

Penerapan *Data Mining* Algoritma *K-Means* Untuk Rekomendasi Pemilihan Bidang Studi Perguruan Tinggi Pada Siswa SMKN 1 Kota Jambi

Vikki Darsono¹, Amroni², Ari Andrianti³

^{1,2}Ilmu Komputer, Teknik Informatika, Universitas Dinamika Bangsa, Jambi, Indonesia

³Ilmu Komputer, Sistem Informasi, Universitas Dinamika Bangsa, Jambi, Indonesia

Email: vikkidarsono2@gmail.com, bh36be@gmail.com, ariandrianti@gmail.com

Abstrak– Dalam memilih jurusan di perguruan tinggi para siswa-siswi SMK Negeri 1 Kota Jambi cenderung memilih jurusan hanya berdasarkan pilihan orang tua, mengikuti pilihan teman atau hanya memilih secara acak tanpa mengetahui jurusan itu sendiri. Para siswa cenderung tidak dapat memperhitungkan potensi yang mereka miliki, sehingga komitmen untuk belajar pada bidang yang sesuai dengan minat mereka tidak berjalan dengan lancar, padahal jurusan yang akan dipilih tidak sesuai dengan potensi yang dimiliki. Oleh karena itu, penulis melakukan analisis *data mining* menggunakan data nilai siswa kelas XII dari semester satu sampai empat dan kuisioner yang penulis bagikan. Dalam melakukan analisis penulis menggunakan alat bantu *tools SPSS*. Metode yang digunakan adalah metode *k-means clustering* dengan data 233 siswa, 21 atribut dan 10 cluster. Jumlah cluster pada perhitungan manual dan menggunakan *tools SPSS* adalah, C1 terdapat 32 data direkomendasikan masuk bidang Keagamaan, C2 terdapat 12 data direkomendasikan masuk bidang Hukum, C3 terdapat 5 data direkomendasikan masuk bidang Sosial dan politik, C4 terdapat 10 data direkomendasikan masuk bidang Komunikasi, C5 terdapat 45 data direkomendasikan masuk bidang Ekonomi dan Bisnis, C6 terdapat 44 data direkomendasikan masuk bidang Pendidikan, C7 terdapat 36 data direkomendasikan masuk bidang Sains dan Teknologi, C8 terdapat 39 data direkomendasikan masuk bidang Seni dan Budaya, C9 terdapat 9 data direkomendasikan masuk bidang Teknik, C10 terdapat 1 data direkomendasikan masuk bidang Kedokteran.

Kata Kunci: *Data Mining*, *K-Means*, Rekomendasi, Bidang Studi Perguruan Tinggi, *SPSS*

Abstract– In choosing a major in higher education, the students of SMK Negeri 1 Jambi City tend to choose a major only based on the choice of their parents, following the choice of friends or just choosing randomly without knowing the major itself. Students tend not to be able to take into account their potential, so that the commitment to study in fields that match their interests does not run smoothly, even though the majors to be chosen are not in accordance with their potential. Therefore, the authors conducted data mining analysis using data on grade XII students from semesters one to four and the questionnaires that the authors distributed. In conducting the analysis, the author uses SPSS tools. The method used is the *k-means clustering* method with data of 233 students, 21 attributes and 10 clusters. The number of clusters in manual calculations and using SPSS tools is, C1 there are 32 data recommended to enter the field of Religion, C2 there are 12 data recommended to enter the field of Law, C3 there are 5 data recommended to enter the field of Social and political, C4 there are 10 data recommended to enter the field of Communication, C5 there are 45 data recommended to enter the field of Economics and Business, C6 there are 44 data recommended to enter the field of Education, C7 there are 36 data recommended to enter the field of Science and Technology, C8 there are 39 data recommended to enter the field of Arts and Culture, C9 there are 9 data recommended to enter the field of Engineering, C10 there is 1 data recommended to enter the field of Medicine.

Keywords: *Data Mining*, *K-Means*, Recommendation, Field of College Studies, *SPSS*

1. PENDAHULUAN

Pada umumnya siswa yang telah lulus dari SMA / SMK / sederajat lainnya akan melanjutkan studi ke Perguruan Tinggi baik Perguruan Tinggi Negeri (PTN) maupun Perguruan Tinggi Swasta (PTS). Pada perguruan tinggi terdapat berbagai bidang studi yang dapat diambil. Banyak siswa yang telah lulus mungkin akan mengalami kebingungan dalam hal memilih bidang studi di perguruan tinggi. Permasalahan ini juga dialami siswa di SMK Negeri 1 Kota Jambi.

Saat akan mengikuti seleksi masuk perguruan tinggi negeri seperti biasanya siswa akan diberi kesempatan memilih tiga program studi berbeda dari dua universitas yang berbeda. Pada saat memilih program studi pertama mungkin siswa SMKN 1 Kota Jambi sedikit tidak merasa bingung karena siswa mungkin akan memilih program studi yang sesuai dengan jurusan mereka sebelumnya di SMKN 1 Kota Jambi, kemudian siswa akan dihadapkan dengan pilihan kedua dan ketiga, disinilah siswa terkadang salah dalam memilih program studi lainnya karena siswa tidak tahu bidang studi yang sesuai dengan kemampuan akademiknya. Begitupula saat siswa akan ke perguruan tinggi swasta, siswa akan dihadapkan dengan banyak pilihan program studi di berbagai perguruan tinggi swasta. Jika siswa memilih salah satu perguruan tinggi swasta yang mereka inginkan, belum tentu ada bidang studi di perguruan tinggi swasta tersebut yang sesuai dengan jurusan siswa sebelumnya di SMK.

Dalam memecahkan permasalahan tersebut penulis menggunakan metode *data mining*. *Data mining* (penambangan data) menjadi sebuah metode yang dapat mengolah arsip data menjadi sebuah informasi atau

pengetahuan (*knowledge*) yang bermanfaat untuk mendukung suatu pengambilan keputusan. *Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, diantaranya yaitu Deskripsi/*Description* (Mengidentifikasi pola dan kecenderungan), Estimasi/*Estimation* (Penargetan), Prediksi/*Prediction* (Nilai yang diperkirakan pada masa yang akan datang), Klasifikasi/*Classification* (Pengkategorian data), Pengklasteran/*Clustering* (Pengelompokan data), Asosiasi/*Association* (Menemukan atribut yang muncul dalam satu waktu). Dalam hal ini penulis memilih menggunakan metode Pengklasteran/*Clustering* (Pengelompokan data). Dikarenakan *Clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode data mining yang bersifat tanpa arahan (*unsupervised*).

Salah satu Algoritma pada *data mining* yang digunakan dalam perhitungan manual dan menggunakan *tools SPSS* pada penelitian ini adalah *k-means clustering*. *K-Means* dapat membagi data kemudian mengelompokkannya kedalam beberapa *cluster* yang memiliki kemiripan dan memisahkan setiap *cluster* berdasarkan perbedaan antar masing-masing *cluster*. Hasil berupa *cluster* data inilah yang dapat digunakan sebagai bahan pertimbangan.

Beberapa penelitian telah membahas tentang topik yang terkait dengan penelitian penulis. Banyak penelitian dilakukan menggunakan teknik data mining untuk menggali berbagai informasi dari sebuah *database*. Algoritma yang digunakan pada penelitian penulis juga telah digunakan di beberapa penelitian. Penelitian sejenis pertama menghasilkan sebuah rekomendasi kelompok penerima bantuan beasiswa PPA di UDINUS. [1] Penelitian kedua memberi informasi mengenai pengelompokan penduduk miskin pada Kecamatan Jambi Selatan dengan 3 kategori yaitu hampir miskin, miskin, dan sangat miskin. [2] Penelitian ketiga menghasilkan suatu aplikasi yang bisa memilah data dengan cepat untuk siapa saja yang layak menerima bantuan BPNT. [3] Penelitian keempat menghasilkan sistem yang dapat mengelompokkan siswa ke kelas – kelas unggulan yang ada pada sekolah tersebut. [4] Dan pada penelitian kelima menghasilkan rekomendasi siswa yang kompeten untuk mengikuti OSN dari pengolahan semua data akademik siswa tersebut. [5]

Dari penelitian sejenis tersebut dapat disimpulkan bahwa yang membedakan penelitian ini dengan penelitian sebelumnya adalah penerapan metode *clustering K-Means* pada kasus yang berbeda. Sementara penelitian penulis akan merekomendasikan penentuan bidang studi perguruan tinggi untuk siswa/i SMK Negeri 1 Kota Jambi dengan cara mengolah data nilai dari SMK Negeri 1 Kota Jambi serta dataset siswa/i terdiri dari Jurusan, Hobi (Seni, Olahraga, Teknologi, Kewirausahaan, Komunikasi), Minat (*Science/Ilmiah*, Teknologi, Seni, Olahraga), Bakat dan seluruh nilai mata pelajaran dari semester satu sampai empat di SMK Negeri 1 Kota Jambi yang telah di rata-ratakan, penelitian penulis menggunakan 233 buah data dengan 21 atribut dan menggunakan *tools SPSS (Statistical Product for Service Solutions)* untuk hasil analisis dan visualisasinya. Dengan adanya penelitian ini diharapkan tingkat persentase kelulusan dalam menentukan bidang studi perguruan tinggi pada siswa/i SMK Negeri 1 Kota Jambi lebih tinggi, berdasarkan atribut terpilih yang kemudian di analisa dengan metode *K-Means Clustering*. Hasil yang diperoleh dapat dijadikan sebagai acuan dalam menentukan bidang studi perguruan tinggi pada siswa/i kelas XII di SMK Negeri 1 Kota Jambi.

Dalam hal ini, rekomendasi pemilihan bidang studi di perguruan tinggi ini dapat membantu siswa dalam memilih bidang studi di perguruan tinggi negeri maupun swasta. Hasil penelitian ini diharapkan dapat membantu pihak sekolah maupun siswa dalam menentukan bidang studi apa yang akan di pilih oleh siswa dalam melanjutkan pendidikan keperguruan tinggi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Untuk membantu dalam penyusunan penelitian ini, maka perlu adanya tahapan – tahapan yang jelas. Tahapan penelitian ini merupakan langkah – langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas. Maka dapat diuraikan pembahasan masing – masing tahapan dalam penelitian adalah sebagai berikut :

1. Identifikasi Masalah

Tahap pertama dalam penelitian adalah mengidentifikasi masalah yang terdapat pada SMK Negeri 1 Kota Jambi adalah kebingungan siswa dalam hal memilih bidang studi di perguruan tinggi. Masalah - masalah yang ada akan dipecahkan menggunakan data mining dengan algoritma *K-Means Clustering*.

2. Studi Literatur

Tahap kedua dalam penulisan penelitian ini adalah studi literatur yang digunakan untuk mengumpulkan bahan referensi mengenai data mining terutama bagaimana sebuah Algoritma *K-Means* bisa digunakan dalam

merekomendasikan keputusan, dikumpulkan dari berbagai buku, *e-book*, jurnal, artikel dan beberapa referensi lainnya yang ada dipergustakaan UNAMA JAMBI.

- a. Pengertian Penerapan
Medana Pratama dkk menyatakan bahwa “Penerapan merupakan proses untuk mengetahui atau memahami nilai bobot suatu zat, kata dan lain sebagainya di sini penerapan dapat diartikan sebagai penentuan nilai ukuran file yang akan ditetapkan pada suatu algoritma”. [6]
 - b. Pengertian *Data Mining*
Menurut Rusdiana dkk Menyatakan bahwa :
“Data adalah fakta atau bagian dari fakta yang mengandung arti sehubungan dengan kenyataan, symbol – symbol, gambar – gambar, angka – angka, huruf, atau simbol yang menunjukkan suatu ide, objek, kondisi, atau situasi dan lainnya yang didapatkan melalui suatu observasi atau secara data diartikan sebagai keterangan tentang sesuatu”. [7] Retno Tri Vulandari menyatakan bahwa “*Data mining* adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data”. [8] Alfa Saleh menyatakan bahwa “*Data mining* merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi – informasi yang nantinya dapat digunakan”. [9]
 - c. Pengertian *Clustering*
Roni Setiawan menyatakan bahwa: “*Clustering* merupakan salah satu metode data mining yang bersifat tanpa arahan (*unsupervised*) yang memiliki arti bahwa karakteristik tiap *cluster* tidak ditentukan sebelumnya melainkan berdasarkan kemiripan atribut – atribut dari suatu kelompok atau *cluster*”. [10]
Purnamaningsih dkk menyatakan bahwa “*Clustering* merupakan proses pengelompokan sekumpulan objek fisik ataupun abstrak kedalam kelas – kelas yang mempunyai kemiripan”. [11]
 - d. Pengertian Algoritma *K-Means*
Nasution menyatakan bahwa “Algoritma *K-Means* adalah algoritma *clustering* yang paling populer dan banyak digunakan dalam dunia industri. Algoritma ini disusun atas dasar ide yang sederhana”. [12]
Fauziah Nur, M. Zarlis meyatakan bahwa “Algoritma *K-Means* merupakan algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah *K cluster* yang sudah ditetapkan di awal”. [13]
 - e. Pengertian Bidang Studi
Menurut Kamus Besar Bahasa Indonesia Bidang Studi adalah pengelompokan sejumlah mata pelajaran yang sejenis atau memiliki ciri yang sama (mata pelajaran yang telah berkorelasi satu dengan yang lain). [14]
 - f. Pengertian Perguruan Tinggi
Menurut UU No. 22 Tahun 1961 Pasal 1 “Perguruan tinggi adalah lembaga ilmiah yang mempunyai tugas menyelenggarakan pendidikan dan pengajaran di atas perguruan tingkat menengah, dan yang memberikan pendidikan dan pengajaran berdasarkan kebudayaan kebangsaan Indonesia dan dengan cara ilmiah”. [15]
 - g. *SPSS (Statistical Product for Service Solutions)*
Menurut Hasyim dan Listiawan *SPSS (Statistical Product for Service Solutions*, dulunya *Statistical Packedge for Social Sciences*) merupakan program komputer statistik yang mampu memproses data statistik secara cepat dan akurat. *SPSS* menjadi sangat populer karena memiliki bentuk pemaparan yang baik (berbentuk grafik dan tabel), bersifat dinamis (mudah dilakukan perubahan data dan *update* analisis) serta mudah dihubungkan dengan aplikasi lain (misalnya ekspor/impor data ke/dari *Excel*). [16]
3. Pengumpulan Data
Pada pengumpulan data, penulis melakukan pengamatan langsung ke lapangan, menyebarkan kuesioner dan melakukan wawancara kepada pihak sekolah SMK Negeri 1 Kota Jambi.
 4. Perhitungan dan Analisis Hasil Dengan Metode Algoritma *K-Means*
Pada tahap ini dilakukan analisa data dari hasil observasi, wawancara, dan kuesioner yang telah dilakukan sebelumnya. Data – data yang telah dikumpulkan oleh penulis selanjutnya akan dilakukan perhitungan dan analisis dengan metode *K-Means clustering* secara manual menggunakan bantuan aplikasi *Microsoft Excel*. Selanjutnya penulis melakukan evaluasi dan menganalisis hasil analisis perhitungan penulis dengan hasil perhitungan menggunakan *tools SPSS (Statistical Product for Service Solutions)*.
 5. Pembuatan Laporan
Pada tahapan terakhir dari proses penelitian ini adalah penyusunan laporan penelitian dengan menggunakan teknik pengumpulan data primer (hasil observasi, wawancara serta data nilai rapor siswa dan kuesioner) dan data sekunder (studi - studi yang telah dilakukan sebelumnya yang relevan dengan penelitian ini).

2.2 Metode Analisa Data

Metode yang digunakan adalah metode *K-Means clustering* yang mana algoritma ini mempunyai kelebihan yaitu mudah diterapkan dan dijalankan, relatif cepat, mudah untuk diadaptasi, dan paling banyak dipraktekkan dalam tugas *data mining*. Metode *K-Means clustering* membagi data kemudian mengelompokkannya kedalam

beberapa *cluster*/kelompok yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan memisahkan setiap *cluster* berdasarkan perbedaan antar masing-masing *cluster*.

1. Proses Pengolahan *Data Mining*

Menurut Ridwan dkk. Beberapa tahapan yang dimiliki *data mining* berdasarkan tugas yang bisa dilakukan yaitu :

- a. Pembersihan data (*data cleaning*)
Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau tidak relevan.
- b. Integrasi data (*data integration*)
Integrasi data merupakan penggabungan data dari berbagai *database* kedalam satu *database* baru.
- c. 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*.
- d. Transformasi dari *database*
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*.
- e. Proses mining
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
- f. Evaluasi pola (*patren evaluation*)
Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan.
- g. Presentasi pengetahuan (*knowledge presentation*)
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.[17]

2. Proses Perhitungan Menggunakan *K-Means*

Algoritma standar dari *K-Means* adalah sebagai berikut:

1. Langkah 1 : Tentukan jumlah *clustering* yang diinginkan (misalkan : $C = 3$).
2. Langkah 2 : Pilih *centroid* awal secara acak. Pada langkah ini secara acak akan dipilih 3 buah data sebagai *centroid*.
3. Langkah 3 : Hitung jarak dengan *centroid*.
Pada langkah ini setiap data akan ditentukan *centroid* terdekatnya, dan data tersebut akan ditetapkan sebagai anggota kelompok yang terdekat dengan *centroid*.
Untuk menghitung jarak ke *centroid* masing-masing *cluster* Misalkan data (x,y) , *centroid* M1 : (a_1,b_1) , *centroid* M2 : (a_2,b_2) , *centroid* M3 : (a_3,b_3) .

$$DM1 = \sqrt{(x - a_1)^2 + (y - b_1)^2} = ? \quad (1)$$

$$DM2 = \sqrt{(x - a_2)^2 + (y - b_2)^2} = ? \quad (2)$$

$$DM3 = \sqrt{(x - a_3)^2 + (y - b_3)^2} = ? \quad (3)$$

Buat tabel hasil perhitungan jarak selengkapnya antara masing-masing data dengan *centroid*, maka di dapatkan keanggotaan dari masing – masing *cluster*.

Pada langkah ini dihitung pula rasio antara *BCV* (*Between Cluster Variation*) dengan *WCV* (*Within Cluster Variation*) :

Karena *centroid* M1 = (a_1,b_1) , M2 = (a_2,b_2) , M3 = (a_3,b_3) .

$$d(m_1, m_2) = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2} = ? \quad (4)$$

$$d(m_1, m_3) = \sqrt{(a_1 - a_3)^2 + (b_1 - b_3)^2} = ? \quad (5)$$

$$d(m_2, m_3) = \sqrt{(a_2 - a_3)^2 + (b_2 - b_3)^2} = ? \quad (6)$$

$$BCV = d(m_1,m_2) + d(m_1,m_3) + d(m_2,m_3) = ? \quad (7)$$

Dalam hal ini $d(m_i,m_j)$ menyatakan jarak *euclidean* dari m ke m_j Menghitung *WCV* yaitu dengan memilih jarak terkecil yang terdapat pada tabel keanggotaan.

$$WCV = c^1+c^2+c^3+ N = ? \quad (8)$$

$$\text{Sehingga besar rasio} = BCV/WCV = ? \quad (9)$$

Karena langkah ini merupakan iterasi 1 maka lanjutan ke langkah berikutnya.

4. Langkah 4 : pembaruan *centroid* dengan menghitung rata – rata nilai pada masing – masing *cluster*. Setelah menghitung rata-rata nilai pada masing-masing *cluster* didapatkan *centroid* baru yaitu : $M1 = (a1,b1)$, $M2 = (a2,b2)$, $M3 = (a3,b3)$.
5. Langkah 5 : (iterasi ke 2) kembali kelangkah ke 3, jika masih ada data yang berpindah *cluster* atau jika nilai *centroid* diatas ambang, atau jika pada fungsi objektif yang digunakan masih diatas ambang. Jika tidak maka iterasi dihentikan. [18]

3. HASIL DAN PEMBAHASAN

3.1 Analisis dan Perhitungan Data Mining Menggunakan K-Means Clustering

Pada tahap ini penulis akan mengolah data siswa-siswi SMK Negeri 1 Kota Jambi dengan menggunakan teknik *clustering k-means* secara manual menggunakan bantuan *tools Microsoft Excel*. Kemudian akan diuji dan dibandingkan dengan *tools* yang di gunakan, yaitu *tools SPSS (Statistical Product for Service Solutions)*.

3.1.1 Representasi Data

Penelitian ini menggunakan menggunakan 233 data siswa kelas XII SMK Negeri 1 Kota Jambi pada semester 1 sampai 4 atau T.A. 2021/2022. Maka didapatkan beberapa data siswa kelas XII SMK Negeri 1 Kota Jambi yang telah terpilih dan mengisi kuesioner. Dibawah ini adalah gambar tabel data siswa.

Tabel 1. Data Siswa/Siswi SMKN 1 Kota Jambi Kelas XII T.A. 2021/2022

No.	Nama Lengkap	Jenis Kelamin	Kelas-Jurusan	Kesulitan dalam memilih Bidang Studi	Memilih Bidang Studi berdasarkan kemampuan akademik, hobi, minat dan bakat	Siapa yang memutuskan pemilihan Bidang Studi	Perlu rekomendasi	HOBI			MINAT			BAKAT				Pelajaran Jurusan			
								Hobi Bidang Seni	...	Hobi Bidang Komunikasi	Minat Bidang Seni	...	Minat Bidang Sains/Ilmiah	1	2	3	4	1	2	3	4
1	Agus Purnama	L	XII AP	Ya	Ya	Diri Sendiri	Ya	Lainnya	Public Speaking	Musik	Lainnya	Public Speaking	86	88	90	95	81	84	84	81	
24	Zahwa radisty	P	XII AP	Tidak	Ya	Diri Sendiri	Ya	Musik	Bahasa Asing	Musik	Lainnya	Voly	83	83	80	90	82	85	86	38	
25	Abela Fitri Rahman	P	XII AK	Tidak	Ya	Diri Sendiri	Ya	Benyanyi	Public Speaking	Benyanyi	Matematika	Public Speaking	90	90	88	89	87	88	91	90	
68	Vannya Puspa	P	XII AK	Ya	Ya	Diri Sendiri	Ya	Musik	tidak keduanya	Menari	IPS	Lainnya	80	85	84	84	82	79	77	81	
69	Alya Ramadhani fitri	P	XII MM	Ya	Ya	Diri Sendiri	Tidak	Menari	Public Speaking	Menari	Lainnya	Voly	85	77	78	83	80	79	86	81	
85	Tri Rizki putra	L	XII MM	Tidak	Ya	Diri Sendiri	Ya	Melukis	Public Speaking	Melukis	Lainnya	Kuliner	75	80	75	80	84	82	88	89	
86	ADELIA	P	XII TN	Ya	Ya	Diri Sendiri	Ya	Musik	Public Speaking	Musik	IPA	Public Speaking	85	83	50	80	81	80	79	78	
141	Zulkipi nurdin	L	XII TN	Ya	Ya	Diri Sendiri	Ya	Musik	Public Speaking	Musik	Lainnya	Otomotif	75	84	80	80	82	83	82	85	
142	Abdhe Febrian	L	XII TKJ	Tidak	Ya	Diri Sendiri	Ya	Musik	Lainnya	Musik	IPS	Game	79	75	78	66	82	83	84	83	
233	Yohanes Roma harapan	L	XII TKJ	Ya	Ya	Diri Sendiri	Tidak	Benyanyi	Public Speaking	Benyanyi	Lainnya	Futsal	85	77	79	75	80	80	78	78	

3.1.2 Transformasi Data

Data siswa/siswi di SMK Negeri 1 Kota Jambi penulis menggunakan 21 atribut yang akan dipakai yaitu: Jurusan, Hobi (Seni, Olahraga, Teknologi, Kewirausahaan, Komunikasi), Minat (Seni, Olahraga, Teknologi, Science/Ilmiah), Bakat dan nilai rata – rata dari mata pelajaran siswa, yaitu : Pendidikan Agama dan Budi Pekerti, Pendidikan Pancasila dan Kewarganegaraan, Bahasa Indonesia, Bahasa Inggris, Matematika, Penjas Orkes, Seni Budaya, Simulasi Digital dan Komunikasi Data, Produk Kreatif dan Kewirausahaan, dan Pelajaran Jurusan. Agar data di atas dapat diolah dengan menggunakan metode *k-means clustering*, maka data yang berjenis data nominal seperti hobi, minat, dan bakat harus diinisialisasikan terlebih dahulu dalam bentuk angka dan nilai rapor dari semester 1 sampai 4 pada setiap mata pelajaran di rata – ratakan. Dapat dilihat dibawah ini.

Tabel 2. Inisialisasi Data

1. Jurusan			2. Hobi			3. Minat			4. Bakat					
No	Nama Jurusan	Inisial	No	Bidang	Nama Hobi	Inisial	No	Bidang	Nama Minat	Inisial	No	Nama Bakat	Inisial	
1	Administrasi Perkantoran (AP)	1	1		Musik	1	1		Musik	1	1	Akademik	1	
2	Akuntansi (AK)	2	2		Bernyanyi	2	2		Bernyanyi	2	2	Musik	2	
3	Multimedia (MM)	3	3	Seni	Menari	3	3	Seni	Menari	3	3	Bernyanyi	3	
4	Tata Niaga (TN)	4	4		Melukis	4	4		4	Melukis	4	4	Menari	4
5	Teknik Komputer dan Jaringan (TKJ)	5	5		Teater	5	5		5	Teater	5	5	Melukis	5
			6		Lainnya	6	6		6	Lainnya	6	6	Teater	6
			7		Sepak bola	7	7		7	Sepak bola	7	7	Sepak bola	7
			8	Futsal	8	8	8	Futsal	8	8	Futsal	8		
			9	Voly	9	9	9	Voly	9	9	Voly	9		
			10	Bela diri	10	10	10	Bela diri	10	10	Bela diri	10		
			11	Olahraga	Basket	11	11	Olahraga	Basket	11	11	Basket	11	
			12		Badminton	12	12		12	Badminton	12	12	Badminton	12
			13		Renang	13	13		13	Renang	13	13	Renang	13
			14		Lari	14	14		14	Lari	14	14	Lari	14
			15		Lainnya	15	15		15	Lainnya	15	15	Game	15
			16	Teknologi	Game	16	16	Teknologi	Game	16	16	Photography	16	
			17		Photography	17	17		17	Photography	17	17	Videography	17
			18		Videography	18	18		18	Videography	18	18	Desain Grafis	18
			19		Desain Grafis	19	19		19	Desain Grafis	19	19	Kuliner	19
			20		Lainnya	20	20		20	Lainnya	20	20	Fashion	20
			21	Kewirausahaan	Kuliner	21	21	Sains/Ilmiah	Matematika	21	21	Elektronik	21	
			22		Fashion	22	22		22	IPA	22	22	Okomotif	22
			23		Elektronik	23	23		23	IPS	23	23	Public Speaki	23
			24		Okomotif	24	24		24	Kedokteran	24	24	Bahasa Asing	24
			25		Lainnya	25	25		25	Lainnya	25	25	Lainnya	25
			26	Komunikasi	Public Speaki	26	26							
			27		Bahasa Asing	27	27							
			28		Lainnya	28	28							

Keterangan Tabel :

- | | |
|--|--|
| 1. J = Jurusan | 13. PKN = Pendidikan Pancasila dan Kewarganegaraan |
| 2. HBS = Hobi Bidang Seni | 14. Bina = Bahasa Indonesia |
| 3. HBO = Hobi Bidang Olahraga | 15. Bing = Bahasa Inggris |
| 4. HBT = Hobi Bidang Teknologi | 16. MTK = Matematika |
| 5. HBKwu = Hobi Bidang Kewirausahaan | 17. PJOK = Pendidikan Jasmani, Olahraga, dan Kesehatan |
| 6. HBKom = Hobi Bidang Komunikasi | 18. SB = Seni Budaya |
| 7. MBS = Minat Bidang Seni | 19. SD&KD = Simulasi Digital Dan Komunikasi Data |
| 8. MBO = Minat Bidang Olahraga | 20. PK = Produk Kreatif dan Kewirausahaan |
| 9. MBT = Minat Bidang Teknologi | 21. PJ = Pelajaran Jurusan |
| 10. MBSI = Minat Bidang Sains/Ilmiah | |
| 11. B = Bakat | |
| 12. PAg&BP = Pendidikan Agama dan Budi Pekerti | |

Tabel 3. Hasil Transformasi Data

No	J	HBS	HBO	HBT	HBKwu	HBKom	MBS	MBO	MBT	MBSI	B	PAg & BP	PKN	Bina	Bing	MTK	PJOK	SB	SD&KD	PK	PJ
1	1	6	2	4	5	1	1	2	4	5	23	90	85	86	82	83	86	90	78	78	82
24	1	1	3	1	2	2	1	3	1	5	9	84	81	85	79	80	86	88	83	82	85
25	2	2	7	3	3	1	2	7	3	1	23	89	83	89	85	84	84	91	90	95	89
68	2	1	8	1	1	3	3	7	2	3	25	83	81	84	85	85	83	94	84	78	80
69	3	3	3	1	2	1	3	3	2	5	9	81	79	80	78	78	80	81	90	78	81
85	3	4	4	2	1	1	4	4	2	5	19	78	78	80	77	79	78	75	83	85	86
86	4	1	6	4	1	1	1	6	4	2	23	75	78	80	79	81	83	85	79	82	79
141	4	1	6	1	4	1	1	6	1	5	22	80	78	79	79	79	80	91	79	88	83
142	5	1	2	1	1	3	1	2	1	3	15	75	79	79	76	76	78	77	81	87	83
233	5	2	1	1	4	1	2	2	1	5	8	79	45	79	78	70	83	82	81	75	79

3.1.3 Perhitungan Manual K-Means Clustering

Pertama pada tahap awal menentukan jumlah cluster yang diinginkan (C = 10). maka penulis merekomendasikan 10 cluster yang mana untuk cluster 1 siswa/siswi SMK Negeri 1 Kota Jambi direkomendasikan masuk Bidang Keagamaan, cluster 2 direkomendasikan masuk Bidang Hukum, cluster 3 direkomendasikan masuk Bidang Sosial dan Politik, cluster 4 direkomendasikan masuk Bidang Komunikasi, cluster 5 direkomendasikan masuk Bidang Ekonomi dan Bisnis, cluster 6 direkomendasikan masuk Bidang Pendidikan, cluster 7 masuk Bidang Sains dan Teknologi, cluster 8 direkomendasikan masuk Bidang Seni dan Budaya, cluster 9 direkomendasikan masuk Bidang Teknik, cluster 10 direkomendasikan masuk Bidang Kedokteran. Selanjutnya memilih centroid awal secara acak. Pada langkah ini secara acak akan dipilih 10 buah data sebagai centroid dari 233 data, data {88, 108, 185, 103, 29, 6, 120, 216, 226, 201}. Seperti gambar di bawah ini.

Sehingga dari tabel 8. diatas didapatkan *centroid baru* diambil dari nilai rata – rata masing – masing cluster yaitu pada tabel berikut ini :

Tabel 8. *Centroid Baru* (Untuk Perhitungan Iterasi ke-2)

Centroid1	2,9	2,6	5,7	2,0	1,8	1,5	2,6	5,6	2,1	3,4	14,6	83,2	81,3	83,8	80,8	81,3	83,0	87,9	82,1	84,1	82,2
Centroid2	4,4	2,0	3,5	1,8	2,6	1,3	2,0	3,5	1,6	3,3	19,0	65,0	69,6	77,9	76,8	76,1	68,0	77,5	79,4	77,4	77,3
Centroid3	3,8	3,3	5,0	1,3	2,0	1,0	2,8	4,5	2,0	3,3	11,8	81,5	78,5	81,5	78,8	82,5	84,5	86,5	80,0	40,5	70,3
Centroid4	3,6	2,0	3,6	1,8	2,6	1,8	3,2	3,6	1,8	3,8	8,2	66,0	71,0	74,2	67,4	76,2	68,4	78,4	81,4	86,4	83,8
Centroid5	3,6	3,3	4,4	2,3	1,8	1,6	3,1	4,7	2,3	3,6	6,3	74,6	78,3	79,7	79,6	78,8	82,2	84,4	82,3	84,6	82,8
Centroid6	4,0	2,8	5,5	1,7	2,1	1,4	2,2	5,8	1,8	3,5	14,1	77,3	76,6	77,9	76,7	76,4	78,8	79,8	83,2	82,0	81,6
Centroid7	4,0	2,8	6,5	1,7	1,8	1,6	2,6	6,5	1,9	3,3	18,6	79,9	78,5	82,1	78,9	79,6	81,8	83,4	80,9	72,3	76,4
Centroid8	4,5	3,0	3,9	1,5	2,8	1,2	2,4	4,3	1,5	3,7	9,1	70,5	71,5	79,5	76,5	74,7	79,1	78,4	80,7	78,0	77,4
Centroid9	5,0	3,0	1,6	1,6	3,2	1,6	3,0	1,8	1,6	3,8	9,4	75,4	46,0	79,6	72,8	71,2	77,6	79,4	80,4	75,8	79,4
Centroid10	5,0	6,0	2,0	1,0	2,0	3,0	6,0	2,0	1,0	5,0	8,0	60,0	59,0	59,0	59,0	58,0	78,0	81,0	86,0	78,0	80,0

Selanjutnya perhitungan kembali kelangkah menghitung jarak ke *centroid* menggunakan *centroid* baru (Iterasi 2), jika masih ada data yang berpindah iterasi dilanjutkan, jika data sudah tidak berpindah lagi maka iterasi dihentikan (dikarenakan pada iterasi 1 masih perhitungan awal, jadi belum ada data yang berpindah). Selanjutnya pada langkah ini dilakukan penempatan lagi data dalam *centroid* terdekat sama seperti yang dilakukan dilangkah sebelumnya. Untuk menghitung jarak ke *centroid* masing – masing pada data siswa/siswi No. 1 sampai No. 233 SMK Negeri 1 Kota Jambi. Semua langkah perhitungan ini dilakukan kembali hingga iterasi ke-15.

Sehingga didapatkan perbandingan jumlah pengelompokkan pada setiap iterasi, rasio setiap iterasi, centroid final dan hasil akhir perhitungan manual sebagai berikut :

Tabel 9. Jumlah Pengelompokkan Setiap Iterasi

Pengelompokkan	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	Jumlah
Iterasi 1	95	8	4	5	19	35	30	31	5	1	233
Iterasi 2	76	12	5	6	32	38	31	26	6	1	233
Iterasi 3	64	12	5	6	44	39	30	24	8	1	233
Iterasi 4	55	12	5	6	50	40	31	25	8	1	233
Iterasi 5	45	12	5	6	56	41	32	27	8	1	233
Iterasi 6	41	12	5	6	56	42	31	31	8	1	233
Iterasi 7	40	12	5	6	50	43	32	36	8	1	233
Iterasi 8	40	12	5	7	47	44	31	38	8	1	233
Iterasi 9	40	12	5	8	45	44	30	39	9	1	233
Iterasi 10	40	12	5	9	45	44	29	39	9	1	233
Iterasi 11	37	12	5	9	45	44	32	39	9	1	233
Iterasi 12	35	12	5	9	45	44	33	40	9	1	233
Iterasi 13	33	12	5	10	45	44	35	39	9	1	233
Iterasi 14	32	12	5	10	45	44	36	39	9	1	233
Iterasi 15	32	12	5	10	45	44	36	39	9	1	233

Tabel 10. Rasio Setiap Iterasi

Nilai	Iterasi 1	Iterasi 2	Iterasi 3	Iterasi 4	Iterasi 5	Iterasi 6	Iterasi 7	Iterasi 8	Iterasi 9	Iterasi 10	Iterasi 11	Iterasi 12	Iterasi 13	Iterasi 14	Iterasi 15
BCV	1934,5	1429,3	1399,8	1404,3	1413,8	1425,4	1432,6	1439,3	1440,2	1432,2	1431,1	1434,3	1437,2	1434,3	1435
WCV	95251	44850	41693	40483	40087	39369	39125	38895	38538	38228	38132	38020	37958	37890	37878
Rasio	0,0203	0,0319	0,0336	0,0347	0,0353	0,0362	0,0366	0,037	0,0374	0,0375	0,0375	0,0377	0,0379	0,0379	0,0379

Politik, C4 sebanyak 10 siswa direkomendasikan masuk Bidang Komunikasi, C5 sebanyak 45 siswa direkomendasikan ke Bidang Ekonomi dan Bisnis, C6 sebanyak 44 siswa direkomendasikan masuk Bidang Pendidikan, C7 sebanyak 36 siswa direkomendasikan masuk Bidang Sains dan Teknologi, C8 sebanyak 39 siswa direkomendasikan masuk Bidang Seni dan Budaya, C9 sebanyak 9 siswa direkomendasikan masuk Bidang Teknik, C10 sebanyak 1 siswa direkomendasikan masuk Bidang Kedokteran dan Kesehatan.

REFERENCES

- [1] A. Salam, D. Adiatma, and J. Zeniarja, "Implementasi Algoritma K-Means Dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS," *JOINS (Journal Inf. Syst.*, vol. 5, pp. 62–68, 2020.
- [2] D. Sunia, Kurniabudi, and P. . Jusia, "PENERAPAN DATA Min. UNTUK Clust. DATA Pendud. MISKIN MENGGUNAKAN Algoritm. K-MEANS," vol. 53, pp. 1689–1699, 2019, doi: 10.14710/jtsiskom.3.2.2015.247-252.
- [3] S. Ghousi, P. Ali, and I. . Achmadi, "Klasifikasi Penentuan Penerima Bantuan Pangan Non Tunai Menggunakan Metode K-Means Clustering," pp. 1–8, 2020.
- [4] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, p. 25, 2021.
- [5] W. Sirait, S. Defit, and G. W. Nurcahyo, "Implementasi Algoritma K-Means untuk Klasterisasi Peserta Olimpiade Sains Nasional Tingkat SMA," *J. Sistim Inf. dan Teknol.*, vol. 1, pp. 25–30, 2019.
- [6] A. Medana Pratama, N. Astuti Hasibuan, and E. Buulolo, "Penerapan Algoritma Huffman Dan Shannon-Fano Dalam Pemampatan File Teks," *Maj. Ilm. INTI*, vol. 5, 2017.
- [7] H. A. Rusdiana, M. M. Moch, S. T. Irfan, and H. M. A. Ramdhadi, "Sistem Informasi Manajemen Sistem Informasi Manajemen Pustaka Setia Pengantar: Penerbit PUSTAKA SETIA Bandung," 2014.
- [8] R. T. Wulandari, "Data Mining," *Yogyakarta Gaya Media*, 2017.
- [9] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Citec Journal.*, vol. 2, no. 3, 2015.
- [10] R. Setiawan, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik Lp3i Jakarta)," vol. 3, no. 1, 2016, doi: 10.21831/elinvo.v2i2.17314.
- [11] C. Purnamaningsih, R. Saptono, A. Aziz, and K. Kunci, "Pemanfaatan Metode K-Means Clustering dalam Penentuan Penjurusan Siswa SMA," vol. 3, no. 1, p. 2014, 2014.
- [12] M. Nasution, "mplementasi Data Mining K-Means Untuk Mengukur Kemampuan Logika Mahasiswa (Studi Kasus : Amik Labuhan Batu)," 2017.
- [13] F. Nur, P. M. Zarlis, and D. B. B. N, "Penerapan Algoritma K-Means Pada Siswa Baru Sekolah menengah Kejuruan Untuk Clustering Jurusan, 100–105," pp. 100–105, 2017, doi: 10.1038/132817a0.
- [14] "Kamus Besar Bahasa Indonesia (KBBI) Kamus versi online/daring (dalam jaringan," doi: 10.33633/joins.v4i2.3054.
- [15] "Republik Indonesia. Undang – Undang Nomor 22 Tahun 1961 Pasal 1 Tentang Perguruan Tinggi."
- [16] Y.- Hasyim, "Implementasi Sistem Navigasi Robot Wall Following Dengan Metode Fuzzy Logic Untuk Robot Pemadam Api Divisi Berkaki Onix Ii Pada Krpai Tahun 2017," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 2, no. 1, pp. 26–31, 2017, doi: 10.29100/jipi.v2i1.181.
- [17] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademi Mahsiswa Menggunakan Algoritma Naive Bayes Classifier," vol. 7, no. 1, 2013, doi: 10.9744/informatika.9.2.115-121.
- [18] F. M. Irfan, "Penerapan Clustering Data Mining Untuk Rekomendasi Penentuan Jurusan Perguruan Tinggi Pada Siswa SMA Negeri 2 Kota Jambi Menggunakan Metode K-means," 2019.