



Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode *Clustering K-Means*

Erdi Amos Saputra¹, Yessica Nataliani²

Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Salatiga, Indonesia
E-mail: ¹682017091@student.uksw.edu; ²yessica.nataliani@uksw.edu

Abstrak

Di dunia pendidikan seringkali terjadi permasalahan bagaimana cara menentukan tingkat prestasi murid dengan kategori rendah, cukup, dan tinggi serta menemukan *top rank* murid unggulan dalam kelas. Dalam hal ini *data mining* dengan pendekatan *K-Means Clustering* dapat digunakan untuk mengelompokkan data menjadi kumpulan data. Dalam sistem analisis, pendekatan *data mining* berdasarkan algoritma *K-Means* dapat digunakan untuk pengelompokan prestasi murid. Dalam penelitian ini data nilai siswa kelas X-XII Bahasa SMAN 1 Tenganan tahun 2014-2017, dari semester satu sampai lima dikelompokkan berdasar nilai rapor. *Clustering* digunakan dalam pembangunan program analitik ini untuk menilai dampak data murid terhadap kecenderungan keberhasilan murid di setiap kelompok yang dapat dibuktikan dengan kelulusan murid yang menduduki *top rank* serta dari hasil wawancara guru pengajar maupun wali kelas serta data nilai yang diperoleh dari Dapodik. Hasil dari penelitian ini membuktikan bahwa teknik *clustering K-Means* dapat dimanfaatkan oleh pengajar untuk mengkategorikan murid berdasarkan nilai mata pelajaran dan absensi, serta menggunakannya untuk menganalisis prestasi murid dengan mengelompokkan dari kategori prestasi rendah, rata-rata, dan tinggi. Selanjutnya, dengan metode *Simple Additive Weighting (SAW)* dicari *top rank* dari *cluster* tinggi untuk menemukan murid unggulan dan membentuk kelas yang ideal untuk meningkatkan prestasi dan memotivasi murid.

Kata kunci : *Data Mining, Clustering K-Means, Data Prestasi Murid, SAW, Top Rank*

1. Pendahuluan

SMAN 1 Tenganan merupakan sekolah negeri yang berada di Kabupaten Semarang, tepatnya di Jl. Jend. Soemitro, Wedilelo Kulon, Karangduren Kecamatan Tenganan, Jawa Tengah, 50775. Kapasitas kelas di SMAN 1 Tenganan cukup besar, mempunyai ruang kelas dan fasilitas yang memadai, seperti gedung olahraga, laboratorium Komputer, Kimia, dan Biologi yang digunakan untuk murid kelas X-XII dengan tiga jurusan, yaitu MIPA, Sosial, dan Bahasa. Jumlah guru dan staf pun juga mencukupi sehingga suasana akademik tercipta dengan baik.



Tingginya prestasi murid dan minimnya angka murid yang tidak berprestasi mencerminkan keunggulan sektor pendidikan. Sektor Pendidikan kini diharapkan mampu bersaing dengan memanfaatkan kemajuan SI/TI, yang dapat menunjang peningkatan daya saing dan menunjang operasional sehari-hari serta pengambilan keputusan strategis. Secara umum, keberhasilan murid dievaluasi berdasarkan evaluasi pelajaran teoritis dan praktis, serta kehadiran dan ketidakhadiran murid selama di dalam kelas. Penilaian dibagi menjadi tiga kategori yaitu pengetahuan, bakat, dan sikap. Pengajar menilai semua murid yang mengikuti pelajaran yang diberikan guna mengevaluasi dan menganalisis prestasi belajar murid.

Jumlah data murid terus bertambah setiap tahunnya. Hal ini mengakibatkan akumulasi data yang belum diolah secara baik digunakan untuk mengungkap pengetahuan dan informasi baru dengan *pattern* yang terbentuk sebagai hasil akumulasi data. Jumlah pemrosesan data yang terus meningkat mengharuskan penggunaan strategi dan metode sehingga dapat ditransformasikan menjadi informasi dan pengetahuan yang dapat dimanfaatkan oleh pendidik dalam proses pembuatan kebijakan. Hal ini menjadi permasalahan bagi guru maupun wali kelas yaitu bagaimana cara menentukan tingkat prestasi murid yang rendah, cukup, dan tinggi serta menemukan *top rank* murid unggulan dalam kelas, agar dapat membentuk kelas yang ideal untuk meningkatkan prestasi maupun memotivasi murid.

Data mining merupakan teknik untuk mengungkap tautan, *pattern*, dan tren baru dengan memfilter kumpulan banyak data yang disimpan di penyimpanan dan menerapkan teknik pengenalan *pattern* seperti prosedur statistik dan matematik [1]. *K-Means* adalah algoritma data *mining* yang dapat digunakan untuk mengelompokkan/mengklasterkan data yang sangat besar atau bertumpuk yang dapat ditangani dengan salah satu dari beberapa cara, termasuk *clustering* [2].

Beberapa penelitian tentang penerapan pendekatan *Means Clustering* telah dilakukan untuk menganalisis prestasi belajar murid yang menghasilkan kelompok murid berprestasi. Pengelompokan tersebut didasarkan data nilai tes, Tes Tengah Semester (TTS), Tes Akhir Semester (TAS), maupun keaktifan absensi murid di beberapa sekolah [3,4,5,6]. Pendekatan *K-Means* digunakan untuk rekomendasi jurusan calon mahasiswa berdasarkan nilai untuk pengelompokan data pemilihan jurusan bagi Perguruan Tinggi [7,8].

Fokus pada penelitian ini adalah pada pengelompokan murid berdasarkan nilai mata pelajaran dan absensi menjadi kategori tinggi, cukup, dan rendah menggunakan metode *K-Means Clustering*. Selanjutnya dari *cluster* tinggi akan dicari siswa unggul menggunakan metode *Simple Additive Weighting* (SAW). Dalam penelitian ini akan dibahas empat bagian yaitu bagian 1 berisi pendahuluan yang menjelaskan mengenai objek penelitian, permasalahan, metode yang digunakan dan manfaat serta berisi penelitian-penelitian terdahulu yang mengangkat topik

clustering K-Means dan teori tentang metode *K-Means*. Pada bagian 2 berisi penjelasan metode yang digunakan dalam melakukan penelitian. Selanjutnya, bagian 3 berisi pembahasan mengenai hasil *clustering* prestasi para murid Kelas X-XII Bahasa SMAN 1 Tengar, dari semester satu sampai lima pada tahun ajaran 2014-2017, menggunakan algoritma *K-Means*. Pada bagian 4 berisi kesimpulan.

2. Metode

Dalam proses pengelompokan data terdapat beberapa metode yang bisa digunakan.

2.1 Data Mining

Data mining adalah proses menggunakan pendekatan tertentu untuk menemukan *pattern* atau informasi tertentu dalam sekumpulan data. *Data mining* atau bisa disebut *knowledge discovery in database* (KDD) merupakan proses yang melibatkan pengumpulan dan analisis data historis untuk mengungkap sistem, *pattern*, apa keterkaitan dalam kumpulan data besar. Hasil pemrosesan *data mining* dapat dimanfaatkan untuk meningkatkan kemampuan dalam pengambilan keputusan di waktu mendatang [9].

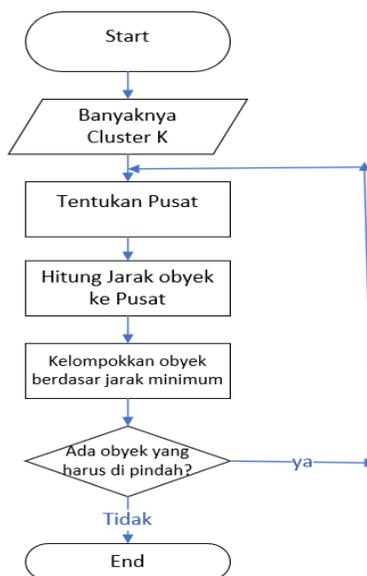
2.2 Clustering

Clustering adalah teknik untuk membedakan kumpulan data menjadi banyak kelompok dilihat dari kecocokan yang diinginkan. *Clustering* dalam *data mining* adalah kumpulan data atau objek dalam *cluster* (grup) dan menjadikan setiap *cluster* memiliki data yang hampir mirip dengan aslinya serta dapat dibedakan dari objek di *cluster* lainnya [10].

Pengelompokan diperlukan karena data yang belum diolah sangat besar dan tidak mudah untuk dilakukan analisis maupun dipelajari. Tujuan pengelompokan dalam skenario ini adalah untuk lebih memahami data dan menganalisis kualitas data. Metode *clustering* yang paling banyak dimanfaatkan dalam penelitian ilmiah dan industri adalah algoritma *K-Means*. Setiap *cluster* diwakili oleh *mean* dari data dalam *cluster* tersebut [11].

2.3 K-Means

Metode yang dapat dimanfaatkan untuk mengelompokkan/*cluster* data adalah *K-Means*. Teknik *clustering* digunakan secara luas, dan popularitasnya melambung diikuti dengan bertambahnya besaran data yang tersedia. *K-Means* adalah metode sederhana untuk analisis *clustering* yang bertujuan untuk menentukan cara terbaik untuk membagi entitas ke-*n* menjadi kelompok yang disebut *cluster* [12]. Diagram alir metode *K-Means* digambarkan pada Gambar 1 berikut.



Gambar 1. Flowchart K-Means

Dari Gambar 1 dapat dilihat diagram alir-*Means* yang dimulai dengan menentukan banyaknya *cluster*. Setelah nilai ditentukan, langkah selanjutnya adalah menentukan pusat *cluster*, dan dilanjutkan dengan menghitung jarak setiap objek setiap pusat *cluster*. Jarak antara tiap titik data yang ada dan tiap *centroid* akan dihitung menggunakan analisis *Euclidean* hingga ditentukan jarak terpendek antara setiap titik data dan *centroid*. Selanjutnya dilakukan pengelompokan objek berdasarkan jarak minimum terhadap pusat *cluster*. Pusat *cluster* tersebut kemudian untuk sementara dijadikan pusat *cluster*, atau *centroid*, *mean*. Jika masih terdapat objek yang harus dipindah ke *cluster* yang lain, maka proses diulang kembali, tetapi jika tidak, maka proses selesai[13].

Setiap *cluster* yang terbentuk akan meningkatkan kriteria partisi, seperti fungsi perbedaan berdasarkan jarak, sehingga objek-objek di dalam *cluster* menjadi mirip, dan objek-objek pada *cluster* yang berbeda ditemukan tidak mirip dalam hal atribut dataset. Jarak *Euclidean* digunakan sebagai ukuran jarak dalam pendekatan *K-Means* untuk menyoroti kesamaan antara setiap *cluster* dengan jarak terkecil dan kesamaan tertinggi. Jarak *Euclidean* antara titik $a = (a_1, a_2, \dots, a_k)$ dan titik $b = (b_1, b_2, \dots, b_n)$ dapat dihitung dengan Rumus (1).

$$d(b_i, a_t) = \sqrt{\sum_{j=1}^l (b_{ij} - a_{tj})^2} \quad (1)$$

dimana:

d = jarak antara nilai data dan nilai pusat *cluster*

b_i = nilai data, $i = 1, 2, \dots, n$, n = jumlah data

a_t = nilai pusat *cluster*, $t = 1, 2, \dots, K$, K = jumlah *cluster*

l = jumlah atribut atau dimensi [14]

Adapun algoritma *K-Means* adalah sebagai berikut [15]:

- Menentukan jumlah *cluster* yang akan dibuat, misalkan $K = 3$.
- Secara acak, membuat *centroid* awal atau titik pusat *cluster*, misalkan *centroid* C1: (79,84,80,84,82,80,82,82,8), *centroid* C2: (79,85,80,85,82,80,84,92,5), dan *centroid* C3: (80,82,78,82,85,80,83,82,7).
- Dengan jarak Euclidean antara dua objek, dihitung jarak setiap data b_i ke setiap *centroid* a_t . Misalkan data yang digunakan adalah murid bernama Alfadhea (A) yang mempunyai nilai (82,91,80,91,85,82,87,93,8), maka jarak data dengan masing-masing *centroid* adalah:

$$d(A, C1) = \sqrt{(82-79)^2 + (91-84)^2 + (80-80)^2 + (91-84)^2 + (85-82)^2 + (82-80)^2 + (87-82)^2 + (93-82)^2 + (8-8)^2}$$

$$d(A, C1) = \sqrt{266} = 16,31$$

$$d(A, C2) = \sqrt{(82-79)^2 + (91-85)^2 + (80-80)^2 + (91-85)^2 + (85-82)^2 + (82-80)^2 + (87-84)^2 + (93-92)^2 + (8-5)^2}$$

$$d(A, C2) = \sqrt{113} = 10,63$$

$$d(A, C3) = \sqrt{(82-80)^2 + (91-82)^2 + (80-78)^2 + (91-82)^2 + (85-85)^2 + (82-80)^2 + (87-83)^2 + (93-82)^2 + (8-7)^2}$$

$$d(A, C3) = \sqrt{312} = 17,66$$

- Urutkan data ke dalam kelompok-kelompok tergantung pada jarak terpendek antara data tersebut dengan setiap *centroid*. Tabel 1 merupakan hasil perhitungan jarak masing-masing data dengan nilai *centroid*.

Tabel 1. Penentuan keanggotaan data terhadap *cluster*

Nama	Perhitungan jarak data terhadap <i>centroid</i>			Jarak terdekat		
	C1	C2	C3	C1	C2	C3
Alfadhea	16,31	10,63	17,66		√	

- Hitung nilai rata-rata dari data yang ada pada *cluster* yang sama untuk mendapatkan posisi *centroid* baru. Setelah semua murid dihitung jarak terhadap semua *centroid*-nya dan ditentukan keanggotaan *cluster*-nya, maka dihitung posisi *centroid* yang baru. Tabel 2 merupakan hasil nilai pembaharuan/pengolahan data yang menggunakan nilai *centroid* di atas dengan menghitung rata-rata nilai pada masing-masing *cluster*.

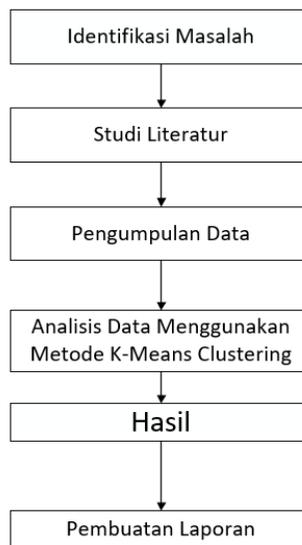
Tabel 2. Nilai *centroid* baru

<i>Cluster 1</i>	78	81	77	81	80	70	80	77	7
<i>Cluster 2</i>	81	83	82	83	84	80	83	80	7
<i>Cluster 3</i>	80	85	81	85	84	81	84	91	5

- g. Jika posisi *centroid* baru berbeda dari *centroid* sebelumnya, kembali ke langkah c, jika tidak maka selesai.

Perhitungan *K-Means* dalam perhitungan nilai murid pada penelitian ini menggunakan rumus dalam Excel yang kemudian diterapkan menggunakan program RStudio.

Struktur langkah-langkah yang jelas diperlukan untuk memberikan pedoman dalam pengembangan penelitian ini. Kerangka ini menguraikan langkah-langkah yang diambil dalam penelitian ini. Kerangka kerja terdiri dari pedoman yang digambarkan pada Gambar 2.

**Gambar 2.** Metode Penelitian

- a. Identifikasi masalah
Tahap awal dalam penyelidikan ini adalah untuk identifikasi masalah. Identifikasi masalah bertujuan untuk menjabarkan dan menjelaskan problem mengenai masalah yang akan diteliti dan berkembang pada objek yang diteliti sehingga latar belakang penelitian ini dapat diidentifikasi dengan bagaimana mengatasi pengelompokan prestasi murid berdasarkan nilai mata pelajaran.

- b. Studi literatur
Untuk mengumpulkan data dan informasi untuk penelitian, dilakukan studi pustaka dengan memanfaatkan studi ilmiah maupun literatur yang berkaitan dengan penelitian ini.
- c. Pengumpulan data
Penelitian dilakukan terhadap 25 data murid kelas X-XII Bahasa, dari semester satu sampai lima. Data tersebut didapat dari nilai rapor satu angkatan tahun ajaran 2014-2017 selama tiga tahun. Observasi lapangan dan analisis dokumen digunakan untuk memperoleh data. Observasi lapangan adalah metode pengumpulan data yang melibatkan terjun langsung ke lapangan dan melakukan *interview* dan observasi. *Interview* dilakukan ke pihak terkait diantaranya Kepala Sekolah, Dapodik, dan Wali Kelas Kelas X-XII Bahasa di SMA Negeri 1 Tenganan. *Interview* dilakukan untuk mengumpulkan data lisan dan keterangan yang akurat serta dapat dipertanggung jawabkan, tentang hasil belajar dan minat murid di SMAN 1 Tenganan. Untuk mengumpulkan informasi yang dibutuhkan, data digabungkan dengan hasil observasi di lokasi penelitian.
- d. Metode *K-Means* digunakan untuk menganalisis data
Analisis data dilakukan pada tahap ini dengan menggunakan survei yang lalu dan temuan wawancara. Data tersebut akan dianalisis dengan pendekatan *K-Means clustering* untuk mengetahui apakah prestasi belajar murid tergolong rendah, cukup, atau tinggi.
- e. Hasil
Data yang telah dianalisis dengan metode *K-Means clustering* menghasilkan penentuan pengelompokan prestasi murid rendah, cukup, dan tinggi.
- f. Pembuatan laporan
Pembuatan laporan berdasarkan hasil penelitian yang kemudian ditransformasikan menjadi laporan penelitian yang memuat penjabaran komprehensif.

3. Hasil dan Pembahasan

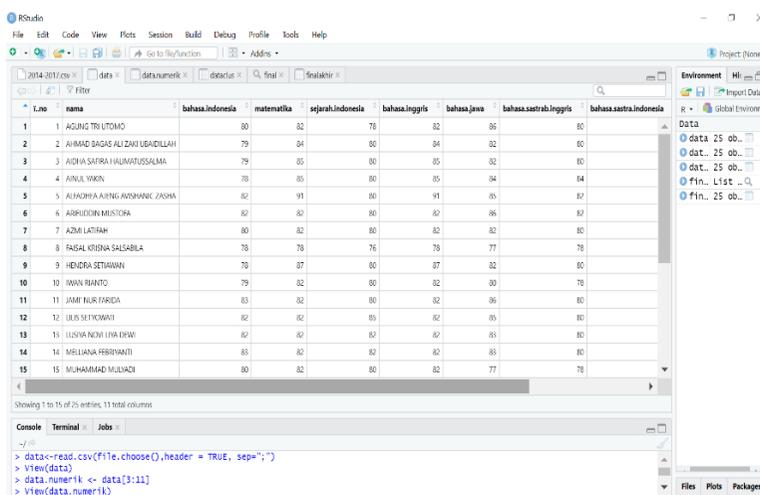
Software RStudio digunakan dalam penelitian ini untuk membahas pengelompokan menggunakan algoritma *K-Means* dengan menggunakan 25 data nilai dan absensi murid kelas X-XII Bahasa SMAN 1 Tenganan dari semester satu sampai lima. Data nilai dan absensi tersebut dijadikan satu data dengan mengambil rata-ratanya.

Tabel 3 merupakan data nilai rata-rata mata pelajaran dan absensi murid kelas X-XII Bahasa, dimana BI: Bahasa Indonesia, MAT: Matematika, SIND: Sejarah Indonesia, BING: Bahasa Inggris, BJW: Bahasa Jawa, BSING: Bahasa dan Sastra Inggris, BSIND: Bahasa dan Sastra Indonesia, BSJP: Bahasa dan Sastra Jepang, dan ABS: Absensi.

Tabel 3. Data nilai rata-rata mata pelajaran murid kelas X-XII Bahasa

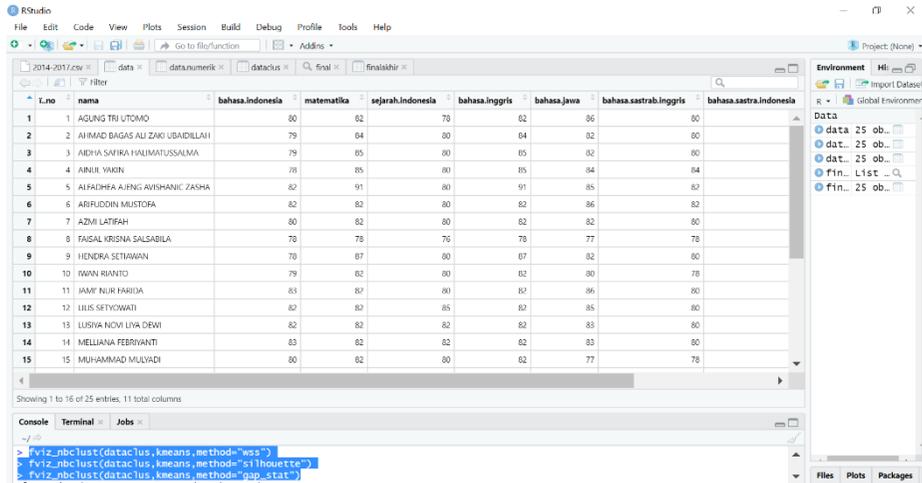
Nama	BI	MAT	SIND	BING	BJW	BSING	BSIND	BSJP	ABS
AGUNG TRI U	80	82	78	82	86	80	83	82	7
AHMADBAGAS	79	84	80	84	82	80	82	82	8
AIDHASAFIRA	79	85	80	85	82	80	84	92	5
AINUL YAKIN	78	85	80	85	84	84	85	92	8
ALFADHEA	82	91	80	91	85	82	87	93	6
ARIFUDDIN	82	82	80	82	86	82	85	80	7
AZMILATIFAH	80	82	80	82	82	80	80	80	5
FAISAL K	78	78	76	78	77	78	80	75	6
HENDRA S	78	87	80	87	82	80	83	82	7
IWAN RIAN TO	79	82	80	82	80	78	80	75	6
JAMP' NUR F	83	82	80	82	86	80	82	82	8
LILIS S	82	82	85	82	85	80	85	80	7
LUSIYA NOVI	82	82	82	82	83	80	86	82	8
MELLIANA F	83	82	82	82	83	80	82	82	7
MUHAMMAD	80	82	80	82	77	78	80	80	7
OKTARI N	80	82	85	82	85	80	83	80	8
RADITYA R	78	82	76	82	82	78	80	80	9
RIAN BUDI	78	84	76	84	82	80	80	75	8
RIDHA K	78	87	80	87	85	80	87	80	7
ROBBY ELVIN	78	82	77	82	82	78	82	75	11
SEKAR SAFIT	82	82	80	82	86	80	82	91	5
SITINURJANAH	82	82	85	82	84	80	85	80	6
SRI PUJI L	83	85	85	85	84	80	82	80	7
WIDY DWI	82	84	85	84	84	80	82	91	3
YENI LASTRI	82	85	85	85	83	80	85	80	6

Gambar 3 merupakan data nilai dan absensi murid yang sudah dimasukkan dalam bentuk Excel dan dipanggil ke RStudio.



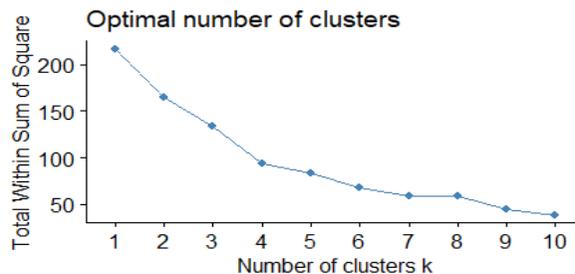
Gambar 3. Data nilai murid

Ada beberapa cara yang digunakan untuk menentukan jumlah *cluster* yaitu menggunakan metode *Elbow*, *Silhouette*, dan *Gap*. Gambar 4 merupakan cara menentukan *cluster* dengan tiga metode yaitu *wss/Elbow*, *Silhouette*, dan *Gap* pada RStudio.



Gambar 4. Penentuan jumlah *cluster* pada RStudio

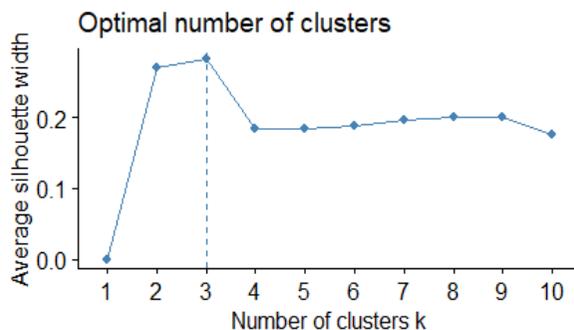
Cara pertama menggunakan metode *Elbow*. Gambar 5 menunjukkan diagram garis hasil metode *Elbow*. Pada gambar tersebut terdapat titik tumpul yang terbentuk diantara titik tiga dan lima, setelah titik tiga sudah tidak ada lagi penurunan yang terlalu dalam namun di angka delapan terdapat kenaikan sedikit yang kemudian turun lagi, sehingga jumlah *cluster* dari hasil metode *Elbow* adalah sejumlah tiga *cluster*, $K=3$.



Gambar 5. Diagram garis hasil metode *Elbow*

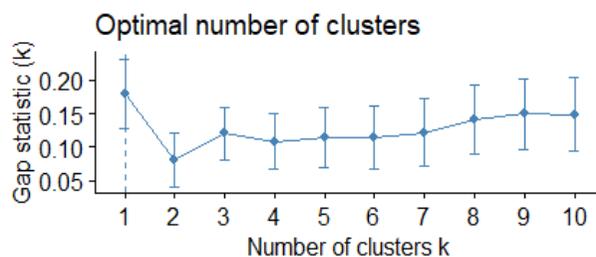
Cara kedua yaitu menggunakan metode *Silhouette*. Gambar 6 menunjukkan diagram garis hasil metode *Silhouette*. Pada gambar tersebut terdapat titik tertinggi yang terbentuk diantara titik dua dan empat, setelah titik tiga penurunan yang terlalu dalam tidak terjadi namun di angka delapan terdapat kenaikan sedikit yang

kemudian turun lagi, sehingga jumlah *cluster* menurut hasil metode *Silhouette* adalah sebanyak tiga *cluster*.



Gambar 6. Diagram garis hasil metode Silhouette

Cara ketiga menggunakan metode *Gap*. Gambar 7 menunjukkan diagram garis hasil metode *Gap*. Pada gambar tersebut terdapat titik-titik yang tertuju di angka satu. Setelah titik satu terdapat penurunan yang terlalu dalam yang kemudian naik secara perlahan, sampai jumlah *cluster* menurut hasil metode *Gap* adalah sebanyak satu *cluster*.

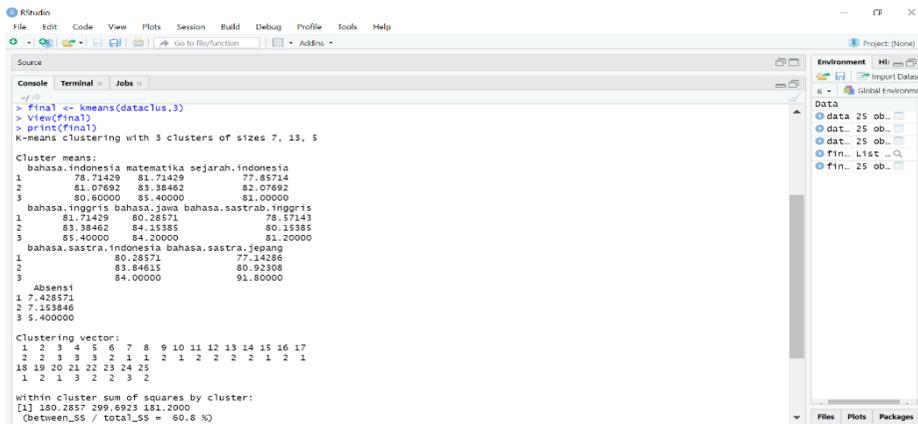


Gambar 7. Diagram garis hasil metode Gap

Dari ketiga cara di atas untuk membuktikan jumlah *cluster*, digunakan nilai paling optimal dari *Elbow*, *Silhouette*, dan *Gap*. Nilai sama didapat dari metode *Elbow* dan *Silhouette* yang menghasilkan nilai tiga, sehingga jumlah *cluster* yang dipakai dalam penelitian ini adalah tiga.

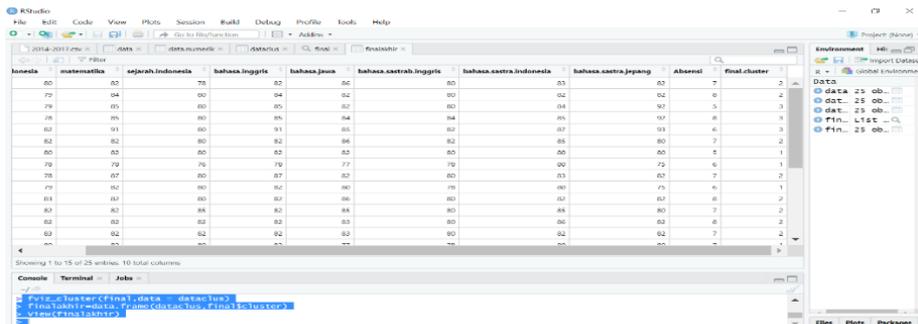
3.1 Metode Clustering Menggunakan Metode K-Means

Dari Tabel 3 akan diteruskan datanya ke dalam RStudio serta akan dimasukkan beberapa perintah untuk menampilkan hasil *cluster* yang ada di Gambar 8. Gambar 8 merupakan hasil *cluster* data nilai murid berdasarkan nilai mata pelajaran. Gambar 9 merupakan hasil pengelompokan nilai murid menggunakan metode *K-Means*. Dari temuan final *cluster* yang ada di Gambar 9 akan ditampilkan di Tabel 4 secara keseluruhan dari nilai mata pelajaran dan absensi.



Gambar 8. Hasil cluster

Gambar 8 merupakan hasil cluster data nilai murid berdasarkan nilai mata pelajaran dan absensi yang telah di proses melalui RStudio serta menggunakan insert code yang diterapkan dalam program ini.



Gambar 9. Hasil akhir

Gambar 9 merupakan hasil pengelompokan nilai murid menggunakan metode K-Means. Dari temuan final cluster yang ada di Gambar 9 akan ditampilkan di Tabel 4 secara keseluruhan dari nilai mata pelajaran dan absensi.

Tabel 4. Hasil akhir cluster nilai murid

Nama	Cluster
AGUNG TRI U	2
AHMADBAGAS	2
AIDA SAFIRA	3
AINUL YAKIN	3
ALFADHEA	3
ARIFUDDIN	2
AZMI LATIFAH	1
FAISAL K	1

HENDRA S	2
IWAN RIAN TO	1
JAMI' NUR F	2
LILIS S	2
LUSIYA NOVI	2
MELLIANA F	2
MUHAMMAD	1
OKTARI N	2
RADITYA R	1
RIAN BUDI	1
RIDHA K	2
ROBBY ELVIN	1
SEKAR SAFIT	3
SITINURJANA H	2
SRI PUJI L	2
WIDY DWI	3
YENI LASTRI	2

Tabel 4 merupakan hasil akhir *cluster* nilai murid berdasarkan nilai mata pelajaran dan absensi pada Gambar 9.

Tabel 5. Hasil pusat *cluster*

<i>Cluster</i>	<i>Cluster means</i>								
	BI	MAT	SIN	BING	BJW	BSING	BSIN	BSJP	ABS
1	78	81	77	81	80	70	80	77	7
2	81	83	82	83	84	80	83	80	7
3	80	85	81	85	84	81	84	91	5

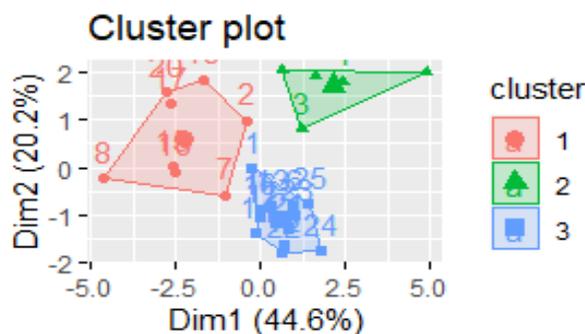
Berdasarkan hasil pada Tabel 5 terdapat pusat *cluster* sebanyak tiga, dimana *cluster* 1 merupakan *cluster* murid dengan prestasi Rendah, *cluster* 2 merupakan *cluster* murid dengan prestasi cukup, dan *cluster* 3 merupakan *cluster* murid dengan prestasi Tinggi. Tabel 6 menunjukkan pembagian tiga *cluster* untuk data nilai murid Kelas X-XII Bahasa menggunakan algoritma *K-Means*, dimana jumlah anggota *cluster* 1 dengan kategori rendah sebanyak 12 anggota, *cluster* 2 dengan kategori cukup sebanyak lima anggota, dan *cluster* 3 dengan kategori tinggi sebanyak delapan anggota.

Tabel 6. Hasil *cluster* data murid

<i>Cluster</i>	Kategori	Nama
1	Rendah	JAMI' NUR FARIDA AGUNG TRI UTOMO RIDHA KURNIA AMANAH AHMAD BAGAS ALI ZAKI HENDRA SETIAWAN AZMI LATIFAH MUHAMMAD MULYADI RIAN BUDI S IWAN RIAN TO

		ROBBY ELVIN YEREMIA
		RADITYA ARI WIBOWO
		FAISAL KRISNA S
2	Cukup	AIDHA SAFIRA H
		OKTARI NUR K
		MELLIANA FEBRIYANTI
		LUSIYA NOVI LIYA DEWI
		LILIS SETYOWATI
3	Tinggi	ALFADHEA AJENG A
		AINUL YAKIN
		WIDY DWI ASTUTI
		SRI PUJI LESTARI
		SITI NURJANAH
		SEKAR SAFITRI
		ARIFUDDIN MUSTOFA
		YENI SULASTRI

Hasil *cluster plot* menggunakan algoritma *K-Means* ditunjukkan pada Gambar 10. Dari gambar tersebut, terlihat bahwa terdapat dimensi yang merupakan Dim1 sebesar 44,6% dan Dim2 sebesar 20,2%, yang meliputi delapan mata pelajaran dan satu nilai absensi, serta 25 titik yang membentuk gambar, tersebar di *cluster* 1,2, dan 3.



Gambar 10. Diagram *cluster plot*

Hasil pengelompokan siswa ini dapat digunakan sekolah untuk menganalisis keberhasilan murid di setiap kelompok. Hasil dari *cluster* tinggi dapat digunakan untuk memberikan penghargaan kepada para siswa yang berprestasi. Sementara itu, hasil dari *cluster* cukup dan rendah dapat dimanfaatkan juga untuk memotivasi murid untuk lebih giat belajar agar bisa mencapai prestasi yang lebih tinggi.

Selanjutnya, dari *cluster* tinggi, dicari *top rank* berdasarkan rata-rata nilai mata pelajaran dan absensinya. Pemilihan *top rank* menggunakan metode *Simple Additive Weighting* (SAW), dimana untuk setiap mata pelajaran diberi bobot sama yaitu 0.12

dan absensi diberi bobot 0.04. Tabel 7 merupakan nilai siswa yang berada pada *cluster* tinggi.

Tabel 7. Nilai siswa di *cluster* Tinggi

Alternatif	Kriteria								
	BI	MAT	SIND	BING	BJW	BSING	BSIND	BSJP	ABS
Bobot	0.12	0.04							
AINUL YAKIN	78	85	80	85	84	84	85	92	8
ALFADHEA	82	91	80	91	85	82	87	93	6
ARIFUDDIN	82	82	80	82	86	82	85	80	7
SEKAR SAFIT	82	82	80	82	86	80	82	91	5
SITINURJANA	82	82	85	82	84	80	85	80	6
SRI PUJIL	83	85	85	85	84	80	82	80	7
WIDY DWI	82	84	85	84	84	80	82	91	3
YENI LASTRI	82	85	85	85	83	80	85	80	6

Karena semua kriteria merupakan kriteria benefit, maka normalisasi dari Tabel 7 pada perhitungan SAW mengikuti rumus $r_{ij} = \frac{x_{ij}}{x_{ij}^*}$, dimana x_{ij} merupakan data pada Tabel 7 dan r_{ij} adalah rating kinerja ternormalisasi dari alternatif A_i pada kriteria C_j , dengan $i = 1, 2, \dots, m$ dan $j = 1, 2, \dots, n$, m adalah banyak alternatif dan n adalah banyak kriteria.

$$R = \begin{pmatrix} 0.94 & 0.93 & 0.94 & 0.93 & 0.98 & 1.00 & 0.98 & 0.99 & 1.00 \\ 0.99 & 1.00 & 0.94 & 1.00 & 0.99 & 0.98 & 1.00 & 1.00 & 0.75 \\ 0.99 & 0.90 & 0.94 & 0.90 & 1.00 & 0.98 & 0.98 & 0.86 & 0.88 \\ 0.99 & 0.90 & 0.94 & 0.90 & 1.00 & 0.95 & 0.94 & 0.98 & 0.63 \\ 0.99 & 0.90 & 1.00 & 0.90 & 0.98 & 0.95 & 0.98 & 0.86 & 0.75 \\ 1.00 & 0.93 & 1.00 & 0.93 & 0.98 & 0.95 & 0.94 & 0.86 & 0.88 \\ 0.99 & 0.92 & 1.00 & 0.92 & 0.98 & 0.95 & 0.94 & 0.98 & 0.38 \\ 0.99 & 0.93 & 1.00 & 0.93 & 0.97 & 0.95 & 0.98 & 0.86 & 0.75 \end{pmatrix}$$

Selanjutnya, nilai preferensi untuk setiap alternatif (V_i) mengikuti rumus $V_i = \sum_{j=1}^n w_j r_{ij}$, dimana w_j adalah bobot masing-masing kriteria j . Tabel 8 merupakan perhitungan nilai V .

Tabel 8. Perhitungan V

Alternatif	Perhitungan	V
AINUL YAKIN	$0.12*0.94 + 0.12*0.93 + 0.12*0.94 + 0.12*0.93 + 0.12*0.98 + 0.12*0.98 + 0.12*1.00 + 0.12*0.99 + 0.04*1.00$	0.96
ALFADHEA	$0.12*0.99 + 0.12*1.00 + 0.12*0.94 + 0.12*1.00 + 0.12*0.99 + 0.12*0.98 + 0.12*1.00 + 0.12*1.00 + 0.04*0.75$	0.98
ARIFUDDIN	$0.12*0.99 + 0.12*0.90 + 0.12*0.94 + 0.12*0.90 + 0.12*1.00 + 0.12*0.98 + 0.12*0.98 + 0.12*0.86 + 0.04*0.88$	0.94
SEKAR SAFIT	$0.12*0.99 + 0.12*0.90 + 0.12*0.94 + 0.12*0.90 + 0.12*1.00 + 0.12*0.95 + 0.12*0.94 + 0.12*0.98 + 0.04*0.63$	0.94
SITINURJANA	$0.12*0.99 + 0.12*0.90 + 0.12*1.00 + 0.12*0.90 + 0.12*0.98$	0.94

	$+ 0.12*0.95 + 0.12*0.98 + 0.12*0.86 + 0.04*0.75$	
SRI PUJI L	$0.12*1.00 + 0.12*0.93 + 0.12*1.00 + 0.12*0.93 + 0.12*0.98$	0.95
	$+ 0.12*0.95 + 0.12*0.94 + 0.12*0.86 + 0.04*0.88$	
WIDY DWI	$0.12*0.99 + 0.12*0.92 + 0.12*1.00 + 0.12*0.92 + 0.12*0.98$	0.94
	$+ 0.12*0.95 + 0.12*0.94 + 0.12*0.98 + 0.04*0.38$	
YENI LASTRI	$0.12*0.99 + 0.12*0.93 + 0.12*1.00 + 0.12*0.93 + 0.12*0.97$	0.94
	$+ 0.12*0.95 + 0.12*0.98 + 0.12*0.86 + 0.04*0.75$	

Dengan mengurutkan nilai V dari besar ke kecil, maka didapatkan *top rank*-nya adalah Alfadhea, diikuti Ainul Yakin, kemudian Sri Puji L. Berdasarkan hasil wawancara langsung dengan guru maupun wali kelas, didapatkan bahwa Alfadhea, Ainul Yakin, dan Widy Dwi secara bergantian menduduki *top rank* di setiap semesternya. *Top rank* dapat digunakan sekolah untuk memberikan penghargaan kepada siswa-siswinya.

4. Kesimpulan

Metode *K-Means* dapat mengelompokkan data nilai murid dalam kumpulan kecil maupun besar dan efektif dalam mengelompokkan dokumen dengan tepat, seperti yang terlihat di atas. Penentuan *centroid* (titik pusat) pada tahap awal algoritma *K-Means* memiliki pengaruh besar pada penemuan *cluster*, seperti yang ditunjukkan dengan pengujian memanfaatkan 25 *record* yang *centroid*-nya berbeda menghasilkan *cluster* yang bervariasi. Berdasarkan hasil penelitian ini didapat tiga *cluster* yaitu *cluster* 1 yang merupakan kategori rendah sebanyak 12 anggota, *cluster* 2 yang merupakan kategori cukup sebanyak lima anggota dan *cluster* 3 yang merupakan kategori tinggi sebanyak delapan anggota. Dari *cluster* tinggi, dapat dilanjutkan dengan mencari siswa unggulan (*top rank*). Dengan metode SAW, didapatkan Alfadhea yang menduduki *top rank*. Hasil ini dapat dimanfaatkan sekolah untuk menganalisis prestasi dari murid juga bisa membantu guru maupun wali kelas dalam membentuk kelas yang ideal dalam arti memotivasi murid untuk belajar lebih giat lagi.

Penggunaan metode lain untuk menghitung data yang dijelaskan di atas, atau memperluasnya dengan lebih banyak data, diharapkan dapat diaplikasikan dalam penelitian selanjutnya. Dapat diprediksi bahwa penelitian ini akan mengarah pada aplikasi maupun *tools* tambahan yang tetap digunakan dalam menyelesaikan permasalahan data murid di sekolah.

Daftar Pustaka

- [1] Netriwati, “*K-Means – Penerapan, Permasalahan dan Metode Terkait*,” *J. Pendidik. Mat.*, vol. 3, no. Pebruari, hal. 47–60, 2017.
- [2] M. Silalahi, “Computer Based Information System Journal Analisis *Clustering* Menggunakan Algoritma *K-Means*,” *Comput. Based Inf. Syst. Journa*, vol. 02, hal. 20–35, 2018.

- [3] A. Utomo, "Implementasi Metode *K-Means Clustering* Untuk Analisa Prestasi Siswa Berdasarkan Data Siswa," hal. 1–9, 2016.
- [4] A. E. Wicaksono, "Implementasi data mining dalam pengelompokan data peserta didik di sekolah untuk memprediksi calon penerima beasiswa dengan menggunakan algoritma k- means (studi kasus sman 16 bekasi)," *J. Teknol. rekayasa*, vol. 21, no. 3, 2016.
- [5] N. Jannah dan T. Yulianto, "Mengelompokkan Siswa Berprestasi Akademik dengan Menggunakan Metode K Means Kelas VII MT," *Math J.*, vol. 2, no. 2, hal. 41–45, 2016.
- [6] F. L. Sibuea dan A. Sapta, "Pemetaan Siswa Berprestasi Menggunakan Metode *K-Means Clustering*," *Jurteksi*, vol. IV, no. 1, 2017.
- [7] P. A. Jusia dan F. M. Irfan, "*Clustering* Data Untuk Rekomendasi Penentuan Jurusan Perguruan Tinggi Menggunakan Metode *K-Means*," *J. IKRA-ITTH Inform.*, vol. 3, no. 3, hal. 75–84, 2019.
- [8] S. Sabariah, Marrylinteri Istoningtyas, "Penentuan Jurusan ke Perguruan Tinggi Menggunakan Metode *Clustering* di SMAN 3 Kuala Tungkal," *J. Process.*, vol. 13, no. 2, 2018.
- [9] Y.Mardi, "Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD)," *J. Edik Inform.*, vol. 2, 2016.
- [10] A. Ramadhan dan Z. Efendi, "Perbandingan *K-Means* dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.* 9, hal. 18–19, 2017.
- [11] B. M. Metisen dan H. L. Sari, "Analisis *Clustering* Menggunakan Metode *K-Means* Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, hal. 110–118, 2015.
- [12] A. & R.Andrian, "Penerapan Metode *K-Means* Untuk *Clustering* Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *J. Ilm. Tek.*, vol. 18, no. 1, hal. 76–82, 2015.
- [13] Y. Darmi dan A. Setiawan, "Penerapan Metode *Clustering K-Means* Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, hal. 148–157, 2016.
- [14] D.H.Margareth and H.Precise, *Data Mining introductory and advanced topic*Data Mining introductory and advanced Topics, vol. 9780130888. 2020.
- [15] D .T. Larose and C. D. Larose, *Discovering Knowledge in Data : An Introduction to Data Mining:Second edition*, vol. 978047098. 2014.