

Pemanfaatan Data Mining untuk Melihat Minat Siswa Setelah Menyelesaikan Pendidikan Sekolah Menengah Atas (SMA) dengan Algoritma *K-Means Clustering*

Rahmat Hidayat
Universitas Putra Bangsa
email: ayat150190@gmail.com

Abstrak

Sekolah Menengah Atas (SMA) merupakan tingkatan terakhir Pendidikan menengah yang harus ditempuh siswa sebelum melanjutkan jenjang Pendidikan tinggi. Baik SMA, SMK ataupun MA memiliki level atau jenjang yang sama dan harus ditempuh dalam waktu 3 tahun oleh siswa. Siswa yang telah menyelesaikan Pendidikan setingkat SMA dapat melanjutkan Pendidikan ketingkat yang lebih tinggi ataupun memilih bekerja, karena juga banyak lowongan yang memberikan persyaratan minimal untuk mengisi suatu formasi yaitu minimal lulus SMA. Banyak hal yang mempengaruhi siswa dalam menentukan pilihan, apakah ingin melanjutkan Pendidikan yang lebih tinggi ataupun lebih memilih untuk bekerja. Salah satu yang memiliki pengaruh yang besar pemilihan tersebut ialah, pekerjaan orang tua yang pastinya juga berbanding lurus dengan pendapatannya, lingkungan sekitar, dll. Dari permasalahan tersebut penulis tertarik untuk melakukan penelitian dengan menggunakan metode data mining dengan algoritma *K-Means Clustering* untuk mengelompokan minat serta bakat siswa dalam menentukan pilihan setelah menyelesaikan Pendidikan setingkat SMA. Data yang digunakan dalam penelitian ini merupakan data yang didapatkan dari data kuesioner yang diisi langsung oleh siswa kelas XII Madrasah Aliyah DR. H. Abdullah Ahmad PGAI Padang. Dari sebanyak 25 orang siswa yang mengisi kuesioner. Maka dapat disimpulkan bahwa kecendrungan siswa untuk ingin melanjutkan Pendidikan kejenjang yang lebih tinggi sangat tinggi, apabila dipersentasekan dari *Clustering 1 (C1)*, dimana yang menjadi kelompok cluster 1 adalah siswa dengan penghasilan orang tua < Rp. 1.000.000 yaitu sebanyak 67 % ingin melanjutkan Pendidikan kejenjang yang lebih tinggi dan 33% ingin bekerja. Sedangkan dari cluster 2 (*C2*), dimana cluster ini merupakan cluster dengan nilai maximum dan bisa dikatakan siswa yang memiliki orang tua yang berpenghasilan > Rp. 5.000.000, terdapat 69% siswa memilih melanjutkan Pendidikan kejenjang yang lebih tinggi dan 31% lebih memilih untuk bekerja. Diharapkan hasil dari penelitian ini dapat membantu perguruan tinggi yang terdapat diwilayah padang atau Sumatra barat dalam menjanging minat siswa untuk bergabung melanjutkan Pendidikan kejenjang Pendidikan yang lebih tinggi.

Kata kunci: Data Mining; *K-Means*; *Clustering*; SMA

Abstract

High School (SMA) is the last level of secondary education that students must take before continuing with higher education. Both SMA, SMK or MA have the same level or levels and must be taken within 3 years by students. Students who have completed high school education can continue their education to a higher level or choose to work, because there are also many vacancies that provide a minimum requirement to fill a formation, namely a minimum of high school graduation. Many things influence students in making choices, whether they want to continue their higher education or prefer to work. One of the things that has a big influence on the election is, the work of parents which of course is also directly proportional to their income, the surrounding environment, etc. From these problems the authors are interested in conducting research using data mining methods with the K-Means Clustering algorithm to classify students' interests and talents in making choices after completing high school education. The data used in this study is data obtained from questionnaire data which was filled in directly by class XII students of Madrasah Aliyah DR. H. Abdullah Ahmad PGAI Padang. Of the 25 students who filled out the questionnaire. So it can be concluded that the tendency of students to want to continue their education to a higher level is very high, if it is a percentage of Clustering 1 (C1), where the cluster 1 group is students with parental income < Rp. 1,000,000, i.e. 67% want to continue their education to a higher level and 33% want to work. Meanwhile, from cluster 2 (C2), where this cluster is the cluster with the maximum value and it can be said that students who have parents who earn > Rp. 5,000,000, there are 69% of students who choose to continue their education to a higher level and 31% prefer to work. It is hoped that the results of this study can help universities in the Padang or West Sumatra region in capturing student interest to join continuing education at a higher level of education.

Keyword: Data Mining; K-Means; Clustering; High School

PENDAHULUAN

Rancangan Undang-Undang (RUU Sisdiknas) mengatur wajib belajar yang dimulai dari usia prasekolah atau usia anak 6 tahun. Didalam RUU yang terbaru ini terdapat perubahan yaitu tentang wajib belajar yang sebelumnya wajib belajar 9 tahun menjadi wajib belajar 13 tahun. Dengan rincian 1 tahun Pendidikan prasekolah, 6 tahun Pendidikan sekolah dasar, 3 tahun Pendidikan sekolah menengah pertama, dan 3 tahun Pendidikan sekolah menengah atas. Tujuan dari peraturan ini antara lain, menciptakan pemerataan Pendidikan diseluruh wilayah Indonesia serta untuk meningkatkan kualitas sumber daya manusia secara global. Setelah menyelesaikan Pendidikan wajib belajar 13 tahun siswa bebas memilih untuk melanjutkan Pendidikan kejenjang yang lebih tinggi ataupun memilih bekerja. Banyak hal yang mempengaruhi siswa dalam menentukan pilihan, kondisi keuangan keluarga yang kurang memadai, factor lingkungan bahkan factor kejenuhan siswa dalam belajar juga memiliki peran penting dalam pemilihan keputusan siswa setelah menyelesaikan Pendidikan setingkat SMA.

Knowledge Discovery in Database (KDD) didefinisikan sebagai ekstraksi informasi potensial, implisit dan tidak dikenal dari sekumpulan data. Proses Knowledge Discovery in Database melibatkan hasil proses dari data mining (proses mengekstraksi kecenderungan suatu pola data), kemudian merubah hasilnya secara akurat menjadi informasi yang mudah dipahami (Harahap. B., 2019). Knowledge Discovery in Database memiliki Teknik menganalisa data untuk digali informasi yang tersembunyi dari jumlah data yang besar dan kompleks, sehingga menghasilkan output berupa karakteristik atau pola dari data tersebut. Salah satu teknik Analisa data mining adalah analisis kelompok (*cluster analysis*) atau yang lebih dikenal dengan *Clustering*.

Data mining sudah banyak diterapkan dalam berbagai bidang, salah satunya metode *Clustering* khususnya metode *K-Means*, antara lain oleh (Darmi.Y, Setiawan.A, 2018) menjelaskan tentang algoritma *K-Means Clustering* yang dapat diimplementasikan dengan model perangkat lunak untuk verifikasi citra sidik jari poin minutiae. *K-Means Clustering* adalah metode untuk mengkategorikan atau mengelompokkan sekelompok objek sesuai attribute yang sama atau karakteristik ke dalam sejumlah groups (jumlah bilangan bulat positif). Ini mendefinisikan sebuah cluster oleh massa yang mewakili mean dari cluster. Dalam data mining analisis cluster populer adalah *K-Means*. Metode *K-Means* merupakan metode kuantitas vector, hal ini sesuai dari yang saya lihat dilapangan. Yakni dengan melakukan identifikasi dari kuesioner yang telah saya bagikan kepada beberapa orang siswa untuk melihat bagaimana kecenderungan yang terjadi terhadap minat siswa dalam melanjutkan Pendidikan apabila didasarkan terhadap beberapa factor yang mempengaruhi.

Dalam penelitian ini, dikemukakan suatu metode pengolahan data dengan mengelompokkan data menggunakan algoritma *K-Means Clustering* dari hasil kuesioner yang telah diisi oleh beberapa siswa kelas XII Madrasah Aliyah DR. H. Abdullah Ahmad PGAI Padang. Dari data yang didapatkan ini dilakukan identifikasi untuk melihat kecenderungan yang terjadi atau factor apa saja yang mempengaruhi siswa dalam menentukan pilihannya setelah menyelesaikan Pendidikan setingkat Sekolah Menengah Atas (SMA). Diharapkan hasil dari penelitian ini dapat membantu perguruan tinggi yang terdapat dikota padang atupun sumatera barat dalam menjaring minat siswa untuk melanjutkan Pendidikan yang lebih tinggi kejenjang perguruan tinggi. Dengan hasil *Clustering* ini pihak perguruan tinggi dapat melakukan pemetaan terhadap sekolah-sekolah untuk menentukan sekolah mana yang memiliki potensi tinggi untuk menjaring minat siswa untuk yang melanjutkan Pendidikan keperguruan tinggi mereka.

METODE

A. Data mining

Data mining merupakan suatu proses dalam mencari korelasi baru, pola dan trend dengan menggali suatu repository data dalam jumlah yang besar dengan menggunakan statistic dan Teknik matematika (Sugiono, dkk, 2019). Data mining juga sering dikenal dengan pattern recognition merupakan suatu metode yang digunakan untuk pengolahan data guna menemukan pola-pola yang tersembunyi dari sekumpulan data yang diolah. Hasil dari olahan data tersebut dengan menggunakan data mining akan menghasilkan suatu informasi baru atau pengetahuan baru yang bersumber dari data lama yang berguna dalam pengambilan keputusan yang akan datang. Data mining merupakan bidang dari beberapa bidang keilmuan yang menyaktukan Teknik dari pembelajaran mesin, pengenalan pola, statistic, database dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar. Adapun beberapa hal penting yang terkait dengan data mining antara lain (Suryadi.U.T, Supriatna.Y, 2019):

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses merupakan data dalam kumpulan besar.
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

B. Clustering

Clustering adalah salah satu metode yang digunakan dalam data mining. *Clustering* dilakukan dengan memprediksi suatu class yang belum diketahui. *Clustering* dapat dideskripsikan sebagai metode untuk membuktikan sebuah objek data sebagai salah satu jenis yang telah dideskripsikan sebelumnya (Zannah. K, Dkk, 2022).

Tujuan dari *Clustering* data dapat dibedakan menjadi dua, yaitu sebagai berikut (Suryadi.U.T, Supriatna. Y, 2019).

1. Pengelompokan untuk pemahaman terbentuk harus merangkap struktur alami data yang bertujuan hanya sebagai awal proses untuk kemudian dilanjutkan dengan pekerjaan inti seperti peringkasan atau summarization (rata-rata standar deviasi), pemberian label kelas pada setiap kelompok bertujuan untuk digunakan sebagai data latih klasifikasi.
2. Pengelompokan untuk penggunaan bertujuan untuk mencari prototipe kelompok yang paling representative terhadap data, memberikan abstraksi dari setiap objek data dalam kelompok dimana sebuah data terletak di dalamnya.

C. Algoritma K-Means

Algoritma *K-Means* merupakan algoritma non hirarki yang bersal dari metode *Clustering* yang paling sederhana dan mempunyai kemampuan mengelompokan data dalam jumlah yang cukup besar dengan waktu komputasi yang cepat dan efisien metode *K-Means* ini mempartisi data dalam kelompok sehingga data bekarakteristik sama

dimasukan kedalam set kelompok yang sama (Mayona.Y, Buaton.R, Simajuntak.M, 2022).

K-Means merupakan suatu metode pengujian komponen populasi data dan mengelompokan data tersebut kedalam suatu cluster yang telah didefinisikan tergantung dari jarak minimum antar komponen populasi dengan masing-masing pusat cluster. Metode *K-Means Clustering* mengelompokan datan memetakan suatu populasi data ke dalam beberapa kelompok cluster dimana tiap-tiap data dari kelompok cluster memiliki karakteristik yang sama dengan kelompoknya dan berbeda dengan kelompok lainnya. Algoritma dasar dari *K-Means Clustering* dapat kita tentukan dengan Langkah-langkah sebagai berikut (Sugiono, Dkk, 2019):

1. Menentukan jumlah cluster yang diinginkan.
2. Memilih cluster secara random dan mengelompokan data yang lainnya ke dalam cluster-cluster tersebut berdasarkan jarak terpendek.
3. Menghitung *Centroid*/ rata-rata dari data yang ada di hasilkan dari masing-masing cluster.
4. Mengalokasikan Kembali masing-masing data kedalam *Centroid*/ rata-rata cluster terdekat.

$$d = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Keterangan:

- d = Euclidean Distance
i = Banyak objek
x,y = Titik koordinat objek
s,t = Titik koordinat *Centroid*

5. Ulangi Langkah ke-3 apabila masih ditemukan data yang berpindah cluster sehingga menimbulkan perubahan nilai *Centroid* cluster.

D. Rapid Miner

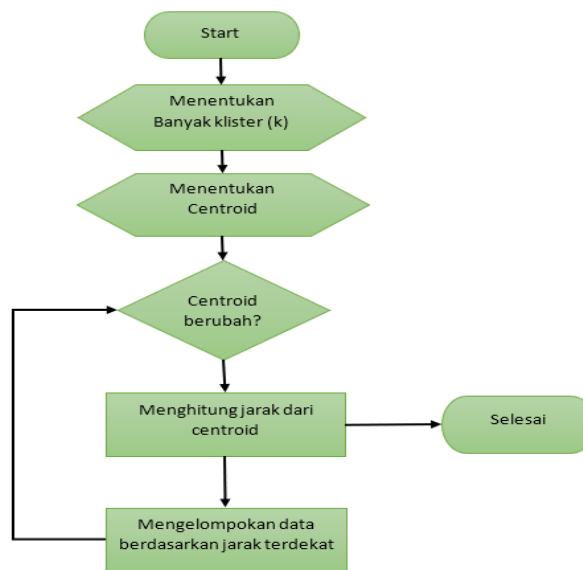
Rapidminer merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Technology Blanchardstown dan Ralf Klinkenberg dari rapid-i.com dengan tampilan GUI (Graphical User Interface) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat open source dan dibuat dengan menggunakan program java dibawah lisensi GNU Public Licence dan rapid miner dapat dijalankan di system operasi manapun (Zannah.K, dkk, 2022). Rapid miner dirancang khusus untuk digunakan untuk analys data dengan data mining. Penggunaan rapid miner tidak diperlukan kemampun dalam coding khusus, karena

semua tool sudah disediakan. Tool yang disediakan juga cukup banyak dan lengkap seperti model Baysian, Modelling, Tree Induction, Neoural Network, dan lain-lain.

HASIL DAN PEMBAHASAN

Implementasi Algoritma *K-Means*

Pada umumnya kinerja metode *K-Means* secara berurutan dapat dilihat pada gambar 1. Diagram alur yang dari metode *K-Means* yang terdapat pada gambar 1 merupakan bagaimana proses pengklusteran antara pendapatan orang tua siswa dengan minat siswa setelah menyelesaikan Pendidikan setingkat Sekolah Menengah Atas (SMA).



Gambar 1. Diagram Flowchart

Pada gambar di atas dapat dijelaskan bahwa dalam proses pengklustersan Langkah pertama yang dilakukan adalah terlebih dahulu ditentukan yaitu menentukan jumlah cluster yang akan dibuat. Setelah jumlah cluster ditentukan barulah masuk ketahap selanjutnya yaitu menentukan *Centroid* dari masing-masing cluster dengan memilih beberapa data yang nantinya akan dijadikan *Centroid* dari masing-masing cluster. Langkah selanjutnya adalah menghitung menghitung jarak dari tiap-tiap data dengan masing-masing *Centroid* cluster yang telah ditentukan sebelumnya. Selanjutnya masuk kelangkah mengelompokkan data dengan masing-masing *Centroid* cluster yang terdekat. Setelah data selesai dikelompokkan maka proses dilanjutkan ke tahap menghitung Kembali titik *Centroid* nya dengan jumlah cluster yang baru terbentuk dari proses sebelumnya. Ketika ditemukan adanya perubahan pada anggota cluster maka kita ulangi lagi Langkah perhitungan jarak antara data dengan *Centroid* yang baru sampai dengan tidak adanya lagi perubahan anggota cluster. Jika sudah tidka ada lagi

perubahan pada anggota cluster maka proses pengclusteran tersebut dinyatakan telah selesai.

Berikut merupakan data yang akan diproses pengelompokan klusternya, berupa kuesioner yang telah di isi oleh siswa kelas XII Madrasah Aliyah Dr. H. Abdullah Ahmad PGAI Padang. Untuk memudahkan dalam proses klusterisasi dilakukan inialisasi pada pekerjaan orang tua, penghasilan orang tua dan minat siswa.

Tabel 1. Data Kuesioner Siswa

No	Nama Lengkap	Pekerjaan Orang Tua	Penghasilan Orang Tua	Minat Setelah Menyelesaikan Pedidikan SMA
1	Wadimah Tussyah'diah Hts	PNS	< Rp. 5.000.000	Kuliah
2	Khairazi Naldi Putra	Wiraswasta	< Rp. 5.000.000	Kuliah
3	Syaiful andri	PNS	< Rp. 5.000.000	Bekerja
4	Hamdan Zoelva	PNS	< Rp. 5.000.000	Bekerja
5	Afdal Zikry	Wiraswasta	< Rp. 3.000.000	Bekerja
6	Ardesmaviona amelya putri	Wiraswasta	< Rp. 2.000.000	Kuliah
7	Fitri suryadi	Ibu rumah tangga	< Rp. 1.000.000	Bekerja
8	Yufri harfandes	Ibu rumah tangga	< Rp. 1.000.000	Kuliah
9	Alfi mukhairin	Wiraswasta	< Rp. 3.000.000	Kuliah
10	Muhammad daffa ramadhan	Swasta	< Rp. 2.000.000	Kuliah
11	Rizkie Abdillah	Petani	< Rp. 1.000.000	Bekerja
12	Mahes saputra	Wiraswasta	> Rp. 5.000.000	Kuliah
13	pedrozaleop	Wiraswasta	> Rp. 5.000.000	Kuliah
14	Qurata Ayuni	Ibu rumah tangga	< Rp. 1.000.000	Kuliah

No	Nama Lengkap	Pekerjaan Orang Tua	Penghasilan Orang Tua	Minat Setelah Menyelesaikan Pendidikan SMA
15	Reski azzamat kalam	Petani	< Rp. 1.000.000	Kuliah
16	Fares saputra	Swasta	< Rp. 2.000.000	Kuliah
17	Hafiz Alim Al-aziz	Petani	< Rp. 1.000.000	Bekerja
18	Hendrimas alfarid	Petani	< Rp. 1.000.000	Bekerja
19	Welmita sari	Ibu rumah tangga	< Rp. 1.000.000	Kuliah
20	Silvina hayati	Petani	< Rp. 1.000.000	Kuliah
21	Kartina	Wiraswasta	< Rp. 3.000.000	Kuliah
22	FADILA JEVITA	Petani	< Rp. 1.000.000	Kuliah
23	May Syahdina Arief	Buruh Lepas	< Rp. 4.000.000	Kuliah
24	May Hendri	Buruh Lepas	< Rp. 2.000.000	Kuliah
25	Novi anjelina	petani	< Rp. 1.000.000	Kuliah

Tabel 2. Inisialisasi Pekerjaan Orang Tua

Kode	Pekerjaan
1	Petani
2	Buruh Lepas
3	Ibu Rumah Tangga
4	Wiraswasta
5	Swasta
6	PNS

Tabel 3. Inisialisasi Pendapatan Orang Tua Siswa

Kode	Penghasilan Orang Tua
1	< Rp. 1.000.000
2	< Rp. 2.000.000
3	< Rp. 3.000.000
4	< Rp. 4.000.000
5	> Rp. 5.000.000

Tabel 4. Inisialisasi Minat Siswa

Kode	Minat Siswa
1	Bekerja
2	Kuliah

Tabel 5. Data yang Telah di Transformasi

No	Pekerjaan Orang Tua (X)	Penghasilan Orang Tua (Y)	Minat Setelah Menyelesaikan Pendidikan SMA (Z)
1	6	5	2
2	4	5	2
3	6	5	1
4	6	5	1
5	4	3	1
6	4	2	2
7	3	1	1
8	3	1	2
9	4	3	2
10	5	2	2
11	1	1	1
12	4	5	2
13	4	5	2
14	3	1	2
15	1	1	2
16	5	2	2
17	1	1	1
18	1	1	1

No	Pekerjaan Orang Tua (X)	Penghasilan Orang Tua (Y)	Minat Setelah Menyelesaikan Pendidikan SMA (Z)
19	3	1	2
20	1	1	2
21	4	3	2
22	1	1	2
23	2	4	2
24	2	2	2
25	1	1	2

Data yang terdapat pada table 5 merupakan data yang telah dilakukan transformasi sesuai dengan ketentuan inialisasi pada masing-masing kolom pada table yang terdapat pada table 2,3 dan 4. Pada tahap selanjutnya ditentukan kluster *Centroid* pada data yang telah ditransformasikan. Pada penelitian ini penulis hanya membuat 2 kluster yaitu data minimal dan maximal dari masing-masing elemen. Berikut ini data yang dijadikan *Centroid* yaitu:

Tabel 6. 2 Kluster Centroid

Kluster	X	Y	Z
Centroid 1 (C1)	1	1	1
Centroid 2 (C2)	6	5	2

Setelah itu dilakukan perhitungan jarak tiap data dengan masing-masing cluster pusat dengan menggunakan persamaan Enclidean Distance dari semua data.

$$d = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Perhitungan Encludian Distance dari semua data ke tiap titik pusat pertama:

$$D_1 = \sqrt{(6 - 1)^2 + (5 - 1)^2 + (2 - 1)^2} = 6,48$$

$$D_2 = \sqrt{(4 - 1)^2 + (5 - 1)^2 + (2 - 1)^2} = 5,10$$

$$D_3 = \sqrt{(6 - 1)^2 + (5 - 1)^2 + (1 - 1)^2} = 6,40$$

$$D_4 = \sqrt{(6 - 1)^2 + (5 - 1)^2 + (1 - 1)^2} = 6,40$$

$$D_5 = \sqrt{(4 - 1)^2 + (3 - 1)^2 + (1 - 1)^2} = 3,61$$

$$D_6 = \sqrt{(4 - 1)^2 + (2 - 1)^2 + (2 - 1)^2} = 3,32$$

$$D7 = \sqrt{(3 - 1)^2 + (1 - 1)^2 + (1 - 1)^2} = 2$$

$$D8 = \sqrt{(3 - 1)^2 + (1 - 1)^2 + (2 - 1)^2} = 2,24$$

$$D9 = \sqrt{(4 - 1)^2 + (3 - 1)^2 + (2 - 1)^2} = 3,74$$

$$D10 = \sqrt{(5 - 1)^2 + (2 - 1)^2 + (2 - 1)^2} = 4,24$$

.....

$$D25 = \sqrt{(1 - 1)^2 + (1 - 1)^2 + (2 - 1)^2} = 1$$

Perhitungan Enclidean Distance dari semua data ke tiap titik pusat kedua.

$$D1 = \sqrt{(6 - 6)^2 + (5 - 5)^2 + (2 - 2)^2} = 0$$

$$D2 = \sqrt{(4 - 6)^2 + (5 - 5)^2 + (2 - 2)^2} = 2$$

$$D3 = \sqrt{(6 - 6)^2 + (5 - 5)^2 + (1 - 2)^2} = 1$$

$$D4 = \sqrt{(6 - 6)^2 + (5 - 5)^2 + (1 - 2)^2} = 1$$

$$D5 = \sqrt{(4 - 6)^2 + (3 - 5)^2 + (1 - 2)^2} = 3$$

$$D6 = \sqrt{(4 - 6)^2 + (2 - 5)^2 + (2 - 2)^2} = 3,61$$

$$D7 = \sqrt{(3 - 6)^2 + (1 - 5)^2 + (1 - 2)^2} = 5,10$$

$$D8 = \sqrt{(3 - 6)^2 + (1 - 5)^2 + (2 - 2)^2} = 5$$

$$D9 = \sqrt{(4 - 6)^2 + (3 - 5)^2 + (2 - 2)^2} = 2,83$$

$$D10 = \sqrt{(5 - 6)^2 + (2 - 5)^2 + (2 - 2)^2} = 3,16$$

.....

$$D25 = \sqrt{(1 - 6)^2 + (1 - 5)^2 + (2 - 2)^2} = 6,40$$

Tabel 7. Hasil Literasi 1

No	(X)	(Y)	(Z)	C1	C2	Keanggotaan
1	6	5	2	6,48	0,00	C2
2	4	5	2	5,10	2,00	C2
3	6	5	1	6,40	1,00	C2
4	6	5	1	6,40	1,00	C2
5	4	3	1	3,61	3,00	C2
6	4	2	2	3,32	3,61	C1
7	3	1	1	2,00	5,10	C1
8	3	1	2	2,24	5,00	C1
9	4	3	2	3,74	2,83	C2
10	5	2	2	4,24	3,16	C2
11	1	1	1	0,00	6,48	C1

No	(X)	(Y)	(Z)	C1	C2	Keanggotaan
12	4	5	2	5,10	2,00	C2
13	4	5	2	5,10	2,00	C2
14	3	1	2	2,24	5,00	C1
15	1	1	2	1,00	6,40	C1
16	5	2	2	4,24	3,16	C2
17	1	1	1	0,00	6,48	C1
18	1	1	1	0,00	6,48	C1
19	3	1	2	2,24	5,00	C1
20	1	1	2	1,00	6,40	C1
21	4	3	2	3,74	2,83	C2
22	1	1	2	1,00	6,40	C1
23	2	4	2	3,32	4,12	C1
24	2	2	2	1,73	5,00	C1
25	1	1	2	1,00	6,40	C1

Dari hasil literasi 1 maka dapat disimpulkan bahwa yang masuk kedalam cluster 1 adalah sebanyak 14 data sedangkan yang masuk ke cluster 2 adalah sebanyak 11 data dari jumlah keseluruhan data 25 data kuesioner siswa. Selanjutnya untuk menentukan *Centroid* baru maka dilakukan penjumlahan data dari banyaknya data yang masuk kedalam masing-masing cluster (C1/C2) lalu dihitung rata-rata dari jumlah data tersebut. Maka didapatkan *Centroid* baru yang dapat dilihat pada table 8. Setelah *Centroid* baru ditemukan Kembali dilakukan literasi 2.

Tabel 8. Centroid Baru

Kluster	X	Y	Z
Centroid 1 (C1)	1	1	1
Centroid 2 (C2)	6	5	2

Tabel 9. Hasil Literasi 2

No	(X)	(Y)	(Z)	C1	C2	Keanggotaan
1	6	5	2	5,47	1,70	C2
2	4	5	2	4,20	1,34	C2
3	6	5	1	5,51	1,83	C2
4	6	5	1	5,51	1,83	C2
5	4	3	1	2,74	1,37	C2

No	(X)	(Y)	(Z)	C1	C2	Keanggotaan
6	4	2	2	2,19	2,06	C2
7	3	1	1	1,34	3,46	C1
8	3	1	2	1,16	3,39	C1
9	4	3	2	2,66	1,20	C2
10	5	2	2	3,15	1,95	C2
11	1	1	1	1,22	4,78	C1
12	4	5	2	4,20	1,34	C2
13	4	5	2	4,20	1,34	C2
14	3	1	2	1,16	3,39	C1
15	1	1	2	1,04	4,74	C1
16	5	2	2	3,15	1,95	C2
17	1	1	1	1,22	4,78	C1
18	1	1	1	1,22	4,78	C1
19	3	1	2	1,16	3,39	C1
20	1	1	2	1,04	4,74	C1
21	4	3	2	2,66	1,20	C2
22	1	1	2	1,04	4,74	C1
23	2	4	2	2,66	2,74	C1
24	2	2	2	0,71	3,34	C1
25	1	1	2	1,04	4,74	C1

Proses pengklusteran terhenti hanya sampai dengan literasi², karena hasil keanggotaan yang terdapat pada literasi 2 yang dapat dilihat pada table 9, sama dengan hasil keanggotaan pada literasi 1 yang terdapat pada table 8. Sehingga tidak perlu lagi dilakukan perhitungan Kembali, maka didapatkan hasil sebanyak 2 group sebagai berikut.

1. Pada *Centroid* 1 (C1) atau *Centroid* dengan nilai minimum terdapat sebanyak 13 anggota, dengan nilai *Centroid* X=1,77, Y=1,31 dan Z=1,69. Yaitu siswa dengan nomor urut (Tabel 1) :7, 8,11,14,15,17,18,19,20,22,23,24,25.
2. Pada *Centroid* 2 (C2) atau *Centroid* dengan nilai maximum terdapat sebanyak 12 anggota, dengan nilai *Centroid* X=4,67, Y=3,75 dan Z=1,75. Yaitu siswa dengan nomor urut (Tabel 1): 1,2,3,4,5,6,9,10,12,13,16,21.

Untuk pembuktian hasil *Clustering* dengan melakukan perhitungan secara manual diatas, selanjutnya kita lakukan pembuktian dengan menggunakan aplikasi rapidminer dengan data yang sama. Dari pengujian dengan menggunakan aplikasi rapidminer yang terdapat pada gambar 2 dan gambar 3, didapatkan 2 cluster. Hanya saja yang membedakan hasil dari pencarian manual dengan rapidminer terletak pada *Centroid*

dengan nilai minimum pada pencarian manual menjadi cluster 1 yang terdiri dari 12 anggota. Sebaliknya *Centroid* dengan nilai maximum menjadi cluster 0 yang terdiri dari 13 anggota. Dan jika diperhatikan hasil dari *Clustering* dengan perhitungan manual dengan menggunakan aplikasi RapidMiner bisa dikatakan 100% sama.

ExampleSet (25 examples, 2 special attributes, 3 regular attributes)

Row No.	id	cluster	(X)	(Y)	(Z)
1	1	cluster_0	6	5	2
2	2	cluster_0	4	5	2
3	3	cluster_0	6	5	1
4	4	cluster_0	6	5	1
5	5	cluster_0	4	3	1
6	6	cluster_0	4	2	2
7	7	cluster_1	3	1	1
8	8	cluster_1	3	1	2
9	9	cluster_0	4	3	2
10	10	cluster_0	5	2	2
11	11	cluster_1	1	1	1
12	12	cluster_0	4	5	2
13	13	cluster_0	4	5	2
14	14	cluster_1	3	1	2
15	15	cluster_1	1	1	2
16	16	cluster_0	5	2	2
17	17	cluster_1	1	1	1
18	18	cluster_1	1	1	1
19	19	cluster_1	3	1	2
20	20	cluster_1	1	1	2
21	21	cluster_0	4	3	2
22	22	cluster_1	1	1	2
23	23	cluster_0	2	4	2
24	24	cluster_1	2	2	2
25	25	cluster_1	1	1	2

```

Cluster 0: 13 items
Cluster 1: 12 items
Total number of items: 25
    
```

Gambar 3. Model Cluster dengan RapidMiner

SIMPULAN

Berdasarkan hasil kuesioner yang di isi oleh siswa kelas XII Madrasah Aliyah Swasta Dr. H. Abdullah Ahmad PGAI Padang, tentang minat siswa setelah menyelesaikan Pendidikan setingkat Sekolah Menengah Atas (SMA), dari sebanyak 25 orang siswa yang mengisi kuesioner. Maka dapat disimpulkan bahwa kecendrungan siswa untuk ingin melanjutkan Pendidikan kejenjang yang lebih tinggi sangat tinggi tinggi, apabila dipersentasekan dari *Clustering* 1 (C1), dimana yang menjadi kelompok cluster 1 adalah siswa dengan penghasilan orang tua < Rp. 1.000.000 yaitu sebanyak 67 % ingin melanjutkan Pendidikan kejenjang yang lebih tinggi dan 33% ingin bekerja. Sedangkan dari cluster 2 (C2), dimana cluster ini merupakan cluster dengan nilai maximum dan bisa dikatakan siswa yang memiliki orang tua yang berpenghasilan > Rp. 5.000.000, terdapat 69% siswa memilih melanjutkan Pendidikan kejenjang yang lebih tinggi dan 31% lebih memilih untuk bekerja.

Adapun saran untuk penelitian selanjutnya:

1. Diharapkan untuk dapat menggunakan data kuesioner yang lebih banyak untuk mendapatkan hasil yang lebih akurat dengan beragamnya data yang diuji untuk meningkatkan akurasi metode klustering dengan algoritma *K-Means*.
2. Menggunakan metode algoritma lain untuk *Clustering* data untuk membandingkan hasil pengujian *Clustering* dengan menggunakan algoritma *K-Means* dengan algoritma lain.
3. Agar hasil *Clustering* ini dapat digunakan oleh orang banyak, perlu adanya pengimplementasian kedalam sebuah Bahasa pemrograman lain, seperti PHP, Visual Basic, dll.

REFERENSI

- Astuti, W., & Widodo, D, A.,(2016). Pemetaan Tindak Kejahatan Jalanan di Kota Semarang Menggunakan Algoritma *K-Means Clustering*. Jurnal Teknik Elktro. Vol.8. No.1.DOI: <https://doi.org/10.15294/jte.v8i1.8788>
- Dewi, S.N., Widarto, A.P., Damanik, I.S., & Satria, H. (2019). Analisa Metode *K-Means* pada Pengelompokan Kriminalitas Menurut Wilayah. Seminar Nasional & Teknologi Informasi (SENSASI). ISBN: 978-602-5272-2-8. Hal: 620-625.
- Dhuhita, W, M, P., (2017). *Clustering* Menggunakan Metode *K-Means* untuk Menentukan Status Gizi Balita. Jurnal Informatika.
- Gultom, D., & dkk., (2020). Penerapan Algoritma *K-Means* untuk Mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar. Jurnal Teknologi Informasi. Vo.4, No.1. DOI: <https://doi.org/10.36294/jurti.v4i1.1263>
- Haviludin, & Dkk. (2021). Implementasi Metode *K-Means* untuk Pengelompokan Rekomendasi Tugas Akhir. Jurnal Ilmiah Ilmu Komputer. Vol. 16, No.1. DOI: <http://dx.doi.org/10.30872/jim.v16i1.5182>

- Priyatma, H., Sajid, F., Haldivany, D., (2019). Klasterisasi Menggunakan Algoritma *K-Means Clustering* untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi dan Penelitian Informatika*. Vol.5, No.1. DOI:[10.26418/jp.v5i1.29611](https://doi.org/10.26418/jp.v5i1.29611)
- Saputro, W., Pahlevi, M. R., Wibowo, A., (2020). Analisis Algoritma *K-Means* untuk Klasterisasi Tindak Pidana Korupsi di Wilayah Hukum Indonesia. *Jurnal Informatika dan Komputer (JIKO)*. Vol.3, No.3. DOI: <http://dx.doi.org/10.33387/jiko.v3i3.1960>
- Sumadikarta, I., Abeiza, E., Penerapan Algoritma *K-Means* pada Data Mining untuk Memilih Produk dan Pelanggan Potensial (Studi Kasus: PT Mega Arvia Utama). *Jurnal Satya Informatika*, Vol.1, No. 1, Hal. 12-22.
- Suryadi, U, T., Supriatna, Y., (2019). Sistem *Clustering* Tindak Kejahatan Pencurian di Wilayah Jawa Barat Menggunakan Algoritma *K-Means*. *Jurnal Teknologi Informasi dan Komunikasi STMIK Subang*. <https://doi.org/10.47561/a.v12i1.147>
- Triyanto, W, A., Suhartono, V., Hirmawan, H. Analisis Keranjang Pasar Menggunakan K-Medoids dan Fp-Growth. *Jurnal Pseudocode*, Vol.2, No.1
- Wijanarko, T.T., Oktavianto, H., Umilasari, R. (2018). Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Kriminalitas Nasional Menggunakan Algoritma *K-Means* dengan Metode Davies Bouldin Index. Departmen Informatic of Engineering.
- Zannah, K., & Dkk. (2022). Model *Clustering* Menggunakan Algoritma K-Mean dalam Menentukan Kriteria Kondisi Gizi Balita dan Anak. *Jurnal Teknik Informatika (JUTEKIN)*. Vol. 10, No.1. DOI: <http://dx.doi.org/10.51530/jutekin.v10i1.568>