0630058202

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 13 Februari 2018  
Dekan



Yun Arifatul Fatimah, ST., MT., Ph.D  
NIK. 987408139

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK  
KEPENTINGAN AKADEMIS**

Sebagai Civitas Akademik Universitas Muhammadiyah Magelang, yang bertanda tangan di bawah ini:

Nama : Widhi Zaefuan  
NPM : 13.0504.0119  
Program Studi : Teknik Informatika S1  
Fakultas : Teknik  
Jenis Karya : Skripsi

Menyatakan bahwa demi pengembangan ilmu pengetahuan, menyetujui memberikan kepada Fakultas Teknik Universitas Muhammadiyah Magelang **Hak Bebas Royalti Noneksklusif (*Non-Exclusive Royalty-Free Right*)** atas karya ilmiah yang berjudul:

PENERAPAN *CLUSTERING* PADA SISTEM PEMBAGIAN ROMBONGAN  
BELAJAR SISWA MENGGUNAKAN ALGORITMA *K-MEANS*

Dengan hak Bebas Royalti Noneksklusif ini Fakultas Teknik Universitas Muhammadiyah Magelang berhak menyimpan, mengalihmedia/memformatkan, mengelola dalam bentuk pangkalan data, merawat, dan mempublikasikan Skripsi tersebut selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenar-benarnya tanpa paksaan dari pihak manapun.

Dibuat Di : Magelang

Pada Tanggal : 13 Februari 2018

Yang Menyatakan,



**Widhi Zaefuan**

NPM: 13.0504.0119

## KATA PENGANTAR

*Alhamdulillah* *rabbi'l'alamin*, puji syukur kehadiran Allah SWT yang telah melimpahkan segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Skripsi yang berjudul “**PENERAPAN *CLUSTERING* PADA SISTEM PEMBAGIAN ROMBONGAN BELAJAR SISWA MENGGUNAKAN ALGORITMA *K-MEANS*”**”.

Penyusunan Skripsi ini dibuat dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Program Studi Teknik Informatika (S1) Fakultas Teknik Universitas Muhammadiyah Magelang.

Penyelesaian Skripsi ini banyak memperoleh bantuan dan bimbingan dari berbagai pihak. Oleh karena itu, diucapkan terima kasih kepada :

1. Ir. Eko Muh Widodo, MT. selaku Rektor Universitas Muhammadiyah Magelang.
2. Yun Arifatul Fatimah, ST., MT., Ph.D selaku Dekan Fakultas Teknik Universitas Muhammadiyah Magelang.
3. R. Arri Widyanto, S.Kom., MT dan Endah Ratna A, S.,Kom., M.Cs selaku Dosen Pembimbing yang telah memberikan waktu dan ilmunya dalam penulisan Skripsi ini.
4. Ayah dan Ibu tercinta serta keluarga yang selalu melapangkan do'a dan kasih sayang sehingga penulis tidak patah semangat dalam menyusun Skripsi ini.
5. Kepala Sekolah Menengah Pertama Negeri 10 Magelang yang telah memeberikan ijin dalam pengambilan data penelitian.
6. Segenap keluarga besar SMP N 10 Magelang yang telah memberikan dukungan dalam penyusunan skripsi.
7. Segenap Dosen dan Staf di lingkungan Fakultas Teknik Universitas Muhammadiyah Magelang.
8. Teman-teman seperjuangan yang selalu berbagi ilmu dalam penyusunan Skripsi ini.

Akhir kata, semoga Allah SWT berkenan membalas segala kebaikan semua pihak yang telah membantu dan semoga Skripsi ini membawa manfaat bagi pengembangan ilmu di masa yang akan datang.

Magelang, 13 Februari 2018



WIDHI ZAEFUAN  
13.0504.0119

## DAFTAR ISI

<b>HALAMAN KULIT MUKA</b> .....	i
<b>HALAMAN JUDUL</b> .....	ii
<b>HALAMAN PENEGASAN</b> .....	iii
<b>HALAMAN PERNYATAAN KEASLIAN</b> .....	iv
<b>HALAMAN PENGESAHAN</b> .....	v
<b>HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI</b> .....	vi
<b>KATA PENGANTAR</b> .....	vii
<b>DAFTAR ISI</b> .....	viii
<b>DAFTAR GAMBAR</b> .....	x
<b>DAFTAR TABEL</b> .....	xi
<b>DAFTAR LAMPIRAN</b> .....	xiii
<b>ABSTRAK</b> .....	xiv
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
A. Latar Belakang .....	1
B. Rumusan masalah .....	2
C. Tujuan Penelitian .....	2
D. Manfaat Penelitian .....	3
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>4</b>
A. Penelitian Relevan .....	4
B. Penjelasan Secara Teoritis Masing-Masing Variabel Penelitian .	6
C. Landasan Teori .....	17
<b>BAB III ANALISIS DAN PERANCANGAN SISTEM</b> .....	<b>18</b>
A. Analisis Sistem .....	18
B. Perancangan Sistem .....	20
<b>BAB IV IMPLEMENTASI DAN PENGUJIAN</b> .....	<b>31</b>
A. Implementasi Data .....	31
B. Implementasi Rapid Miner .....	33
C. Menentukan Kelompok Unggulan .....	39
D. Pembentukan Rombel / Kelas .....	41
E. Pengujian .....	42



<b>BAB V HASIL DAN PEMBAHASAN .....</b>	<b>45</b>
A. Hasil .....	45
B. Pembahasan .....	49
<b>BAB IV PENUTUP .....</b>	<b>50</b>
A. Kesimpulan .....	50
B. Saran .....	50
<b>DAFTAR PUSTAKA .....</b>	<b>51</b>
<b>LAMPIRAN .....</b>	<b>53</b>

## DAFTAR GAMBAR

Gambar 2.1.	Tahap-Tahap Data Mining .....	11
Gambar 3.1.	Diagram Alur Permasalahan .....	19
Gambar 3.2.	Alur Tahapan Penelitian .....	20
Gambar 3.3.	Use Case Pemodelan Sistem .....	21
Gambar 3.4.	Alur Algoritma K-Means .....	22
Gambar 3.5.	State Diagram Sistem .....	28
Gambar 3.6.	Proses Input Data .....	29
Gambar 3.7.	Menghubungkan Antar Operator .....	29
Gambar 3.8.	Tampilan Hasil Output .....	30
Gambar 4.1.	Import Data Dan Seleksi Data .....	33
Gambar 4.2.	Pembersihan Data .....	34
Gambar 4.3.	Transformasi Data .....	34
Gambar 4.4.	<i>K-Means Clustering</i> .....	34
Gambar 4.5.	Ekspor Data .....	35
Gambar 4.6.	Pemodelan <i>K-Means Clustering</i> .....	35
Gambar 4.7.	Jumlah Anggota <i>Cluster</i> Tahun 2016/2017 .....	35
Gambar 4.8.	<i>Centroid</i> Akhir Tahun 2016/2017 .....	36
Gambar 4.9.	Data Anggota <i>Cluster</i> Tahun 2016/2017 .....	37
Gambar 4.10.	Jumlah Anggota <i>Cluster</i> Tahun 2015/2016 .....	38
Gambar 4.11.	Centroid Akhir Tahun 2015/2016 .....	38
Gambar 4.12.	Data Anggota <i>Cluster</i> Tahun 2015/2016 .....	39
Gambar 5.1.	Grafik Jumlah Siswa Masing-Masing <i>Cluster</i> .....	45
Gambar 5.2.	Range Nilai Masing-Masing <i>Cluster</i> Tahun 2016/2017 .....	45

Gambar 5.3.	Range Nilai Masing-Masing <i>Cluster</i> Tahun 2015/2016 .....	46
Gambar 5.4.	Grafik Jumlah Siswa Masing-Masing <i>Cluster</i> .....	46
Gambar 5.5.	Daftar Nama Siswa Tahun 2016/2017 .....	48
Gambar 5.6.	Daftar Nama Siswa Tahun 2015/2016 .....	48

## DAFTAR TABEL

Tabel 3.1	Data Sampel Penelitian Tahun 2016/2017 .....	19
Tabel 3.2	Data Sampel Penelitian Tahun 2015/2016 .....	20
Tabel 3.3	Menentukan Nilai <i>Centroid</i> .....	23
Tabel 3.4	Sampel 15 Data Pertama .....	23
Tabel 3.5	Jarak Masing-Masing Data Dengan <i>Centroid</i> .....	25
Tabel 3.6	Mengelompokkan Data Berdasarkan Jarak Terdekat .....	25
Tabel 3.7	<i>Centroid</i> Baru Pada Iterasi 1 .....	26
Tabel 3.8	Rata-Rata <i>Centroid</i> Kelompok Unggulan.....	27
Tabel 4.1	Nilai <i>Centroid</i> Awal .....	32
Tabel 4.2	Nilai <i>Centroid</i> Akhir Data Tahun Ajaran 2016/2017 .....	32
Tabel 4.3	Nilai <i>Centroid</i> Akhir Data Tahun Ajaran 2015/2016 .....	33
Tabel 4.4	Data Hasil <i>Clustering</i> .....	40
Tabel 4.5	Hasil Penentuan Kelompok Unggul, Sedang Dan Bawah .....	40
Tabel 4.6	Penyebaran Siswa Data Tahun Ajaran 2016/2017 .....	41
Tabel 4.7	Penyebaran Siswa Data Tahun Ajaran 2015/2016 .....	41
Tabel 4.8	Pengujian Jumlah Anggota <i>Cluster</i> Tahun 2016/2017 .....	42
Tabel 4.9	Pengujian <i>Centroid</i> Akhir Data Tahun 2016/2017 .....	42
Tabel 4.10	Pengujian Jumlah Anggota <i>Cluster</i> Tahun 2015/2016 .....	43
Tabel 4.11	Pengujian <i>Centroid</i> Akhir Data Tahun 2015/2016 .....	43

## DAFTAR LAMPIRAN

Lampiran 1.	Nilai penegas tahun ajaran 2016/2017 .....
Lampiran 2.	Nilai penegas tahun ajaran 2015/2016 .....
Lampiran 3.	Hasil Transformasi Data Tahun 2016/2017 .....
Lampiran 4.	Hasil Transformasi Data Tahun 2015/2016 .....
Lampiran 5.	Hasil <i>Clustering</i> Manual 2016/2017 .....
Lampiran 6.	Hasil <i>Clustering</i> Manual 2015/2016 .....
Lampiran 7.	Hasil <i>Clustering</i> Rapid Miner 2016/2017 .....
Lampiran 8.	Hasil <i>Clustering</i> Rapid Miner 2015/2016 .....
Lampiran 9.	Hasil Pembentukan Rombel Tahun 2016/2017 .....
Lampiran 10.	Hasil Pembentukan Rombel Tahun 2015/2016 .....

## ABSTRAK

### PENERAPAN *CLUSTERING* PADA SISTEM PEMBAGIAN ROMBONGAN BELAJAR SISWA MENGGUNAKAN ALGORITMA *K-MEANS*

Oleh : Widhi Zaefuan  
Pembimbing : 1. R. Arri Widyanto, S.Kom., MT  
2. Endah Ratna A, S.Kom., M.Cs

Sistem pembagian rombongan belajar / kelas di SMP Negeri 10 Magelang belum menghasilkan kelompok siswa homogen dalam prestasi belajar sehingga berdampak pada proses pembelajaran yang berjalan menjadi lebih lambat. Untuk mengatasi masalah tersebut dilakukan penelitian dengan menerapkan metode *clustering* algoritma *K-Means* untuk menganalisa data siswa SMP Negeri 10 Magelang sehingga dapat ditentukan kelompok unggulan, kelompok sedang dan kelompok bawah guna menunjang proses pembagian rombongan belajar siswa. *K-Means* merupakan metode *clustering* non hirarki yang mempartisi data ke dalam bentuk dua atau lebih *cluster* dalam jumlah sampel data yang besar. *K-Means* mengelompokkan data yang memiliki karakteristik sama ke dalam satu *cluster* dan mengelompokkan data yang memiliki karakteristik berbeda ke dalam *cluster* lain. Penelitian dilakukan dengan menggunakan atribut nilai mata pelajaran Bahasa Indonesia, Bahasa Inggris, Matematika dan IPA sebagai parameter untuk menjalankan algoritma *K-Means*. Hasil dari penelitian ini menunjukkan bahwa sampel data tahun ajaran 2016/2017 menghasilkan *cluster* model  $c_0=71$ ,  $c_1=26$  dan  $c_2=110$ , sedangkan pada sampel data tahun ajaran 2015/2016 menghasilkan *cluster* model  $c_0=105$ ,  $c_1=71$  dan  $c_2=26$ . *Centroid* akhir yang dihasilkan menunjukkan bahwa pada tahun 2016/2017  $c_1$  adalah *Cluster* Unggul,  $c_0$  adalah *Cluster* Sedang, dan  $c_2$  adalah *Cluster* Bawah sedangkan *centroid* akhir pada tahun 2015/2016 menunjukkan  $c_2$  adalah *Cluster* Unggul,  $c_1$  adalah *Cluster* Sedang dan  $c_0$  adalah *Cluster* Bawah. Pada tahun 2016/2017 *Cluster* Unggul terdapat 26 siswa yang cukup untuk menempati 1 kelas saja yaitu kelas IX A, *Cluster* Sedang dengan jumlah 71 siswa dibagi ke dalam 2 kelas yaitu kelas IX B dan IX C masing-masing 32 siswa. Sisa 7 siswa dari *Cluster* Sedang ditentukan berdasarkan jarak terdekat terhadap *centroid Cluster* Bawah, kemudian ditempatkan di kelas IX D, IX E, IX F dan IX G bersama dengan anggota *Cluster* Bawah.

**Kata Kunci :** *K-Means*, *Clustering*, Pembagian Kelas.

## **ABSTRACT**

### **IMPLEMENTATION OF CLUSTERING IN STUDENT GROUP DISTRIBUTION SYSTEM USING K-MEANS ALGORITHM**

By : Widhi Zaefuan  
Supervisor : 1. R. Arri Widyanto, S.Kom., MT  
2. Endah Ratna A, S.Kom., M.Cs

*The distribution of the study group / class at SMP Negeri 10 Magelang has not been able to produce a group of homogeneous students in the learning achievement so that the impact on the learning process runs more slowly. To overcome the problem, the research was done by applying clustering method of K-Means algorithm to analyze the data of the students of SMP Negeri 10 Magelang so that it can be determined the Excellent Group, the Middle Group and the Lower Group in order to support the student class distribution. K-Means is a non-hierarchical clustering method that divides the data into two or more clusters in large amounts of data samples. K-Means classifies data that share the same characteristics into one cluster and classifies data that have different characteristics into other clusters. The study was conducted by using the marks of Indonesian, English, Mathematics and Science Subjects as parameters to run K-Means algorithm. The result of this research showed that the sample of data of academic year 2016/2017 produced cluster model  $c_0 = 71$ ,  $c_1 = 26$  and  $c_2 = 110$  while the sample of academic year 2015/2016 produced cluster model  $c_0 = 105$ ,  $c_1 = 71$  and  $c_2 = 26$ . The final centroid showed that in 2016/2017  $c_1$  was Excellent Cluster,  $c_0$  was Medium Cluster, and  $c_2$  was Lower Cluster while the final centroid in 2015/2016 showed that  $c_2$  was Excellent Cluster,  $c_1$  was Medium Cluster and  $c_0$  was Lower Cluster. In 2016/2017 in the Excellent Cluster there were 26 students, which are occupied only one class that was class IX A, Middle Cluster with a total of 71 students divided into 2 classes, they were IX B and IX C. Each class consisted of 32 students. There were 7 students remaining from Middle Cluster determined based on the closest distance to centroid of Lower Cluster, then placed in class IX D, IX E, IX F and IX G along with members of the Lower Cluster.*

**Keywords:** *K-Means, Clustering, The Student Class Distribution.*

# **BAB I**

## **PENDAHULUAN**

### **A. Latar Belakang**

SMP Negeri 10 Magelang yang berlokasi di Jl. Sukarno Hatta no 2 Magelang, pada awal tahun 2017 memiliki siswa sebanyak 628 siswa yang terbagi dalam 21 rombongan belajar (Rombel). Salah satu program yang diagendakan di SMP Negeri 10 Magelang adalah program bimbingan belajar (pemantapan) untuk siswa kelas IX yang akan diterapkan langsung pada rombongan belajar / kelas. Bimbingan belajar ini bertujuan untuk lebih memantapkan kesiapan siswa kelas IX dalam menghadapi Ujian Nasional Berbasis Komputer (UNBK). Sistem pembagian kelompok bimbingan belajar yang ada saat ini masih berdasarkan daftar absen kelas yang sudah ada yang disusun secara konvensional oleh guru BK.

Berdasarkan hasil wawancara, sistem pembagian kelompok rombongan belajar yang ada saat ini dirasa kurang efektif oleh pihak sekolah karena pembagian kelompok rombongan belajar masih berdasar sikap dan perilaku siswa. Hasil dari pengelompokan rombongan belajar siswa yang sudah dilakukan sebelumnya belum memenuhi kriteria yang diharapkan, yaitu belum membentuk rombel yang homogen. Masalah tersebut mengakibatkan sulitnya proses belajar mengajar dan penilaian siswa oleh guru.

Dengan mulai diterapkannya UNBK pada tahun ajaran ini, maka waktu yang digunakan untuk proses pembelajaran selama satu tahun ajaran menjadi semakin sempit. Waktu pembelajaran semakin berkurang untuk melaksanakan Simulasi UNBK tahap I sampai dengan tahap III dan Ujian Sekolah Berbasis Nasional. Sementara kegiatan Ujian Kertas dan Ujian Praktek tetap berjalan seperti biasa. Sehingga kendala yang muncul jika pembagian rombongan belajar tidak bersifat homogen adalah proses pembelajaran yang kurang efisien dari segi waktu.

Belum adanya sistem pendukung yang dapat membantu pembagian rombongan belajar yang homogen di SMP Negeri 10 Magelang, mengakibatkan proses pengelompokan rombongan belajar tidak seperti yang



diharapkan. Banyaknya nilai yang digunakan sebagai nilai pembanding untuk mengelompokkan siswa, membuat pemecahan masalah ini semakin kurang *valid*.

Untuk mewujudkan hal tersebut, maka SMP Negeri 10 Magelang memerlukan informasi baru yang terkait data-data siswa yang sudah ada untuk dievaluasi bersama oleh guru dan wali kelas guna mendukung sistem pembagian kelas homogen. Sehingga kebijakan yang diambil untuk mengelompokkan dan membentuk rombongan belajar yang baru dapat dilakukan tepat sasaran dan efektif. Untuk mengatasi masalah yang tengah dihadapi tersebut, maka akan dilakukan penelitian dengan menerapkan metode *clustering* Algoritma *K-Means* untuk menganalisa data siswa SMP Negeri 10 Magelang. *K-Means* merupakan metode data *clustering* non hirarki yang berusaha mengelompokkan data ke dalam satu atau lebih *cluster* atau kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. *K-Means* memiliki kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat.

## **B. Rumusan Masalah**

Berdasarkan latar belakang masalah diatas, maka permasalahan yang dapat dirumuskan adalah bagaimana menerapkan *clustering* dengan metode algoritma *K-Means* yang dapat menghasilkan informasi mengenai pengelompokan data siswa berdasarkan nilai mata pelajaran, sehingga dapat ditentukan kelompok unggulan, kelompok sedang dan kelompok bawah dalam rangka pembentukan rombongan belajar siswa.

## **C. Tujuan Penelitian**

Adapun tujuan dari penelitian ini adalah menerapkan metode *clustering* dengan algoritma *K-Means* untuk pengelompokan data siswa kelas IX di SMP Negeri 10 Magelang sehingga dapat ditentukan kelompok unggulan, kelompok sedang dan kelompok bawah, guna menunjang proses pembagian rombongan belajar siswa.

#### **D. Manfaat Penelitian**

Manfaat yang diperoleh dari hasil penelitian ini adalah:

1. Mengetahui karakteristik kecenderungan hasil belajar siswa yang dihasilkan dari implementasi dengan algoritma *K-Means clustering*, sehingga dapat digunakan untuk menentukan pembagian rombongan belajar dengan mudah.
2. Memberikan informasi sebagai dasar pertimbangan pengambilan kebijakan dalam melakukan tindak lanjut pembelajaran siswa.
3. Memberikan wawasan ilmu pengetahuan tentang konsep data mining dengan teknik *clustering*, khususnya algoritma *K-Means*.

## BAB II

### TINJAUAN PUSTAKA

#### A. Penelitian Relevan

1. Penelitian yang dilakukan oleh Fauziah Nur, M. Zarlis, dan Benny Benyamin Nasution (2017) yang berjudul Penerapan Algoritma *K-Means* Pada Siswa Baru Sekolah Menengah Kejuruan Untuk *Clustering* Jurusan, menyatakan bahwa pada pengujian yang dilakukan dalam penelitian tersebut, proses *clustering* data siswa baru SMK terjadi sebanyak 3 kali iterasi. Kemudian berdasarkan hasil *cluster* dengan menerapkan beberapa kriteria dari calon siswa menggunakan *K-Means* dapat diambil pengelompokan dengan rata-rata jurusan yang diambil adalah rekayasa perangkat lunak dan sedikit jumlah siswa yang tidak lulus. Bahkan ada beberapa jurusan yang tidak dibuka dikarenakan kriteria-kriteria siswa tidak dapat lulus dalam jurusan tersebut.

Kekurangan yang terdapat pada penelitian ini adalah bahwa terdapat ketidaksesuaian antara hasil *cluster* dengan kriteria atribut data. Dimana hasil *clustering* menunjukkan kelompok tidak lulus, rekayasa perangkat lunak dan teknik komputer jaringan, sedangkan pada atribut data tidak berhubungan dengan ketiganya yaitu penghasilan orang tua, tanggungan anak, dan hasil tes. Sedangkan kelebihan dari penelitian ini adalah diterapkan cara untuk menentukan banyaknya *cluster* yaitu dengan menghitung akar dari setengah jumlah data yang ada.

2. Penelitian yang dilakukan oleh Ari Muzakir (2014) yang berjudul Analisa Dan Pemanfaatan Algoritma *K-Means Clustering* Pada Data Nilai Siswa Sebagai Penentuan Penerima Beasiswa, menyatakan bahwa Algoritma *K-Means* dapat melakukan pengelompokan dokumen dalam jumlah yang banyak akan tetapi belum efisien dalam mengelompokan dokumen secara tepat. Penentuan *centroid* (titik pusat) pada tahap awal Algoritma *K-Means* sangat berpengaruh pada hasil *cluster* seperti pada hasil pengujian yang dilakukan dengan menggunakan 100 *record* dengan *centroid* yang berbeda menghasilkan hasil *cluster* yang berbeda juga.

Kekurangan yang terdapat pada penelitian ini adalah penentuan banyaknya iterasi *K-Means* yang dibatasi hanya 3 kali iterasi saja, sehingga apabila terdapat data yang sangat banyak dan centroid awal ditentukan secara acak maka akan menghasilkan anggota *cluster* yang jauh berbeda. Sedangkan kelebihan dari penelitian ini yaitu telah menggunakan alur algoritma yang jelas.

3. Penelitian yang dilakukan oleh Asroni dan Ronald Adrian (2015) yang berjudul Penerapan Metode *K-Means* Untuk *Clustering* Mahasiswa Berdasarkan Nilai Akademik Dengan *Weka Interface* Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang, menyatakan bahwa algoritma *K-Means* bisa digunakan untuk mengelompokkan mahasiswa berdasarkan IPK dan beberapa atribut mata kuliah.

Kekurangan yang terdapat pada penelitian ini adalah alur proses sistem yang kurang jelas dan kurangnya penjelasan tahap demi tahap mengenai algoritma *K-Means*. Sedangkan kelebihan dari penelitian ini yaitu adanya hitungan persentase hasil *clustering* sehingga data dapat diketahui penyebarannya dengan jelas.

Yang menjadi pembeda antara penelitian relevan dengan penelitian yang sedang dilakukan adalah:

1. Obyek penelitian yang sedang dilakukan merupakan data nilai siswa SMP Negeri 10 Magelang.
2. Tujuan dari penelitian ini adalah pengelompokan data yang dilakukan untuk menunjang sistem pembagian rombongan belajar, sedangkan pada penelitian relevan memiliki tujuan sebagai penentu beasiswa, penentu peserta lomba, dan untuk pengelompokan jurusan.

Beberapa kelebihan pada penelitian yang sedang dilakukan adalah:

1. Dilihat dari tujuan yang sudah disebutkan maka penelitian relevan cenderung memiliki manfaat yang hanya dirasakan oleh sebagian obyek *cluster* saja. Sedangkan pada penelitian yang sedang dilakukan memiliki manfaat yang dapat dirasakan oleh seluruh obyek cluster yaitu adanya kemudahan selama proses pembelajaran hingga menghadapi Ujian

Nasional. Bahkan manfaatnya juga dapat dirasakan oleh obyek di luar cluster yaitu guru pengampu mata pelajaran.

2. Hasil klasterisasi masih sangat berhubungan dengan atribut kriteria yang *dicluster*.
3. Jumlah iterasi tidak dibatasi dan penentuan *centroid* awal berdasarkan pembagian rata terhadap jumlah *cluster* agar menghasilkan *cluster* yang optimal.

## **B. Penjelasan Secara Teoritis Masing-masing Variabel Penelitian**

### **1. Siswa**

Dalam Undang-undang Pendidikan no.2 Tahun 1989, siswa atau murid juga disebut dengan peserta didik. Dalam hal ini siswa dilihat sebagai seseorang (subjek didik), yang mana nilai kemanusiaan sebagai individu, sebagai makhluk sosial yang mempunyai identitas moral, harus dikembangkan untuk mencapai tingkatan optimal dan kriteria kehidupan sebagai manusia warga negara yang diharapkan.

Siswa adalah setiap orang yang secara resmi terdaftar untuk mengikuti pelajaran di dunia pendidikan (Sarwono, 2007). Dari pendapat tersebut bisa dijelaskan bahwa siswa adalah status yang disandang oleh seseorang karena hubungannya dengan dunia pendidikan yang diharapkan menjadi calon-calon intelektual untuk menjadi generasi penerus bangsa.

Siswa yang menjadi obyek penelitian ini adalah siswa SMP Negeri 10 Magelang kelas VIII yang naik ke kelas IX pada tahun 2016/2017 sejumlah 207 siswa.

### **2. Rombongan Belajar (Rombel)**

Peraturan Menteri Pendidikan Nasional Republik Indonesia no. 24 Tahun 2007 menyatakan bahwa rombongan belajar adalah kelompok peserta didik yang terdaftar pada satu satuan kelas. Dimana dalam rombongan belajar ini bisa dilakukan pada pembagian kelas di dalam satu sekolah. Jadi terdapat kelas unggulan dan ada pula kelas reguler di dalam satu sekolah.

### 3. Nilai Mata Pelajaran / Prestasi Belajar

Prestasi belajar atau hasil belajar adalah realisasi atau pemekaran dari kecakapan potensial atau kapasitas yang dimiliki seseorang. Penguasaan prestasi bila dilihat dari perilakunya, baik perilaku dalam bentuk penguasaan pengetahuan, ketrampilan berpikir maupun kemampuan motorik (Sukmadinata, 2005). Prestasi belajar pada dunia pendidikan adalah hasil pencapaian seseorang selama mengikuti pelajaran di sekolah yang berbentuk skor atau nilai (Sukmana, 2004).

### 4. *Data Mining*

Data mining adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya namun dapat dipahami dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. Data mining biasa juga disebut dengan “*Data knowledge discovery*” atau menemukan pola tersembunyi pada data. Data mining adalah proses dari menganalisa data dari prespektif yang berbeda dan menyimpulkannya ke dalam informasi yang berguna (Prasetyo, 2013).

Pada prosesnya data mining akan mengekstrak informasi yang berharga dengan cara menganalisis adanya pola-pola ataupun hubungan keterkaitan tertentu dari data-data yang berukuran besar. Data mining berkaitan dengan bidang ilmu-ilmu lain, seperti *Database System*, *Data Warehousing*, *Statistic*, *Machine Learning*, *Information Retrieval*, dan Komputasi Tingkat Tinggi. Selain itu data mining didukung oleh ilmu lain seperti *Neural Network*, Pengenalan Pola, *Spatial Data Analysis*, *Image Database*, dan *Signal Processing*.

Beberapa survei tentang proses pemodelan dan metodologi menyatakan bahwa, “Data mining digunakan sebagai penunjuk, dimana data mining menyajikan intisari atas sejarah, deskripsi dan sebagai standar petunjuk mengenai masa depan dari sebuah proses model data mining”.

a. Karakteristik data mining

- 1) Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- 2) Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dapat dipercaya.
- 3) Data mining berguna untuk membuat keputusan kritis.

b. Fungsi Data Mining

Teknik-teknik data mining telah digunakan untuk menemukan pola yang tersembunyi dan memprediksi tren masa depan. Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan yaitu :

1) Deskripsi

Model prediksi berkaitan dengan pembuatan sebuah model yang dapat melakukan pemetaan dari setiap himpunan variabel kesetiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang didapat. Deskripsi dari pola dan kecenderungan yang terdapat dalam data sering memberikana kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2) Estimasi

Memperkirakan suatu hal dari sejumlah sample yang dimiliki. Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik dari pada kearah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

### 3) Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, namun perbedaannya adalah nilai dalam hasil prediksi akan terjadi dimasa yang akan datang (memperkirakan hal yang belum terjadi).

### 4) Klasifikasi

Kegiatan menggolongkan, dengan menggunakan data historis. Klasifikasi digunakan untuk menempatkan bagian yang tidak diketahui pada data ke  $n$  dalam kelompok yang sudah diketahui. Klasifikasi menggunakan variabel target dengan nilai nominal. Dalam satu set pelatihan, variabel target sudah diketahui. Dengan pembelajaran dapat ditemukan hubungan antara fitur dengan variabel target. Ada dua langkah dalam proses klasifikasi (Han and Kamber, 2006):

- a) Pembelajaran (*learning*): pelatihan data dianalisis oleh algoritma klasifikasi.
- b) Klasifikasi: data yang diujikan digunakan untuk mengkalkulasi akurasi dari aturan klasifikasi. Jika akurasi dianggap dapat diterima, aturan dapat diterapkan pada klasifikasi data *tuple* yang baru.

### 5) Klasterisasi

Klasterisasi merupakan pengelompokan data, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan record yang memiliki kemiripan satu dengan yang lainya dan memiliki ketidak miripan dengan record-record antar *cluster*. Klasterisasi berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklasteran. Pengklasteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai variabel target. Akan tetapi, algoritma pengklasteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan.



## 6) Asosiasi

Asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Asosiasi digunakan untuk menemukan pola yang menggambarkan kekuatan hubungan fitur dalam data. Pola yang ditemukan biasanya mempresentasikan bentuk aturan implikasi atau subset fitur. Tujuannya adalah untuk menemukan pola yang menarik dengan cara yang efisien.

## c. Arsitektur Data Mining

Data mining merupakan proses pencarian pengetahuan yang menarik dari data berukuran besar yang disimpan dalam basis data, data warehouse atau tempat penyimpanan informasi lainnya. Dengan demikian arsitektur sistem data mining memiliki komponen-komponen utama yaitu:

- 1) *Database, data warehouse, World Wide Web*, atau tempat penyimpanan informasi lainnya: bisa berbentuk satu atau banyak database, *data warehouse, spreadsheet*, ataupun tempat penyimpanan informasi lainnya. *Data Cleaning, Data Integration* dan *Data Selection* dapat dijalankan pada data tersebut. *Database* dan *data warehouse server*, komponen ini bertanggung jawab dalam pengambilan data yang relevan, berdasarkan permintaan pengguna.
- 2) *Knowledge Based*, komponen ini merupakan *domain knowledge* yang digunakan untuk memandu pencarian atau mengevaluasi pola-pola yang dihasilkan. Pengetahuan tersebut meliputi hirarki konsep yang digunakan untuk mengorganisasikan atribut atau nilai atribut kedalam level abstraksi yang berbeda. Pengetahuan tersebut juga dapat berupa kepercayaan pengguna (*user belief*), yang dapat digunakan untuk menentukan kemenarikan pola yang diperoleh.
- 3) *Data mining engine*. Bagian ini merupakan komponen penting dalam arsitektur sistem data mining. Komponen ini terdiri dari modul-modul fungsional seperti karakterisasi, asosiasi, klasifikasi, dan analisis cluster.

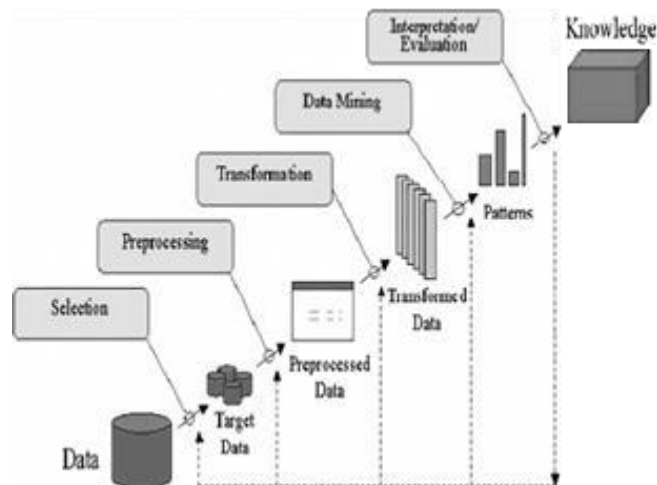
4) *Graphical user interface* (GUI). Modul ini berkomunikasi dengan pengguna dan data mining. Melalui komponen ini, pengguna berinteraksi dengan sistem menggunakan query.

d. Klasifikasi Sistem Data Mining

Data Mining merupakan suatu pendekatan dalam pemecahan masalah dengan menggunakan tinjauan berbagai sudut pandang ilmu secara terpadu yaitu, *database system*, *statistics*, *machine learning*, *visualization*, dan *information system*.

e. Tahap-Tahap Data mining

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap yang bersifat interaktif. Pengguna terlibat langsung atau dengan perantaraan knowledge base.



Gambar 2.1. Tahap-tahap data mining

Sumber : (Jurnal STMIK AMIKOM, 2015)

Dalam penyusunannya data mining ada beberapa tahapan yaitu :

1) Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama,

jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya.

2) Seleksi Data (*Data Selection*)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analysis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

3) Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

4) Transformasi data (*Data Transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5) Proses mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6) Evaluasi pola (*pattern evaluation*)

Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

7) Presentasi pengetahuan (*knowledge presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining. (Han, 2006)

## 5. Algoritma

Algoritma adalah urutan langkah-langkah logis penyelesaian masalah yang disusun secara sistematis dan logis. Kata Logis merupakan kata kunci dalam Algoritma. Langkah-langkah dalam Algoritma harus logis dan harus dapat ditentukan bernilai salah atau benar (Rosa dan Shalahuddin 2010).

Kriteria Algoritma yang baik :

- a. Tepat, benar, sederhana, standar dan efektif.
- b. Logis, terstruktur dan sistematis.
- c. Semua operasi terdefinisi.
- d. Semua proses harus berakhir setelah sejumlah langkah dilakukan.
- e. Ditulis dengan bahasa yang standar dengan format pemrograman agar mudah untuk diimplementasikan dan tidak menimbulkan arti ganda.

Algoritma adalah jantung ilmu komputer atau informatika. Banyak cabang ilmu komputer yang diacu dalam terminologi algoritma. Namun algoritma tidak selalu identik dengan ilmu komputer saja. Dalam kehidupan sehari-hari pun banyak terdapat proses yang dinyatakan dalam suatu algoritma (Alex Budiyanto, 20003).

Secara umum, pihak (benda) yang mengerjakan proses disebut pemroses (*processor*). Pemroses tersebut dapat berupa manusia, komputer, robot atau alat-alat elektronik lainnya. Pemroses melakukan suatu proses dengan melaksanakan atau mengeksekusi algoritma yang menjabarkan proses tersebut (Munir, 2007).

## 6. *K-Means*

*K-Means* merupakan salah satu metode *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster dalam jumlah sampel data yang besar. Metode ini mempartisi data ke dalam cluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang

mempunyai karakteristik yang berbeda di kelompokkan ke dalam *cluster* yang lain (Prasetyo, 2013).

Algoritma K-Means merupakan algoritma klasterisasi yang mengelompokkan data berdasarkan titik pusat klaster (*centroid*) terdekat dengan data. Tujuan dari K-Means adalah pengelompokkan data dengan memaksimalkan kemiripan data dalam satu klaster dan meminimalkan kemiripan data antar klaster. Ukuran kemiripan yang digunakan dalam klaster adalah fungsi jarak. Sehingga pemaksimalan kemiripan data didapatkan berdasarkan jarak terpendek antara data terhadap titik *centroid* (Asroni, 2015).

Algoritma K-Means sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaannya dalam praktek (Fauziah, 2017).

Pembentukan rombongan belajar / kelas siswa yang homogen adalah kunci dari dilakukan penelitian ini, sehingga penerapan *K-Means clustering* ini dianggap sudah tepat.

Secara umum algoritma dasar dari *K-Means Clustering* adalah sebagai berikut :

a. Tentukan jumlah *cluster*

Untuk menentukan banyaknya cluster k dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak cluster.

b. Tentukan *centroid*/ titik pusat setiap *cluster*. Untuk menentukan *centroid* awal dilakukan dengan cara pemilihan data secara acak.

c. Hitung nilai jarak antara masing-masing data dengan *centroid*. Adapun persamaan yang dapat digunakan salah satunya yaitu *Euclidean Distance Space*. *Euclidean distance space* sering digunakan dalam perhitungan jarak, hal ini dikarenakan hasil yang diperoleh merupakan jarak terpendek antara dua titik yang diperhitungkan. Adapun persamaannya adalah sebagai berikut.

$$d_{ij} = \sqrt{\sum_{k=1}^p \{x_{ik} - x_{jk}\}^2} \dots\dots\dots(2.1)$$

Dimana:

$d_{ij}$  = Jarak objek antara objek i dan j

P = Dimensi data

$X_{ik}$  = Koordinat dari obyek i pada dimensi k

$X_{jk}$  = Koordinat dari obyek j pada dimensi k

- d. Alokasikan masing-masing data ke *centroid*/ rata-rata terdekat, sehingga terbentuk pengelompokan data awal.
- e. Tentukan nilai *centroid* baru dan bandingkan dengan nilai *centroid* yang sudah ada sebelumnya. Jika terjadi perubahan nilai *centroid* maka proses akan kembali ke langkah c. Jika tidak terjadi perubahan nilai *centroid* maka hentikan prosesnya.

Pembaharuan suatu titik *centroid* dapat dilakukan dengan menggunakan rumus berikut:

$$\frac{\sum}{n} \dots\dots\dots(2.2)$$

Di mana:

$m_i$  = titik *centroid* dari *cluster* ke-i

c = *cluster*

n = banyaknya data pada *cluster*

## 7. *Rapid Miner*

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 kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi (Dennis, 2013).

Alasan penelitian ini menggunakan *Rapid Miner* sebagai tool pengolah data adalah karena *rapid miner* memiliki banyak operator data mining, berbasis GUI dan dapat dijalankan di banyak sistem operasi seperti windows dan linux.

### C. Landasan Teori

1. Tujuan utama analisis *cluster* adalah mengelompokkan obyek-obyek data berdasarkan kemiripan karakteristik. Pada penelitian ini juga akan dilakukan *cluster* data siswa untuk menentukan kelas unggulan berdasarkan hasil belajar siswa.
2. Pada penelitian ini akan dibahas proses klasterisasi data siswa untuk menggunakan algoritma K-Means dimana proses klasterisasi data pada metode K-Means dilakukan pembentukan partisi sedemikian rupa sehingga setiap objek berada tepat pada satu partisi. Sehingga status keanggotaan setiap data sangat jelas terhadap satu *cluster* saja.
3. *K-Means* merupakan salah satu metode *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster dalam jumlah sampel data yang besar. Metode ini akan digunakan pada penelitian yang akan dilakukan dengan jumlah data sebanyak 207 dan 202 data.
4. Metode *K-Means* termasuk ke dalam metode *supervised clustering* dimana jumlah pusat *cluster* ditentukan terlebih dahulu dalam proses *clustering*.



# **BAB III**

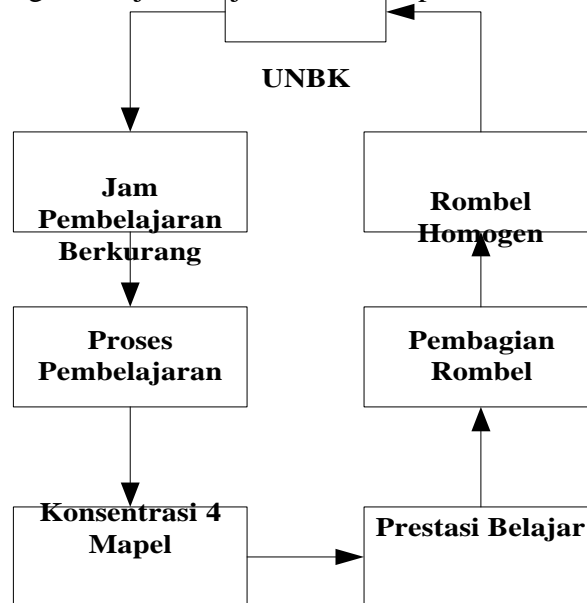
## **ANALISIS DAN PERANCANGAN SISTEM**

### **A. Analisis Sistem**

#### **1. Analisis Masalah**

- a. Rombongan belajar siswa yang homogen adalah kelompok pembelajaran yang terdiri dari kumpulan siswa yang memiliki kesamaan/ kemiripan kriteria. Dalam hal ini kriteria yang diambil adalah nilai dari empat mata pelajaran yang diujikan dalam Ujian Nasional Berbasis Komputer (UNBK), yaitu nilai mata pelajaran Bahasa Indonesia, Bahasa Inggris, Matematika dan Ilmu Pengetahuan Alam. Dengan dilaksanakannya UNBK di SMP Negeri 10 Magelang, maka waktu yang dibutuhkan untuk proses pembelajaran kelas IX selama satu tahun ajaran menjadi semakin menyempit atau berkurang. Sehingga proses pembelajaran perlu dilakukan dalam waktu yang seefisien mungkin. Oleh karena itu rombongan belajar yang homogen perlu diwujudkan dalam proses pembelajaran karena dapat mempermudah proses pembelajaran itu sendiri.
- b. SMP Negeri 10 magelang memiliki 21 rombongan belajar dimana setiap tingkat kelasnya ada 7 rombongan belajar. Jumlah siswa kelas VIII yang naik ke kelas IX setiap tahunnya kurang lebih mencapai 200 siswa sejumlah dengan siswa yang harus dipersiapkan menghadapi UNBK. Hal ini menjadi tantangan bagi petugas pembagi rombongan belajar yang harus mengolah banyaknya jumlah data siswa untuk disebar ke dalam 7 kelompok rombongan belajar berdasar nilai hasil belajar kelas VIII khususnya nilai mata pelajaran yang diujikan dalam UNBK.
- c. Tersedianya data berupa nilai hasil belajar siswa kelas VIII sangat mendukung untuk diolah dalam pengelompokan siswa membentuk rombel yang diharapkan. Namun data nilai tersebut menunjukkan bahwa standar penilaian yang digunakan adalah penilaian skala 100. Hal ini menjadikan data nilai setiap siswa dan setiap mata pelajaran

menjadi sangat bervariasi sehingga kendala pada pembagian rombongan belajar menjadi lebih kompleks.



Gambar 3.1. Diagram Alur Permasalahan

## 2. Analisis Data

Data yang digunakan sebagai objek penelitian ini adalah data nilai hasil belajar siswa kelas VIII SMP Negeri 10 Magelang pada tahun ajaran 2016/2017. Dengan total data sebanyak 207 siswa dengan rincian sebagai berikut:

Tabel 3.1. Data Sampel Penelitian Tahun 2016/2017

No.	Kelas	Jumlah (siswa)
1.	VIII A	29
2.	VIII B	30
3.	VIII C	30
4.	VIII D	28
5.	VIII E	28
6.	VIII F	31
7.	VIII G	31
	<b>Jumlah</b>	<b>207</b>

Kemudian sebagai pembandingan data pada penelitian ini juga akan di gunakan data sampel nilai siswa kelas VIII yang naik ke kelas IX pada tahun ajaran 2015/2016. Dengan jumlah data sebagai berikut:

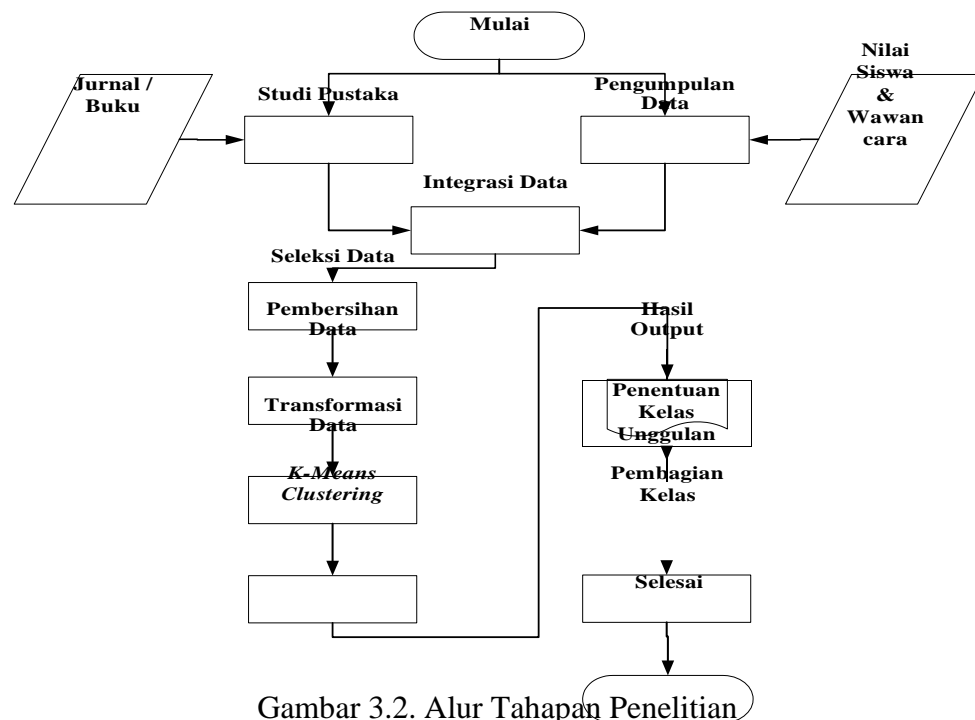
Tabel 3.2. Data Sampel Penelitian Tahun 2015/2016

No.	Kelas	Jumlah (siswa)
1.	VIII A	29
2.	VIII B	29
3.	VIII C	29
4.	VIII D	30
5.	VIII E	28
6.	VIII F	29
7.	VIII G	28
	<b>Jumlah</b>	<b>202</b>

Variabel yang digunakan untuk proses *clustering* adalah jenis kelamin siswa, data nilai mata pelajaran Bahasa Indonesia, Bahasa Inggris, Matematika dan Ilmu Pengetahuan Alam. Proses *clustering* data dilakukan untuk melihat data kecenderungan siswa terhadap mata pelajaran yang diujikan dalam UNBK guna menunjang sistem pembagian rombongan belajar dan sistem pembelajaran kelas IX.

## B. Perancangan Sistem

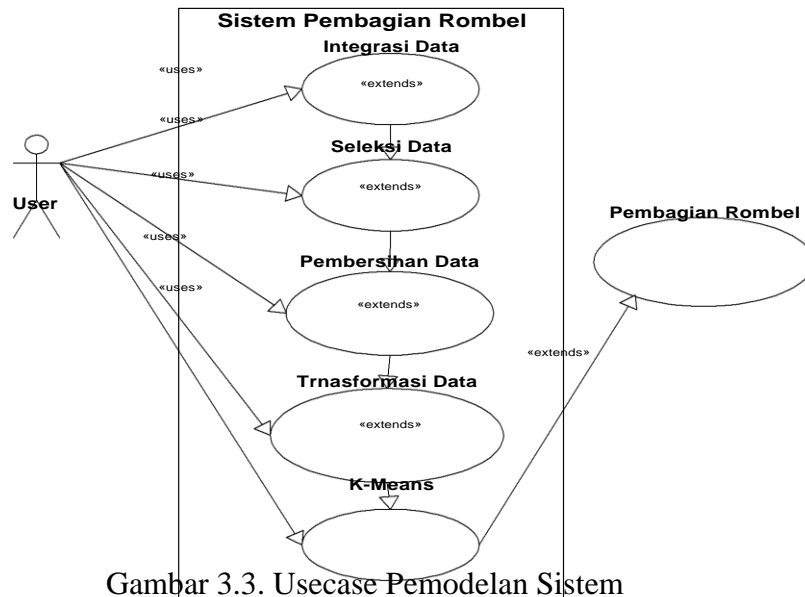
Tahapan penelitian dalam perancangan sistem dapat dilihat pada gambar berikut:



Gambar 3.2. Alur Tahapan Penelitian

## 1. Pemodelan Sistem Menggunakan Use Case Diagram

Berikut ini adalah gambar pemodelan sistem pembagian rombongan belajar yang akan dibuat dengan menggunakan *use case* diagram.



Gambar 3.3. Usecase Pemodelan Sistem

Beberapa proses yang terdapat dalam alur sistem pembagian rombongan belajar tersebut meliputi:

### a) Integrasi Data

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya.

### b) Seleksi Data

Data yang tersedia dalam database sangat banyak dan tidak semuanya dibutuhkan dalam penelitian ini, oleh karena itu data tersebut perlu diseleksi yang sesuai dengan yang akan dianalisis. Data tersebut meliputi Nomor Induk Siswa, Nama Siswa, Jenis Kelamin dan Nilai hasil belajar siswa kelas VIII semester genap (penegas) tahun ajaran 2016/2017 yang terdiri dari mata pelajaran Bahasa Indonesia, Bahasa Inggris, Matematika dan Ilmu Pengetahuan Alam.

### c) Pembersihan Data

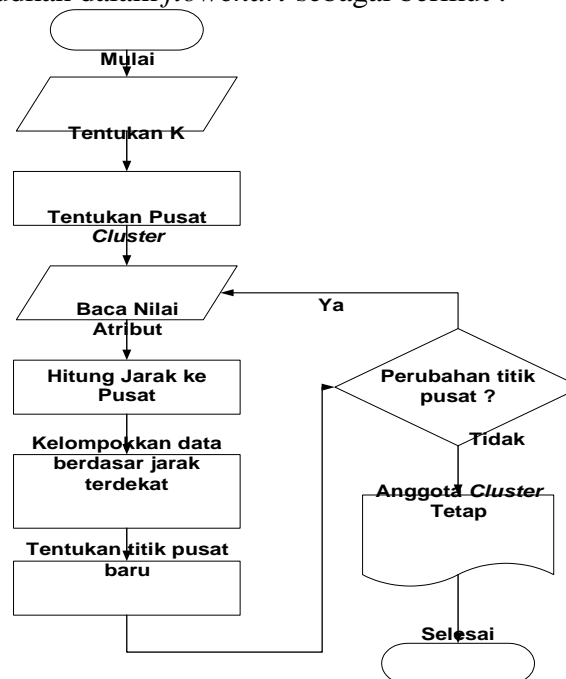
Proses pembersihan data yaitu membersihkan data dari *noise* atau *missing value* yang merupakan data yang tidak konsisten atau tidak relevan. Dalam penelitian ini *missing value data* akan diperlakukan dengan memasukan nilai *mean* pada masing-masing variabel. Sehingga data tersebut menjadi relevan untuk digunakan sebagai objek penelitian sehingga tidak mempengaruhi performa dalam proses pengelompokan data menggunakan algoritma *K-Means*.

### d) Transformasi Data

Setelah data diseleksi maka tahap selanjutnya adalah transformasi data, yaitu proses perubahan tipe data menjadi tipe numeris/ angka agar data dapat diolah menggunakan algoritma *K-Means*. Pada penelitian ini data yang akan diubah tipe datanya adalah atribut Jenis Kelamin dimana jenis kelamin laki-laki akan diubah menjadi 0 dan jenis kelamin perempuan diubah menjadi 1. Dengan asumsi bahwa siswa perempuan lebih unggul prestasi belajarnya dibanding siswa laki-laki.

### e) Proses *Clustering*

Adapun rancangan model tahapan proses *clustering* pengelompokan data siswa menggunakan algoritma *K-Means* diwujudkan dalam *flowchart* sebagai berikut :



Gambar 3.4. Alur Algoritma K-Means

Keterangan *Flowchart* Algoritma K-Means :

- 1) Menentukan jumlah kluster yang akan dibuat. Pada penelitian ini akan ditentukan jumlah *cluster* sebanyak 3 *cluster* untuk menentukan kelompok siswa unggulan, sedang dan bawah.
- 2) Menentukan *centroid* atau titik pusat dari setiap *cluster* yang dilakukan secara acak. *Centroid* pada penelitian ini adalah seperti terlihat dalam tabel berikut ini :

Tabel 3.3. Menentukan nilai *centroid*

NIS	NAMA	JK	B. Ind	B. Ing	Mat	IPA
5776	BAGAS REZKY P	0	79	75	75	75
5781	DEVINA PUTRI	1	75	80	73	75
5752	INDRA DEFANTY	0	79	80	85	78

- 3) Hitung jarak setiap data ke *centroid* atau pusat kluster dengan menggunakan persamaan *euclidean distance* yaitu sebagai berikut:

Tabel 3.4. Sampel 15 data pertama

NIS	NAMA	JK	B. Ind	B. Ing	Mat	IPA
5680	ACHMAD FAUZAN	0	79	75	80	75
5685	AMANDA NABILA AJI	1	83	80	76	76
5652	ANISSA DESY	1	79	80	80	76
5776	BAGAS REZKY	0	79	75	75	75
5743	BUNGA SYALUM	1	82	80	78	76
5656	CANDRA DEWI	1	81	80	80	75
5808	CHANDRA IZA RASYAD	0	84	75	82	76
5781	DEVINA PUTRI	1	75	80	73	75
5720	FARIS TUNGGUL	0	80	75	70	75
5841	FATIHA TUL MAULIDA	1	79	80	75	76
5814	FERNANDA RIZKY	0	74	75	72	75
5752	INDRA DEFANTY	0	79	80	85	78
5724	IRVAN AGUNG	0	77	75	70	75
5817	JUNDA AYU	1	76	80	75	74
5665	KARUNIA INDAH	1	78	78	78	75

Iterasi 1 : Berdasarkan data yang ada dapat di ambil satu contoh untuk proses perhitungan *cluster* (C), di mana :

$$K = 3$$

$$m_1 = (0, 79, 75, 75, 75)$$

$$m_2 = (1, 75, 80, 85, 78)$$

$$m_2 = (0, 79, 80, 85, 78)$$

maka :

Pencarian *cluster* untuk data pertama.

$$\begin{array}{l} \sqrt{(\quad) (\quad) (\quad) (\quad) (\quad)} \\ \sqrt{(\quad) (\quad) (\quad) (\quad) (\quad)} \\ \sqrt{(\quad) (\quad) (\quad) (\quad) (\quad)} \\ \sqrt{\quad} \end{array}$$

$$\begin{array}{l} \sqrt{(\quad) (\quad) (\quad) (\quad) (\quad)} \\ \sqrt{(\quad) (\quad) (\quad) (\quad) (\quad)} \\ \sqrt{(\quad) (\quad) (\quad) (\quad) (\quad)} \\ \sqrt{\quad} \end{array}$$

$$\begin{array}{l} \sqrt{(\quad) (\quad) (\quad) (\quad) (\quad)} \\ \sqrt{(\quad) (\quad) (\quad) (\quad) (\quad)} \\ \sqrt{(\quad) (\quad) (\quad) (\quad) (\quad)} \\ \sqrt{\quad} \end{array}$$

Cara di atas dilakukan untuk masing-masing data hingga data terakhir (*End Of File* / EOF), sehingga menghasilkan jarak data sebagai berikut.

Tabel 3.5. Jarak masing-masing data dengan centroid terdekat

<b>C1</b>	<b>C2</b>	<b>C3</b>
5	9,539392	7,681146
6,63325	8,602325	10,0995
7,211103	8,124038	5,477226
0	6,78233	11,57584
6,708204	8,660254	7,937254
7,416198	9,219544	6,244998
8,660254	13,74773	7,937254
6,78233	0	13,0384
5,09902	7,745967	16,12452
5,196152	4,582576	10,24695
5,830952	5,291503	15,09967
11,57584	13,0384	0
5,385165	6,244998	16,21727
6	2,44949	11,22497
4,472136	6,164414	8

- 4) Mengelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroid* atau pusat klasternya.

Tabel 3.6. Mengelompokkan data berdasar jarak terdekat

<b>NIS</b>	<b>C1</b>	<b>C2</b>	<b>C3</b>
5680	1		
5685	1		
5652			1
5776	1		
5743	1		
5656			1
5808			1
5781		1	
5720	1		
5841		1	
5814		1	
5752			1
5724	1		
5817		1	
5665	1		



- 5) Kemudian tentukan titik *centroid* baru menggunakan rumus (2.2). Perhitungan dilakukan dengan menghitung rata-rata nilai dari masing-masing anggota *cluster*, adapun hasilnya sebagai berikut:

Tabel 3.7. Centroid baru pada iterasi 1

	JK	B. Ind	B. Ing	Mat	IPA
Centroid1	0,428571	79,71429	76,85714	75,28571	75,28571
Centroid2	0,75	76	78,75	73,75	75
Centroid3	0,5	80,75	78,75	81,75	76,25

- 6) Setelah didapatkan nilai titik pusat terbaru. Bandingkan nilai titik pusat tersebut dengan titik pusat yang sudah ditentukan sebelumnya, jika terdapat perubahan nilai titik pusat maka proses kembali ke langkah nomor 3. Namun jika tidak ada perubahan nilai titik pusat maka proses berhenti atau selesai.

#### f) Menentukan Rombongan Belajar

##### 1) Evaluasi Hasil *Cluster*

Dengan berhentinya proses *clustering* maka akan terbentuk pengelompokan data siswa yang terdiri dari tiga kelompok dengan kriteria yang sudah ditentukan. Data *cluster* tersebut kemudian akan dievaluasi apakah data hasil *cluster* sudah sesuai atau belum dengan menampilkan jumlah anggota, nilai *centroid*, serta daftar anggota setiap *cluster* nya. Jika sudah sesuai maka data siswa sudah siap untuk digunakan dalam penentuan rombongan belajar dimana setiap kelompok tersebut memiliki anggota yang saling memiliki kemiripan kriteria. Dalam menentukan rombongan belajar, dari hasil *clustering* akan ditentukan terlebih dulu *cluster* mana yang merupakan kelompok unggulan, sedang dan bawah.

##### 2) Penentuan Kelompok Siswa Unggulan, Sedang dan Bawah

Penentuan kelompok unggulan dilakukan dengan cara menghitung nilai rata-rata *centroid* pada setiap *cluster* yang sudah terbentuk, kemudian rata-rata *centroid* tersebut diurutkan dari nilai terbesar ke nilai terkecil. Untuk *cluster* yang memiliki nilai rata-

rata *centroid* terbesar merupakan kelompok siswa dengan hasil prestasi unggulan.

Tabel 3.8. Rata-rata *centroid* kelompok unggulan

	JK	B. Ind	B. Ing	Mat	IPA	Rata-rata
Centroid1	0,428571	79,71429	76,85714	75,28571	75,28571	61,5142842
Centroid2	0,75	76	78,75	73,75	75	60,85
Centroid3	0,5	80,75	78,75	81,75	76,25	63,6

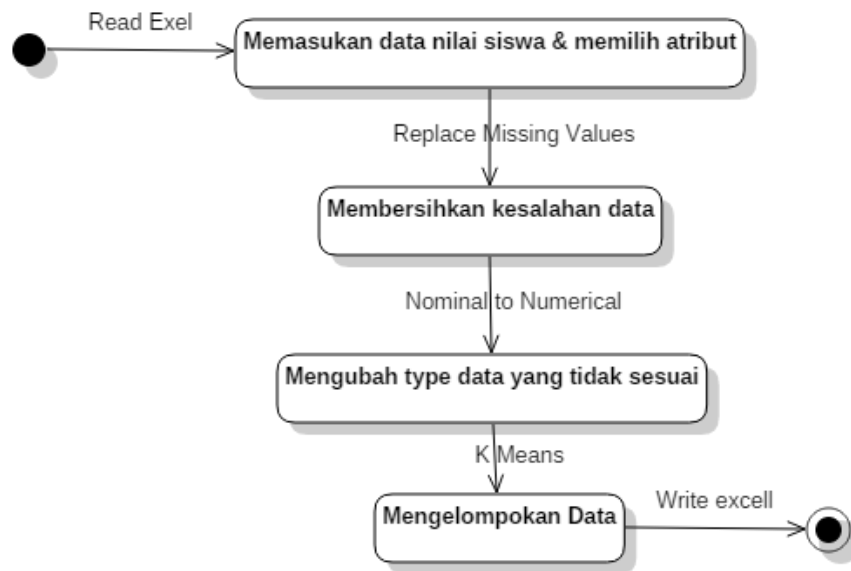
### 3) Pembagian Rombongan Belajar

Untuk penentuan jumlah siswa di dalam rombongan belajar terdapat beberapa aturan dari Dinas Pendidikan Kota Magelang yaitu jumlah minimal siswa pada satu rombongan belajar adalah sebanyak 20 siswa dan jumlah maksimalnya sebanyak 32 siswa. Pada penelitian ini diberlakukan kebijakan apabila hasil cluster tidak memungkinkan untuk dibentuk sesuai jumlah siswa yang ditentukan maka akan dibentuk maksimal 1 rombongan belajar campuran. Pembentukan rombel campuran akan melakukan pemindahan data tertentu dari satu *cluster* ke *cluster* lain yang memiliki jarak data dengan *centroid* terdekat pada *cluster* tujuan.

Tahapan pembagian rombongan belajar adalah sebagai berikut:

- (a) Jumlah anggota *cluster* dibagi 32 (jumlah maksimal)
- (b) Jika sisa bagi adalah antara 10 sampai 19, maka cluster tersebut kekurangan data sebanyak 20 dikurangi sisa bagi, kekurangan data diambil dari *cluster* berikutnya berdasarkan jarak terdekat dengan *centroid* tujuan
- (c) Jika sisa hasil bagi kurang dari 10, maka cluster tersebut kelebihan data sebanyak sisa bagi, kelebihan data dipindah ke *cluster* berikutnya berdasarkan jarak terdekat dengan *centroid* tujuan

## 2. State Diagram



Gambar 3.5. State Diagram Sistem

## 3. Implementasi Rapidminer

Untuk mengimplementasikan proses *K-Means Clustering* digunakan aplikasi Rapidminer untuk memudahkan proses simulasi perubahan data yang terjadi dalam mengelompokkan data siswa di SMP Negeri 10 Magelang. Seperti jumlah jumlah anggota *cluste*, nilai centroid, dan daftar data tiap *cluster* yang dihasilkan untuk dijadikan referensi validitas data yang akan diolah. Langkah-langkahnya dalam implementasi tersebut sebagai berikut :

### a) Input Data

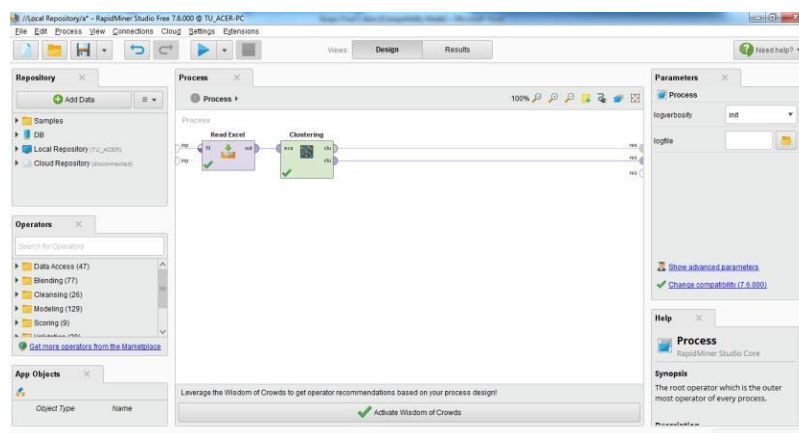
Input data dilakukan dengan cara meng-import data kesiswaan yang telah diseleksi dan dibuat dalam dokumen excel.



Gambar 3.6. Proses import data

## b) Proses Pengklasteran

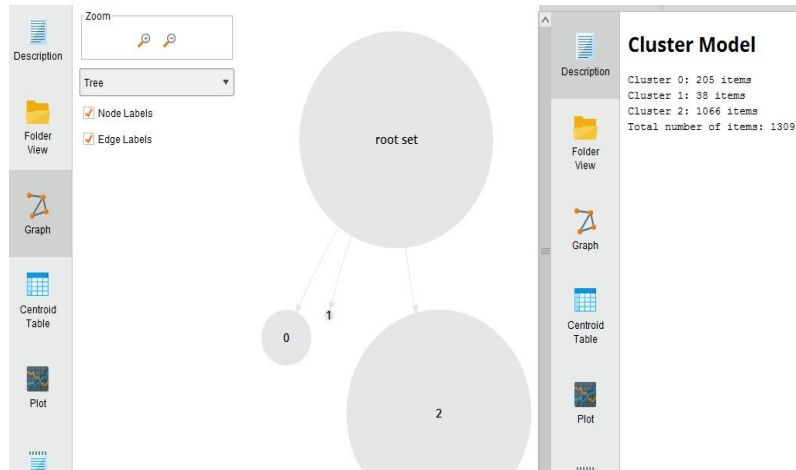
Dalam pemrosesan data diawali dengan memasukkan fungsi operator *K-Means* dan diatur jumlah kluster yang diinginkan, dalam kasus ini diatur 3 kluster. Hubungkan semua operator yang akan digunakan, kemudian jalankan proses.



Gambar 3.7. Menghubungkan antar operator

### c) Hasil Output

Hasil keluaran dari simulasi ini yaitu menunjukkan beberapa informasi yang meliputi jumlah klaster, jumlah anggota di setiap klaster, data mana saja yang menjadi anggota setiap klasternya serta titik pusat (*centroid*) setiap klasternya.



Gambar 3.8. Tampilan hasil output



# BAB VI

## PENUTUP

### A. Kesimpulan

Berdasarkan penelitian yang sudah dilakukan maka dapat disimpulkan bahwa:

1. Penerapan *clustering* menggunakan algoritma *K-Means* bisa untuk mengelompokkan data siswa dan mampu menentukan kelompok siswa unggulan, sedang dan kelompok bawah.
2. Hasil penghitungan *K-Means clustering* dapat diterapkan untuk pembagian rombel siswa dengan melakukan pendekatan jarak terdekat masing-masing data terhadap *centroid cluster* lain.
3. Data yang dihasilkan di masing-masing data sample hampir sama yaitu Cluster Unggul sebanyak 26 data dan Cluster Sedang sebanyak 71 data.
4. Cluster Bawah yang dihasilkan terdapat selisih data yaitu pada tahun ajaran 2016/2017 sebanyak 110 data sedangkan tahun ajaran 2015/2016 sebanyak 105 data.

### B. Saran

Berdasarkan hasil penelitian yang sudah dilakukan ada beberapa saran yang dapat diberikan untuk penelitian relevan yang selanjutnya, diantaranya sebagai berikut :

1. Penelitian agar menggunakan data sample tahun ajaran yang lebih dari 2 tahun agar menghasilkan *cluster* model yang lebih bervariasi yang memungkinkan muncul kasus baru dalam proses tahapan pembentukan rombel. Agar dapat mengembangkan aplikasi yang dapat mengolah data hasil *clustering* untuk menentukan jarak anggota *cluster* terhadap *centroid cluster* lain.
2. Perlu dilakukan penelitian relevan menggunakan metode lain yang lebih baik sebagai pembandingan data yang dihasilkan pada masing-masing metode penelitian.

## Daftar Pustaka

- Asroni. Adriaan, R., 2015. Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang. Yogyakarta: Jurnal Ilmiah Semesta Teknika. Vol. 18, No. 1, 76-82.
- Budiyanto. Alex., 2003. *Pengantar Algoritma dan Pemrograman*. Ilmu Komputer.Com.
- Ediyanto, Muhlasah N. M. dan Neva S., *Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis*. Jurnal Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster) Volume 02 , No. 2 (2013), hal 133 - 136.
- Fauzanu, A., Darwiyanto E., Agung Ary W.G., 2017. *Analisis Web Usage Mining Menggunakan Teknik K-Means Clustering Dan Association Rule*. -: e-Proceeding of Engineering. Vol. 4, No. 2
- Firmansyah, A., Irham Gufroni A., Nur Rachman A., 2013. *Data Mining Dengan Metode Clustering K-Mean Untuk Pengelompokan Mahasiswa Potensial Drop Out Pada Program Studi Teknik Informatika Universitas Siliwangi*. Tasikmalaya: Jurnal Teknik Informatika Universitas Siliwangi Tasikmalaya.
- Khobiri, A., 2016. *Aplikasi Sistem Pembagian Kelas Siswa Baru di MTs Raudlatut Thalabah Berdasarkan Hasil Nilai Akademis Siswa*. Kediri: Artikel Skripsi Universitas Nusantara PGRI Kediri.-
- Muhandis, R.N., 2016. *Pengaruh Prestasi Belajar ..., -* : Skripsi FAI.UMP.
- Muzakir, A., 2014. *Analisa Dan Pemanfaatan Algoritma K-Means Clustering Pada Data Nilai Siswa Sebagai Penentuan Penerima Beasiswa*. Yogyakarta: Prosiding Seminar Nasional Aplikasi Sains & Teknologi.
- Nur, F., Zarlis, M., dan Nasution, B.B., 2017. *Penerapan Algoritma K-Means Pada siswa Baru Sekolah Menengah Kejuruan Untuk Clustering Jurusan*. Medan: Jurnal Nasional dan Teknologi Jaringan. Vol. 1, No. 2.
- Nurhayadi. 2016. *Pembagian Kelas Homogen Menggunakan K-Means Clustering*. -: Jurnal Pendidikan Matematika. Vol. 5, No. 1.
- Peraturan Menteri Pendidikan Nasional Republik Indonesia No. 24 Tahun 2007 halaman 5.
- Rahmat, B., Agidatama Gafar A., Fajriani N., Ramdani U., Rihin Uyun F., Purnama Sari P.Y., Ransi N., 2017. *Implementasi K-Means Clustering Pada Rapidminer Untuk Analisis Daerah Rawan Kecelakaan*. Seminar Nasional Riset Kuantitatif Terapan 2017 Kendari:-.
- Rima D.R., 2014. *Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro*. Yogyakarta: Jurnal Universitas Dian Nuswantoro.



Siyamto, Y., 2017. *Pemanfaatan Data Mining Dengan Metode Clustering Untuk Evaluasi Biaya Dokumen Ekspor Di PT. Winstar*. Batam: Media Informatika Budidarma. Vol. 1, No. 2.

Valentino, F., Bharata Adji T., Erna Permanasari A., 2017. *Komparasi Metode Decision Tree dan K-Means Clustering Dalam Mengatasi Masalah Cold-start Pengguna Baru*. Yogyakarta: Jurnal Departemen Teknik Elektro dan Teknologi Informasi, FT UGM.

