

Data Mining Rekomendasi Sekolah Calon Siswa SMA di Kota Tomohon Menggunakan Metode K-Means Clustering

Civita C. I. Loho¹, Vivi P. Rantung², Gladly C. Rorimpandey³

1,2,3 Program Studi Teknik Informatika UNIMA

Email: 17210145@unima.ac.id

Abstract—Tomohon is a small city in North Sulawesi. In Tomohon City, there are 12 high schools and the equivalent that are officially registered on the reference data site of the Ministry of Education and Culture. In practice, the number of students enrolled in 12 schools in each academic year has a significant difference. This is influenced by the high public perception of schools that are labeled ‘favorite schools’ with other schools. This study aims to group data on prospective high school students in Tomohon City for the 2021/2022 academic year using the K-Means Clustering method for high school recommendations. K-Means is a partition-based Clustering method in Data Mining where data will be grouped based on the level of similarity, so that student data that have the same characteristics are grouped into one cluster, while student data with different characteristics are grouped into different clusters.

Keywords— Information Technology, Data Mining, Clustering, K-Means.

Abstrak—Tomohon merupakan sebuah kota kecil di Provinsi Sulawesi Utara. Di kota Tomohon ada 12 Sekolah Menengah Atas (SMA) dan setaranya yang terdaftar secara resmi di situs Data Referensi Kementerian Pendidikan dan Kebudayaan. Pada pelaksanaannya, jumlah siswa yang masuk di 12 sekolah pada setiap tahun ajaran memiliki perbedaan yang cukup signifikan. Hal ini dipengaruhi oleh tingginya persepsi masyarakat terhadap sekolah yang diberi label ‘sekolah favorit’ dengan sekolah yang lain. Penelitian ini bertujuan untuk mengelompokkan data calon siswa SMA di Kota Tomohon tahun ajaran 2021/2022 menggunakan metode *K-Means Clustering* untuk rekomendasi SMA berdasarkan wilayah. *K-Means* merupakan metode *Clustering* berbasis partisi pada Data Mining dimana data akan dikelompokkan berdasarkan tingkat kemiripan, sehingga data siswa yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* sedangkan data siswa yang memiliki karakteristik berbeda dikelompokkan kedalam *cluster* yang berbeda pula.

Kata Kunci— Teknik Informatika, Data Mining, Klastering, K-Means.

I. PENDAHULUAN

Pendidikan merupakan hal yang sangat penting dalam kemajuan suatu bangsa, sebab kualitas hidup suatu bangsa tercermin dari kualitas pengetahuan masyarakat di dalamnya. Semakin tinggi standar pendidikan di negara tersebut, maka semakin tinggi juga tingkat kesejahteraan masyarakatnya. Pendidikan merupakan sebuah proses untuk melahirkan generasi penerus yang berkualitas. Menjadi bangsa yang maju tentu merupakan cita-cita setiap negara di dunia tak terkecuali Indonesia.

Di kota Tomohon, terdapat 12 Sekolah Menengah Atas (SMA). Pada pelaksanaannya, jumlah siswa yang masuk di 12 sekolah pada setiap tahun ajaran memiliki perbedaan yang cukup signifikan. Hal ini dipengaruhi oleh tingginya persepsi masyarakat terhadap sekolah yang diberi label ‘sekolah favorit’ dengan sekolah yang lain.

Untuk mendukung program pemerintah dalam pemerataan Pendidikan di Indonesia khususnya di kota Tomohon, penulis membuat rekomendasi sekolah bagi calon siswa SMA di kota Tomohon tahun ajaran 2021/2022.

II. TINJAUAN PUSTAKA

A. Data Mining

Data Mining atau yang bisa disebut *Knowledge Discovery in Databases (KDD)* adalah suatu proses pengerukan atau pengumpulan informasi penting dari suatu data yang besar atau dapat dikatakan sebagai suatu proses pengambilan pola pada data yang akan di proses dengan *output* yang berupa informasi penting.

Berdasarkan fungsionalitasnya, tugas-tugas *data mining* dapat dikelompokkan ke dalam enam kelompok yaitu: klasifikasi (*classification*), klasterisasi (*clustering*), regresi (*regression*), deteksi anomaly (*anomaly detection*), pembelajaran aturan asosiasi (*association rule learning*) atau pemodelan ketergantungan (*dependency modelling*), dan perangkuman (*summarization*) (Suyanto, 2019).

B. Metode Clustering

Clustering atau klasterisasi adalah metode pengelompokan data. *Clustering* merupakan salah satu aktifitas analisis data yakni klasifikasi atau pengelompokan data ke dalam beberapa kategori/cluster (flinsetyadiz, 2020).

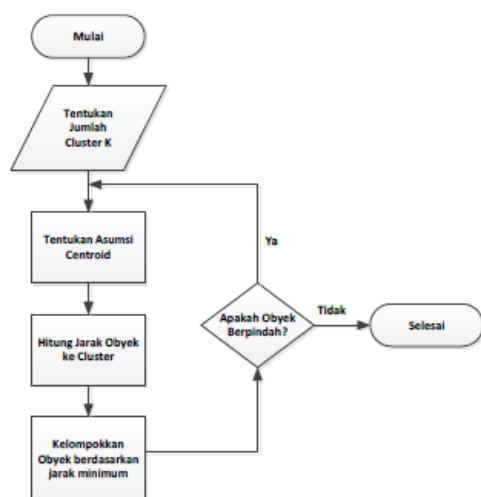
Hasil *clustering* yang baik akan menghasilkan tingkat kesamaan yang tinggi dalam suatu kelas dan tingkat

kesamaan yang rendah antar kelas. Kesamaan yang dimaksud merupakan pengukuran secara *numeric* terhadap dua buah objek. Nilai kesamaan antar kedua objek akan semakin tinggi jika kedua objek yang dibandingkan memiliki kemiripan yang tinggi. Begitu juga dengan sebaliknya. Kualitas hasil *clustering* sangat bergantung pada metode yang dipakai.

C. Algoritma K-Means

K-means merupakan salah satu metode *cluster analysis* non hirarki yang berusaha untuk mempartisi objek yang ada kedalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam *cluster* yang lain.

Data clustering menggunakan metode *K-Means Clustering* ini secara umum dilakukan dengan algoritma dasar sebagai berikut:



Gbr. 1 Alur Algoritma *K-Means Clustering*

1. Tentukan jumlah cluster
2. Alokasikan data ke dalam cluster secara random
3. Hitung centroid/rata-rata dari data yang ada di masing-masing cluster
4. Alokasikan masing-masing data ke centroid/rata-rata terdekat
5. Kembali ke Step 2, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai *threshold* yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai *threshold* yang ditentukan.

D. Transformasi Normal (Normalisasi)

Normalisasi merupakan suatu teknik untuk mengorganisasikan data kedalam tabel-tabel untuk memenuhi kebutuhan pemakai didalam suatu organisasi. Data-data yang dilakukan normalisasi dengan membagi nilai data tersebut dengan nilai range data (nilai data maksimum – nilai data minimum). Tujuan dari normalisasi yaitu:

1. Untuk menghilangkan kerangkapan data
2. Untuk mengurangi kompleksitas
3. Untuk mempermudah pemodifikasian data

$$x_n \frac{x_0 - x_{min}}{x_{max} - x_{min}}$$

Dengan,

X_n = nilai data normal

X_0 = nilai data actual

X_{min} = nilai minimum data aktual keseluruhan

X_{max} = nilai maksimum data aktual keseluruhan

III. METODE

A. Metode Pengumpulan Data

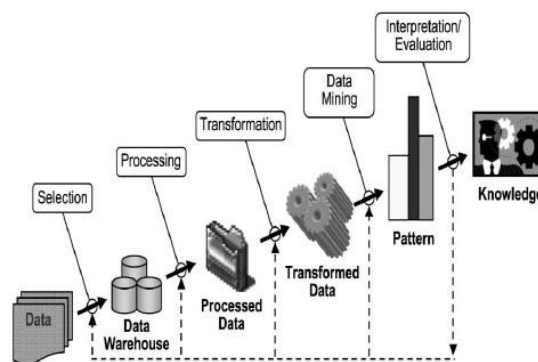
Dalam proses pengembangan penelitian ini, dibutuhkan referensi-referensi dari berbagai sumber untuk mengumpulkan informasi dan data yang terkait serta melakukan pengecekan keakuratan teori-teori yang digunakan guna mendapatkan hasil penelitian yang valid dan konkret. Untuk itu ada beberapa teknik pengumpulan data yang dilakukan dalam penelitian ini yaitu sebagai berikut:

1) Studi Dokumen / Studi Pustaka (*Library Reasearch*): Melakukan pengumpulan data dengan meneliti berbagai macam dokumen yang berguna untuk bahan analisis. Dokumen-dokumen tersebut berupa dokumen primer dan sekunder, dengan memanfaatkan buku-buku maupun media internet.

2) Angket (Kuisisioner): Proses pengumpulan data dengan cara memberikan seperangkat pertanyaan atau pernyataan tertulis kepada responden untuk dijawab.

B. Rancangan Penelitian

Tahapan-tahapan yang ditempuh dalam pelaksanaan penelitian ini adalah sebagai berikut:



Gbr. 2 Tahapan Proses *Data Mining*

1) *Data Selection*: Kegiatan pengumpulan data dilakukan dengan memanfaatkan layanan Google Form dimana proses pengumpulan data dimulai dari penyebaran link Google Form kepada para subjek penelitian yaitu siswa SMP kelas 9 di Kota Tomohon melalui grup kelas daring yang bekerja sama dengan pihak pendidik di masing-masing sekolah.

2) *Pre-processing / Cleaning*: Sebelum data diolah menggunakan metode K-Means Clustering, data terlebih dahulu melewati proses cleaning yang bertujuan untuk

membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

3) *Transformation*: Proses ini dilakukan untuk menyesuaikan nilai range data dengan melakukan inisialisasi variabel-variabel pada data ke dalam nilai-nilai tertentu agar data dapat diolah kedalam metode K-Means Clustering.

4) *Data Mining*: Pada tahap ini, penulis menerapkan metode K-Means Clustering. Metode K-Means Clustering dipilih untuk pengolahan data pada penelitian ini sebagai metode pengolahan data mining yang paling familiar bagi penulis dengan beberapa pengalaman pengelolaan data menggunakan metode K-Means Clustering sebelumnya.

5) *Interpretation / Evaluation*: Pada tahapan akhir ini, hasil pengolahan data dengan metode K-Means Clustering pada Data Mining kemudian ditampilkan untuk dilakukan pemeriksaan.

IV. HASIL DAN PEMBAHASAN

Tujuan dari penelitian ini adalah untuk mengelompokkan data calon siswa SMA di kota Tomohon tahun ajaran 2021/2022 menggunakan metode *K-Means Clustering*. Terdapat beberapa tahapan pengolahan data yang dilakukan sesuai dengan tahapan pengolahan *data mining* yaitu sebagai berikut:

1) *Data Selection*:

Penelitian ini dimulai dari pengumpulan data calon siswa SMA di kota Tomohon tahun ajaran 2021/2022 yang dilakukan di seluruh SMP yang ada di kota Tomohon dengan menjalankan kuisisioner/angket menggunakan platform aplikasi berbasis web bernama Google Form. Kuisisioner/angket ini diberikan kepada siswa kelas 9 di masing-masing sekolah. Dari proses pengumpulan data yang dilakukan, diperoleh sampel data sebanyak 475 yang terdiri dari Nama Siswa, Asal Sekolah, dan Alamat Tempat Tinggal. Berikut contoh data calon siswa SMA tahun ajaran 2021/2022 yang diperoleh.

TABEL. I. DATA KUISISIONER CALON SISWA SMA KOTA TOMOHON TAHUN AJARAN 2021/2022

No	Timestamp	Nama	Alamat	
			Kelurahan	Kecamatan
1	20/01/2021 10:39	Adolf chris leonardo toreh	Tinoor Dua	Tomohon
2	12/01/2021 11:08	agnesya destiny senduk	Lahendong	Tomohon Selatan
3	30/01/2021 11:31	Agusstiven Tumundo	Tinoor Dua	Tomohon Utara
4	30/01/2021 10:25	Agustinus Richardo Aray	Kolongan Satu	Tomohon Tengah
5	30/01/2021 16:21	Aily debora intawijaya	Rurukan	Tomohon Timur
6	13/01/2021 18:11	Aisyah Otoluwa	Paslaten Satu	Tomohon Timur
7	16/01/2021 10:08	Alan Samuel Tristan Mumpel	Kolongan	Tomohon Tengah
8	29/01/2021 20:12	Albertus agung ngala	Taratara Satu	Tomohon Barat
9	06/01/2021 15:32	Alessandro nesta milano tuuk	Taratara Dua	Tomohon Barat

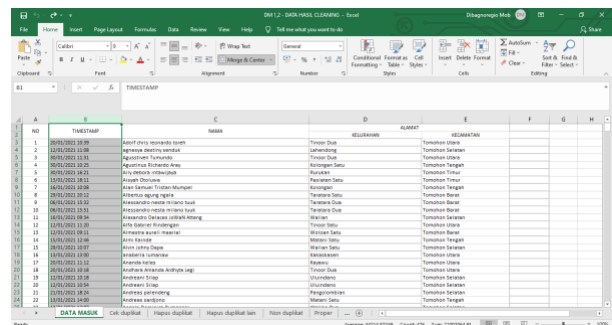
10	06/01/2021 15:51	Alessandro nesta milano tuuk	Taratara Dua	Tomohon Barat
----	------------------	------------------------------	--------------	---------------

2) *Pre-processing / Cleaning*:

Data yang telah terkumpul kemudian di seleksi untuk memisahkan antara data yang layak dan tidak layak diolah. Proses cleaning ini dilakukan secara otomatis dengan memanfaatkan fungsi-fungsi pada *Microsoft Office Excel*.

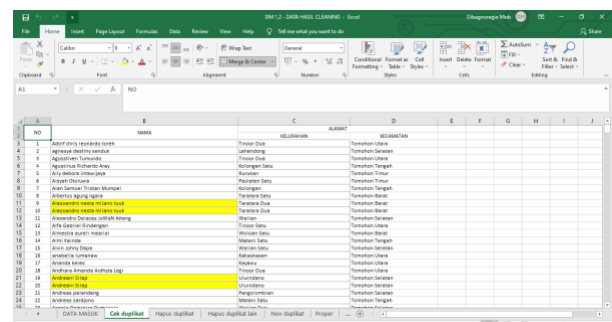
Langkah pertama yang dilakukan adalah menghapus tabel 'timestamp' yang tidak akan digunakan dalam pengolahan data pada penelitian ini sehingga yang tersisa adalah data yang akan diolah yang terdiri dari tabel 'nomor', tabel 'nama' dan tabel 2 variable yaitu:

1. Alamat Kelurahan Tempat Tinggal
2. Alamat Kecamatan Tempat Tinggal



Gbr. 3 Proses Penghapusan Tabel 'timestamp'

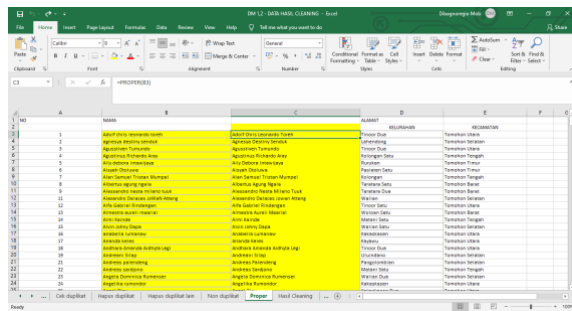
Langkah kedua yaitu menyeleksi data ganda yang dilakukan dengan fungsi "Duplicate Value" di "Highlight Cells Rules" pada tool "Conditional Formatting" menu "Home" dimana fungsi ini bekerja dengan cara membaca data yang sama dan kemudian cel dengan data ganda tersebut ditandai sesuai dengan format cel yang diinginkan. Disini, saya menggunakan format cel "yellow fill" sehingga cel yang berisi data ganda ditandai dengan warna cel kuning. Karena data ganda yang akan di cek adalah data nama siswa, maka fungsi "duplicate value" ini digunakan pada kolom 'nama' di tabel. Setelah data-data ganda ditemukan, dilakukan penghapusan dengan perintah "format only unique or duplicate value" di "New Rule" pada tool yang sama. Berikut adalah gambar proses penyeleksian data ganda yang dilakukan:



Gbr. 4 Proses Seleksi Data Ganda

Langkah terakhir yang dilakukan pada proses ini yaitu merapikan data hasil seleksi data ganda. Untuk merapikan penulisan nama pada kolom 'nama' dilakukan dengan rumus "proper". Setelah data nama dirapikan, data disusun

berdasarkan alfabet A ke Z yang dilakukan dengan fungsi “sort A to Z” pada tool “sort & filter” menu “Home”.



Gbr. 5 Proses Merapikan Data

Setelah melewati proses cleaning didapatkan sebanyak 402 data yang layak untuk masuk ke proses pengolahan, sedangkan sisanya adalah data yang tidak layak untuk diolah. Data yang tidak akan digunakan tersebut berjumlah 73 data.

Contoh data hasil seleksi dengan 2 variabel yang akan digunakan terdapat pada tabel dibawah ini:

TABEL. II. DATA HASIL CLEANING

No	Nama	Alamat	
		Kelurahan	Kecamatan
1	Adolf Chris Leonardo Toreh	Tinoor Dua	Tomohon Utara
2	Agnesya Destiny Senduk	Lahendong	Tomohon Selatan
3	Agusstiven Tumundo	Tinoor Dua	Tomohon Utara
4	Agustinus Richardo Aray	Kolongan Satu	Tomohon Tengah
5	Aily Debora Intawijaya	Rurukan	Tomohon Timur
6	Aisyah Otoluwa	Paslaten Satu	Tomohon Timur
7	Alan Samuel Tristan Mumpel	Kolongan	Tomohon Tengah
8	Albertus Agung Ngala	Taratara Satu	Tomohon Barat
9	Alessandro Nesta Milano Tuuk	Taratara Dua	Tomohon Barat
10	Alexandro Delaces Jowan Atteng	Walian	Tomohon Selatan

3) Transformation:

Transformasi / normalisasi dilakukan agar data dapat diolah kedalam metode K-Means Clustering. Pada proses normalisasi data ini, kedua variable yang akan digunakan masing-masing ditransformasikan dimana untuk variabel alamat kecamatan dikelompokkan menjadi 5 kelompok. Adapun kelompok kecamatan Tomohon Barat ditransformasikan dengan nilai 1, kelompok kecamatan Tomohon Selatan ditransformasikan dengan nilai 2, kelompok kecamatan Tomohon Tengah ditransformasikan dengan nilai 3, kelompok kecamatan Tomohon Timur ditransformasikan dengan nilai 4, dan kelompok kecamatan Tomohon Utara ditransformasikan dengan nilai 5. Untuk variabel alamat kelurahan dikelompokkan menjadi 44 kelompok. Adapun kelompok data variabel alamat kelurahan yaitu sebagai berikut:

TABEL. III. HASIL TRANSFORMASI DATA VARIABEL ALAMAT

Kelurahan	Transformasi dengan nilai
Taratara	1
Taratara Satu	2
Taratara Dua	3
Taratara Tiga	4
Woloan Satu	5
Woloan Satu Utara	6
Woloan Dua	7
Woloan Tiga	8
Kampung Jawa	9
Lahendong	10
Lansot	11
Pangolombian	12
Pinaras	13
Tumatangtang	14
Tumatangtang Satu	15
Tondangow	16
Uluindano	17
Walian	18
Walian Satu	19
Walian Dua	20
Kamasi	21
Kamasi Satu	22
Kolongan	23
Kolongan Satu	24
Matani Satu	25
Matani Dua	26
Matani Tiga	27
Talete Satu	28
Talete Dua	29
Kumelembuai	30
Paslaten Satu	31
Paslaten Dua	32
Rurukan	33
Rurukan Satu	34
Kakaskasen	35
kakaskasen Satu	36
Kakaskasen Dua	37
Kakaskasen Tiga	38
Kayawu	39
Kinilow	40
Kinilow Satu	41
Tinoor Satu	42
Tinoor Dua	43
Wailan	44

Proses transformasi data ini dilakukan secara otomatis dengan memanfaatkan rumus “if” pada *Microsoft Office Excel*. Berikut contoh data hasil transformasi atau normalisasi yang dilakukan.

No	Nama	Alamat	
		Kelurahan	Kecamatan
1	Adolf Chris Leonardo Toreh	43	5
2	Agnesya Destiny Senduk	10	2
3	Agusstiven Tumundo	43	5
4	Agustinus Richardo Aray	24	3
5	Aily Debora Intawijaya	33	4
6	Aisyah Otoluwa	31	4
7	Alan Samuel Tristan Mumpel	23	3
8	Albertus Agung Ngala	2	1
9	Alessandro Nesta Milano Tuuk	3	1
10	Alexandro Delaces Jowan Atteng	18	2

Wilayah	Sekolah
1	SMAN 2 Tomohon
2	SMAK Peter and Paul SMAN 1 Tomohon SMAS Kosgoro Tomohon
3	SMAS Kristen 1 Tomohon SMAS Kristen 2 Tomohon
4	SMA Lentera Harapan Tomohon SMAS Katolik Karitas Tomohon SMAS PGRI Tomohon
5	MAS Mardhatilllah SMAS Katolik Seminari Kakaskasen Tomohon SMAS Lokon ST Nikolaus Tomohon

Titik Pusat		
Cendroid 1	1	1
Cendroid 2	9	2
Cendroid 3	21	3
Cendroid 4	30	4
Cendroid 5	35	5

Dari hasil perhitungan diatas didapatkan jarak data siswa pertama dengan *cluster* tiga adalah 22,09.

Jarak data siswa pertama ke titik pusat *cluster* empat yaitu:

$$\begin{aligned} d(1,4) &= \sqrt{(43-30)^2 + (5-4)^2} \\ &= \sqrt{(13)^2 + (1)^2} \\ &= 13,04 \end{aligned}$$

Dari hasil perhitungan diatas didapatkan jarak data siswa pertama dengan *cluster* empat adalah 13,04.

Jarak data siswa pertama ke titik pusat *cluster* lima yaitu:

$$\begin{aligned} d(1,5) &= \sqrt{(43-35)^2 + (5-5)^2} \\ &= \sqrt{(8)^2 + (0)^2} \\ &= 8 \end{aligned}$$

Tahap 4 :

Berdasarkan hasil perhitungan pada tahap 3 dapat disimpulkan bahwa *cluster* dengan jarak terdekat pada data siswa pertama adalah *cluster* 5, sehingga siswa pertama masuk ke dalam kelompok *cluster* 5. Berikut tabel hasil perhitungan dan hasil *cluster* 20 sampel siswa:

TABEL. VII. HASIL PERHITUNGAN DATA KE SETIAP CLUSTER PADA ITERASI 1

Data ke-i	Jarak Ke Cendroid					Jarak Terdekat	Cluster Diikuti
	1	2	3	4	5		
1	42,19	34,13	22,09	13,04	8,00	8,00	5
2	9,06	1,00	11,05	20,10	25,18	1,00	2
3	42,19	34,13	22,09	13,04	8,00	8,00	5
4	23,09	15,03	3,00	6,08	11,18	3,00	3
5	32,14	24,08	12,04	3,00	2,24	2,24	5
6	30,15	22,09	10,05	1,00	4,12	1,00	4
7	22,09	14,04	2,00	7,07	12,17	2,00	3
8	1,00	7,07	19,10	28,16	33,24	1,00	1
9	2,00	6,08	18,11	27,17	32,25	2,00	1
10	17,03	9,00	3,16	12,17	17,26	3,16	3

Tahap 5 :

Setelah dilakukan penghitungan dan penempatan data pada *cluster* terdekat, dilakukan pengujian dengan melakukan penghitungan kembali jarak titik pusat masing-masing data ke *cluster* pada iterasi selanjutnya yaitu iterasi 2. Apabila cendroid yang baru *konvergen* dengan cendroid yang lama maka iterasi dihentikan. Sebaliknya, apabila cendroid yang baru belum *konvergen* dengan cendroid yang lama maka iterasi dilanjutkan.

Pada iterasi kedua ini, Langkah pertama yang harus dilakukan adalah menentukan titik pusat (cendroid) baru. Berbeda dengan penentuan titik pusat (cendroid) pada iterasi pertama. Apabila pada iterasi pertama penentuan titik pusat (cendroid) dilakukan secara random, pada iterasi kedua dan seterusnya, penentuan titik pusat (cendroid) dilakukan dengan rumus:

$$C = \frac{\sum m}{n}$$

Dimana C adalah cendroid n, m adalah jumlah nilai masing-masing variabel x dan y pada data ke-n, dan n adalah jumlah data anggota *cluster* n.

Sebagai contoh, dilakukan perhitungan untuk mencari titik pusat baru cendroid 1 di iterasi 2. Untuk mendapatkan nilai m dari variabel x dan y, dilakukan dengan menjumlahkan seluruh nilai masing-masing variabel x dan y untuk semua data yang masuk kedalam *cluster* 1 karena titik pusat yang akan dicari adalah titik pusat cendroid 1. Dari hasil penghitungan, nilai m pada variabel x = 240 dan y = 80. Jumlah data yang masuk ke *cluster* 1 pada iterasi pertama adalah 80 sehingga n = 80. Nilai-nilai yang sudah didapatkan tersebut kemudian disubstitusikan kedalam rumus dengan hasil penghitungan sebagai berikut:

$$\text{Titik pusat cendroid 1 variabel x : } \frac{240}{80} = 3$$

$$\text{Titik pusat cendroid 1 variabel y : } \frac{80}{80} = 1$$

Untuk penghitungan titik pusat baru pada cendroid lainnya diuraikan sebagai berikut:

Berikut perhitungan titik pusat baru pada cendroid 2:

$$\text{Titik pusat cendroid 2 variabel x : } \frac{748}{80} = 9,35$$

$$\text{Titik pusat cendroid 2 variabel y : } \frac{119}{80} = 1,49$$

Berikut perhitungan titik pusat baru pada cendroid 3:

$$\text{Titik pusat cendroid 3 variabel x : } \frac{1254}{60} = 20,9$$

$$\text{Titik pusat cendroid 3 variabel y : } \frac{151}{60} = 2,52$$

Berikut perhitungan titik pusat baru pada cendroid 4:

$$\text{Titik pusat cendroid 4 variabel x : } \frac{1033}{35} = 29,51$$

$$\text{Titik pusat cendroid 4 variabel y : } \frac{124}{35} = 3,54$$

Berikut perhitungan titik pusat baru pada cendroid 5:

$$\text{Titik pusat cendroid 5 variabel x : } \frac{5704}{147} = 38,80$$

$$\text{Titik pusat cendroid 5 variabel y : } \frac{717}{147} = 4,88$$

Dari hasil perhitungan diatas didapatkan titik pusat baru iterasi 2 sebagai berikut:

TABEL. VIII. TITIK PUSAT ITERASI 1 SETELAH CLUSTER

Titik Pusat		
Cendroid 1	3	1
Cendroid 2	9,35	1,49
Cendroid 3	20,9	2,52
Cendroid 4	29,51	3,54
Cendroid 5	38,80	4,9

Setelah didapatkan titik pusat (cendroid) baru, dilakukan kembali perhitungan jarak titik pusat (cendroid) baru masing-masing data ke setiap *cluster*. Hasil penghitungan tersebut terdapat pada tabel berikut:

TABEL. IX. HASIL PERHITUNGAN DATA KE SETIAP CLUSTER PADA ITERASI 2

Data ke-i	Jarak Ke Cendroid					Jarak Terdekat	Cluster Diikuti
	1	2	3	4	5		
1	40,20	33,83	22,24	13,56	4,20	4,20	5

2	10,0 5	0,83 3	10,9 1	19,5 8	28,9 5	0,83 3	2
3	43,1 9	33,8 3	22,2 4	13,5 6	4,20 2	4,20 2	5
4	24,0 8	14,7 3	3,14 1	5,54 2	14,9 2	3,14 2	3
5	33,1 4	23,7 8	12,1 9	3,52 1	5,87 1	3,52 1	4
6	31,1 4	21,8 0	10,2 1	1,55 1	7,85 1	1,55 1	4
7	23,0 9	13,7 3	2,15 1	6,54 1	15,9 1	2,15 1	3
8	2,00 3	7,37 6	18,9 6	27,6 3	37,0 1	2,00 1	1
9	3,00 3	6,37 8	17,9 6	26,6 4	36,0 1	3,00 1	1
10	18,0 3	8,67 2	2,95 2	11,6 2	21,0 0	2,95 2	3

Dari hasil perhitungan, didapatkan bahwa hasil *cluster* yang didapatkan belum konvergen karena masih terdapat data-data yang berpindah *cluster*. Data-data tersebut yaitu data ke: 5, 12, 30, 54, 56, 57, 60, 68, 85, 101, 106, 121, 167, 205, 209, 221, 222, 223, 271, 277, 286, 292, 310, 311, 322, 324, 344, 365, 370, 372, 375, dan 393 dengan total data yang berpindah *cluster* yaitu sebanyak 32 data. Karena masih terdapat data yang berpindah *cluster*, maka iterasi dilanjutkan ke iterasi 3.

TABEL. X. HASIL PERHITUNGAN DATA KE SETIAP *CLUSTER* PADA ITERASI 3

Dat a ke-i	Jarak Ke Cendroid					Jarak Terdeka t	Cluste r Diikut i
	1	2	3	4	5		
1	40,6 2	34,4 8	22,2 4	12,2 4	3,44 1	3,44 1	5
2	10,0 5	1,42 1	10,9 1	20,9 0	29,7 1	1,42 1	2
3	43,1 9	34,4 8	22,2 4	12,2 4	3,44 1	3,44 1	5
4	24,0 8	15,3 8	3,14 1	6,87 2	15,6 9	3,14 2	3
5	33,1 4	24,4 4	12,1 9	2,19 1	6,63 1	2,19 1	4
6	31,1 4	22,4 5	10,2 1	0,35 1	8,62 1	0,35 1	4
7	23,0 9	14,3 9	2,15 1	7,86 2	16,6 8	2,15 1	3
8	2,00 3	6,71 6	18,9 6	28,9 6	37,7 7	2,00 1	1
9	3,00 3	5,72 8	17,9 6	27,9 6	36,7 8	3,00 1	1
10	18,0 3	9,32 2	2,95 2	12,9 4	21,7 7	2,95 2	3

Pada iterasi ketiga, masih didapatkan data-data yang berpindah *cluster* diantaranya data ke: 15, 22, 42, 71, 90, 102, 147, 156, 163, 238, 287, 358, 373, dan 399 dengan total data yang berpindah *cluster* yaitu sebanyak 14 data. Karena masih terdapat data yang berpindah *cluster*, maka iterasi dilanjutkan ke iterasi 4.

TABEL. XI. HASIL PERHITUNGAN DATA KE SETIAP *CLUSTER* PADA ITERASI 4

Dat a ke-i	Jarak Ke Cendroid					Jarak Terdeka t	Cluste r Diikut i
	1	2	3	4	5		
1	40,6 2	34,6 2	22,4 3	11,4 5	2,97 1	2,97 1	5
2	10,0 5	1,55 1	10,7 2	21,6 9	30,1 8	1,55 1	2
3	43,1 9	34,6 2	22,4 3	11,4 5	2,97 1	2,97 1	5
4	24,0 8	15,5 2	3,33 1	7,66 1	16,1 5	3,33 1	3
5	33,1 4	24,5 7	12,3 8	1,40 1	7,10 1	1,40 1	4
6	31,1 4	22,5 8	10,4 0	0,60 1	9,08 1	0,60 1	4
7	23,0 9	14,5 2	2,34 1	8,65 1	17,1 4	2,34 1	3
8	2,00 3	6,58 6	18,7 7	29,7 5	38,2 4	2,00 1	1
9	3,00 3	5,58 8	17,7 7	28,7 5	37,2 4	3,00 1	1
10	18,0 3	9,45 2	2,76 1	13,7 4	22,2 3	2,76 1	3

Pada iterasi keempat, masih terdapat beberapa data yang berpindah *cluster* diantaranya data ke: 100, 105, 366, dan 367 dengan total data yang berpindah *cluster* yaitu sebanyak 4 data. Karena masih terdapat data yang berpindah *cluster*, maka iterasi dilanjutkan ke iterasi 5.

TABEL. XII. HASIL PERHITUNGAN DATA KE SETIAP *CLUSTER* PADA ITERASI 5

Dat a ke-i	Jarak Ke Cendroid					Jarak Terdeka t	Cluster Diikuti
	1	2	3	4	5		
1	40,6 2	34,6 2	22,1 1	11,0 8	2,97 1	2,97 1	5
2	10,0 5	1,55 1	11,0 4	22,0 6	30,1 8	1,55 1	2
3	43,1 9	34,6 2	22,1 1	11,0 8	2,97 1	2,97 1	5
4	24,0 8	15,5 2	3,01 1	8,03 1	16,1 5	3,01 1	3
5	33,1 4	24,5 7	12,0 6	1,03 1	7,10 1	1,03 1	4
6	31,1 4	22,5 8	10,0 8	0,97 1	9,08 1	0,97 1	4
7	23,0 9	14,5 2	2,02 1	9,02 1	17,1 4	2,02 1	3
8	2,00 3	6,58 6	19,0 9	30,1 2	38,2 4	2,00 1	1
9	3,00 3	5,58 8	18,1 0	29,1 2	37,2 4	3,00 1	1
10	18,0 3	9,45 2	3,08 1	14,1 1	22,2 3	3,08 1	3

Pada proses iterasi kelima, sudah tidak didapatkan lagi data-data yang berpindah *cluster* sehingga iterasi dihentikan karena hasil *cluster* pada iterasi 5 sudah mencapai titik konvergen.

5) Interpretation / Evaluation :

Dari proses pengolahan data yang dilakukan, didapatkanlah hasil klasterisasi seperti yang terdapat pada tabel di bawah ini:

TABEL. XIII. JUMLAH DATA HASIL *CLUSTER*

CLUSTER	JUMLAH
1	66
2	92
3	66
4	61
5	117
TOTAL DATA	402

Hasil pembentukan *cluster* divisualisasikan kedalam grafik berikut:

Gbr. 6 Grafik Hasil *Cluster*

Berdasarkan hasil klasterisasi diatas, diperoleh data rekomendasi sekolah calon siswa SMA di kota Tomohon tahun ajaran 2020/2021 yang dapat dilihat pada tabel berikut:

TABEL. XIV. TABEL HASIL PENGELOMPOKAN REKOMENDASI SEKOLAH

NO	NAMA	CLUSTER
1	Adolf Chris Leonardo Toreh	5
2	Agnesya Destiny Senduk	2
3	Agusstiven Tumundo	5
4	Agustinus Richardo Aray	3
5	Aily Debora Intawijaya	4
6	Aisyah Otoluwa	4
7	Alan Samuel Tristan Mumpel	3
8	Albertus Agung Ngala	1
9	Alessandro Nesta Milano Tuuk	1
10	Alexandro Delaces Jowan Atteng	3

V. KESIMPULAN

Berdasarkan hasil analisis penelitian yang dilakukan dan telah dibahas pada bab-bab sebelumnya, maka dapat disimpulkan beberapa hal sebagai berikut, yaitu:

- 1) Pengolahan data rekomendasi sekolah calon siswa SMA di kota Tomohon dapat dilakukan dengan menggunakan metode *K-Means Clustering*.
- 2) Pengelompokan data siswa SMP kelas 9 kota Tomohon tahun ajaran 2021/2022 untuk rekomendasi SMA menghasilkan sebanyak 66 siswa direkomendasikan untuk masuk SMA di wilayah satu, sebanyak 92 siswa direkomendasikan masuk SMA di wilayah dua, sebanyak 66 siswa direkomendasikan masuk SMA di wilayah tiga, sebanyak 61 siswa direkomendasikan

masuk SMA di wilayah empat, dan sebanyak 117 siswa direkomendasikan masuk SMA di wilayah lima.

UCAPAN TERIMA KASIH

Penyusunan naskah ini dapat terselesaikan berkat dukungan, topangan, dan arahan dari berbagai pihak. Untuk itu, saya ucapkan banyak terima kasih.

DAFTAR PUSTAKA

- [1] Arthasalina, D. S. (2019, Desember 05). *Pendidikan Indonesia Peringkat 13 di Asia, Ini Daftar Lengkapnya*. Retrieved from IDN TIMES: <https://www.google.co.id/amp/s/www.idntimes.com/life/education/amp/dian-arthasalina/pendidikan-indonesia-peringkat-13-di-asia>
- [2] Caesaria, S. D. (2021, 05 22). *PPDB 2021: Ini Aturan Jalur Zonasi Versi Permendikbud*. Retrieved from KOMPAS.com: <https://www.kompas.com/edu/read/2021/05/22/084502271/ppdb-2021-ini-aturan-jalur-zonasi-versi-permendikbud?page=all>
- [3] Fammaldo, E., & Hakim, L. (2018). PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN TINGKAT KESEJAHTERAAN KELUARGA UNTUK PROGRAM KARTU INDONESIA PINTAR. *JITTER*, 23-31.
- [4] flinsetyadi. (2020, Juli 20). *Algoritma Clustering Data Mining: Metode Partisi*. Retrieved from Flin Setyadi: <https://flinsetyadi.com/algoritma-clustering-dalam-data-mining-metode-partisi/>
- [5] Harusulilo, Y. E. (2019, Januari 29). *8 Dampak Negatif "Favoritisme Sekolah" Menurut Kemendikbud*. Retrieved from KOMPAS.com: <https://edukasi.kompas.com>
- [6] Imron, M. (2019, Januari 14). *Sejarah Awal Mula Terbentuknya Kota Tomohon Sulawesi Utara*. Retrieved from Kuwaluhan.com: <https://www.kuwaluhan.com/2019/01/sejarah-awal-mula-terbentuknya-kota.html?m=1>
- [7] Kalumata, A. (2019, November 13). *Tomohon Menuju Iklim Pendidikan Kondusif, Maju dan Berdaya Saing*. Retrieved from MANADO TODAY: <https://www.google.co.id>
- [8] Lumentut, J. (2014, Februari 23). *Profil Kota Tomohon*. Retrieved from seputarsulut.com: <https://www.seputarsulut.com>
- [9] MAMIKOS. (2021, Januari 13). *Mengenal Apa itu Sistem Zonasi Sekolah di PPDB 2021/2022*. Retrieved from mamikos.com: <https://mamikos.com/info/mengenal-apa-itu-sistem-zonasi/>
- [10] Nur, F., Zarlis, P., & Nasution, D. B. (2017). PENERAPAN ALGORITMA K-MEANS PADA SISWA BARU SEKOLAH MENENGAH KEJURUAN UNTUK CLUSTERING JURUSAN. *Jurnal Nasional Informatika dan Teknologi Jaringan*, 100-105.
- [11] Nurzahputra, A., Muslim, M. A., & Khusniati, M. (2017). Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa. *Techno.COM*, 17-24.
- [12] *Permendikbud 44 tahun 2019 tentang PPDB pada TK, SD, SMP, SMA, dan SMK*. (2020, Januari 09). Retrieved from JOGLOABANG: <https://www.jogloabang.com/pendidikan/permendikbud-44-2019-ppdb-tk-sd-smp-sma-smk>
- [13] *PPDB Sistem Zonasi Bermasalah, Kemendikbud Revisi Aturan*. (2019, Juni 20). Retrieved from CNN Indonesia: <https://www.cnnindonesia.com/nasional/20190620204456-12-405120/ppdb-sistem-zonasi-bermasalah-kemendikbud-revisi-aturan>
- [14] Prahasti. (2017). DATA MINING DALAM PENGELOMPOKAN JENIS DAN JUMLAH PEMBAGIAN ZAKAT DENGAN MENGGUNAKAN METODE CLUSTERING K-MEANS. *Jurnal Teknologi Informasi*, 2011-2018.

- [15] PUSDATIN. (n.d.). *JUMLAH DATA SATUAN PENDIDIKAN (SEKOLAH) PER KABUPATEN/KOTA*. Retrieved from DATA REFERENSI KEMENTERIAN PENDIDIKAN & KEBUDAYAAN: referensi.data.kemendikbud.go.id
- [16] Riadi, M. (2017, September 21). *Pengertian, Fungsi, Proses dan Tahapan Data Mining*. Retrieved from KajianPustaka.com: <https://www.kajianpustaka.com>
- [17] *Sejarah Tomohon*. (n.d.). Retrieved from [tomohon.info: https://www.tomohon.info/sejarah-tomohon/](https://www.tomohon.info/sejarah-tomohon/)
- [18] Siregar, M. (2018). KLASTERISASI PENJUALAN ALAT-ALAT BANGUNAN MENGGUNAKAN METODE K-MEANS . *JURNAL TEKNOLOGI DAN OPEN SOURCE*, 83-91.
- [19] Suwirmayanti, N. L., & Saryanti, I. A. (2019). PENERAPAN TEKNIK CLUSTERING UNTUK PENGELOMPOKAN KONSENTRASI MAHASISWA DENGAN METODE K-MEANS. *SINTESA Prosiding*, 519-526.
- [20] Suyanto, D. (2019). *DATA MINING UNTUK KLASIFIKASI DAN KLASTERISASI DATA*. Bandung: Informatika.
- [21] Tomohon.info. (n.d.). *KECAMATAN*. Retrieved from [tomohon.info: www.tomohon.info](https://www.tomohon.info)
- [22] Yunita, F. (2018). PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PADA PENERIMAAN MAHASISWA BARU. *SISTEMASI*, 238-249.