

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 Vuldari 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	Pendidikan Agama dan Budi Pekerti				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 4. Pemilihan Centroid Awal Secara Acak

No	Nama Lengkap	Centroid	J	HBS	HBO	HBT	HBK wu	HBK om	MBS	MB O	MBT	MBS I	B	PAG &BP	PKN	Bina	BIng	MTK	PJO K	SB	SD& KD	PK	PJ
88	AESYA KOMALASARI	Centroid 1	4	3	6	5	1	1	3	6	5	2	19	87	82	88	80	86	81	91	80	89	85
108	Kurnia Abdul Muhamad	Centroid 2	4	1	6	1	2	1	1	6	1	5	25	64	71	77	79	77	65	82	78	66	70
185	M.Ananta meutuah	Centroid 3	5	2	6	1	1	1	2	6	1	5	12	76	76	73	74	78	87	84	83	39	65
103	Farhan yahya parilli	Centroid 4	4	4	1	3	4	2	4	1	3	1	7	68	71	72	58	76	67	76	78	85	82
29	Binta al atas	Centroid 5	2	3	6	2	1	1	3	6	2	4	4	66	74	79	80	76	88	89	87	93	87
6	Arsita Aftianur	Centroid 6	1	2	2	1	1	1	1	6	2	5	15	81	78	66	79	74	71	80	83	87	85
120	Nikhen agusriani	Centroid 7	4	6	9	3	1	2	6	9	3	4	25	85	81	85	84	80	81	83	77	65	75
216	Riki rediansyah	Centroid 8	5	1	2	1	4	1	3	2	1	1	8	65	63	85	79	69	82	80	86	75	77
226	Syahrani	Centroid 9	5	1	3	2	4	2	1	3	2	5	9	79	39	80	71	71	77	76	86	79	82
201	Najibabul cholik	Centroid 10	5	6	2	1	2	3	6	2	1	5	8	60	59	59	59	58	78	81	86	78	80

Selanjutnya hitung jarak dengan centroid (Iterasi 1) setiap data akan ditentukan centroid terdekatnya, dan data tersebut akan diterapkan sebagai anggota kelompok yang terdekat dengan centroid. Untuk menghitung jarak ke centroid masing – masing dimulai pada siswa/siswi No. 1.

Untuk mengukur jarak dengan pusat cluster digunakan Rumus Euclidian Distance sebagai berikut :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{10}$$

Keterangan : x merupakan data siswa, y merupakan centroid

$$DM1 = \sqrt{(1-4)^2 + (6-3)^2 + (2-6)^2 + (4-5)^2 + (5-1)^2 + (1-1)^2 + (1-3)^2 + (2-6)^2 + (4-5)^2 + (5-2)^2 + (23-19)^2 + (90-87)^2 + (85-82)^2 + (86-88)^2 + (82-80)^2 + (83-86)^2 + (86-81)^2 + (90-91)^2 + (78-80)^2 + (78-89)^2 + (82-85)^2} = 17,088 \tag{11}$$

$$DM2 = \sqrt{(1-4)^2 + (6-1)^2 + (2-6)^2 + (4-1)^2 + (5-2)^2 + (1-1)^2 + (1-1)^2 + (2-6)^2 + (4-1)^2 + (5-5)^2 + (23-25)^2 + (90-64)^2 + (85-71)^2 + (86-77)^2 + (82-79)^2 + (83-77)^2 + (86-65)^2 + (90-82)^2 + (78-78)^2 + (78-66)^2 + (82-70)^2} = 43,4511 \tag{12}$$

Dilanjutkan DM3 – DM10

Perhitungan jarak ke centroid masing-masing dilanjutkan pada data siswa No. 2 sampai No. 233. Dengan perhitungan jarak seluruh data terhadap 10 centroid yang sudah ditetapkan, maka dapat dilihat jarak semua nama yang menjadi sampel ke centroid terdapat pada tabel berikut :

Tabel 5. Hasil Perhitungan Jarak Antara Masing - Masing Data Dengan Centroid (Iterasi 1)

No	Nama Lengkap	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	Jarak Centroid Terdekat	Jarak Centroid Terdekat ^2	Jarak Terdekat
1	Agus Purnama	17,088	43,4511	51,2543	49,0612	39,5854	34,4238	22,4499	42,8252	55,9375	63	17,088007	292	C1
2	Amanda putri	26,6833	35,3553	44,5084	32,3265	26,8887	22,8692	27,6043	26,9072	40,4351	44,8442	22,869193	523	C6
3	Andi Irene Febrianty	27,2947	32,2335	37,2827	47,6865	42,8019	37,4166	13,1529	37,855	53,1601	58,532	13,152946	173	C7
4	Angga Wahyudi	16,6132	48,1248	56,4535	43,9431	28,1957	31,8277	34,322	38,3406	52,2972	58,8473	16,613248	276	C1
5	Apika Lestari	20,2731	35,4542	44,9667	38,4968	30,3974	26,8328	22,2486	29,7825	43,6807	51,517	20,273135	411	C1
"...."	"...."	"...."	"...."	"...."	"...."	"...."	"...."	"...."	"...."	"...."	"...."	"...."	"...."	"...."
233	Yohanes Roma harapa	48,0312	42,2729	51,6527	41,5692	40,3485	41,2795	45,7712	24,7184	14,8324	39,5474	14,832397	220	C9
WCV (Within Cluster Variation) Iterasi 1 :												95251		

Selanjutnya pada langkah ini dihitung pula besaran BCV (Between Cluster Variation). Centroid iterasi 1 digunakan untuk menghitung BCV pada iterasi 1
 Cara menghitung BCV (Between Cluster Variation) :

$d(m1, m2) =$

$$\sqrt{\begin{aligned} &(4 - 4)^2 + (3 - 1)^2 + (6 - 6)^2 \\ &+ (5 - 1)^2 + (1 - 2)^2 + (1 - 1)^2 \\ &+ (3 - 1)^2 + (6 - 6)^2 + (5 - 1)^2 \\ &+ (2 - 5)^2 + (19 - 25)^2 + (87 - 64)^2 \\ &+ (82 - 71)^2 + (88 - 77)^2 + (80 - 79)^2 \\ &+ (86 - 77)^2 + (81 - 65)^2 + (91 - 82)^2 \\ &+ (80 - 78)^2 + (89 - 66)^2 \\ &+ (85 - 70)^2 = 45,0999 \end{aligned}} \quad (13)$$

$d(m1, m3) =$

$$\sqrt{\begin{aligned} &(4 - 5)^2 + (3 - 2)^2 + (6 - 6)^2 \\ &+ (5 - 1)^2 + (1 - 1)^2 + (1 - 1)^2 \\ &+ (3 - 2)^2 + (6 - 6)^2 + (5 - 1)^2 \\ &+ (2 - 5)^2 + (19 - 12)^2 + (87 - 76)^2 \\ &+ (82 - 76)^2 + (88 - 73)^2 + (80 - 74)^2 \\ &+ (86 - 78)^2 + (81 - 87)^2 + (91 - 84)^2 \\ &+ (80 - 83)^2 + (89 - 39)^2 \\ &+ (85 - 65)^2 = 59,7411 \end{aligned}} \quad (14)$$

Perhitungan *BCV* dilanjutkan hingga $d(m9, m10)$. Selanjutnya hasil perhitungan dijumlahkan, hasilnya dapat dilihat pada tabel berikut :

Tabel 6. *BCV (Between Cluster Variation)* Iterasi 1

<i>BCV (Between Cluster Variation) Iterasi 1</i>							
M1,M2	45,0999	M2,M5	46,4004	M3,M9	59,8999	M5,M9	45,6508
M1,M3	59,7411	M2,M6	36,428	M3,M10	57,4108	M5,M10	44,1362
M1,M4	44,5758	M2,M7	31,6544	M4,M5	37,5766	M6,M7	37,2961
M1,M5	32,5423	M2,M8	32,0312	M4,M6	30,1993	M6,M8	36,2905
M1,M6	31,5436	M2,M9	46,8722	M4,M7	49,224	M6,M9	44,7437
M1,M7	30	M2,M10	45,8148	M4,M8	34,4238	M6,M10	42,4264
M1,M8	41,9762	M3,M4	57,9828	M4,M9	40,6448	M7,M8	38,9102
M1,M9	53,2259	M3,M5	60,8276	M4,M10	31,7175	M7,M9	53,4509
M1,M10	61,9112	M3,M6	56,0179	M5,M6	31,4006	M7,M10	59,8415
M2,M3	40,6325	M3,M7	37,855	M5,M7	45,5522	M8,M9	31,1929
M2,M4	38,0657	M3,M8	45,8803	M5,M8	28,8271	M8,M10	36,4829
						M9,M10	40,0749
BCV						1934,4525	

Sehingga dapat dilakukan perhitungan rasio antara besaran *BCV (Between Cluster Variation)* dengan *WCV (Within Cluster Variation)*. Perhitungan *WCV* telah dilakukan pada tabel 5. sebelumnya, pembagian besar rasio pada iterasi 1 = $BCV/WCV = 1934,4525/95251 = 0,0203$. Karena langkah ini merupakan iterasi 1 maka dilanjutkan pada tahap selanjutnya iterasi 2 dengan mengulang langkah yang sama (Iterasi).

Pembaharuan *centroid* untuk iterasi ke-2 dengan menghitung rata-rata nilai pada masing – masing *cluster*. Dapat dilihat pada tabel berikut ini :

Tabel 7. Pemisahan dan Perhitungan Rata – Rata Masing – Masing *Cluster*

Nama Lengkap	J	HB S	HB O	HB T	HB Ku	HB Ko m	MB S	MB O	MB T	MB SI	B	PA g& BP	PK N	Bin a	BIn g	MT K	PJO K	SD SB	&K D	PK	PJ	Jarak Terdekat
Cluster 2																						
Kurmia Abdul Mihar	4	1	6	1	2	1	1	6	1	5	25	64	71	77	79	77	65	82	78	66	70	C2
Muhammad Vieri	4	6	1	1	2	2	6	1	1	1	25	61	70	81	79	76	74	77	77	84	82	C2
Rahmat risky maulana	4	2	6	1	4	2	2	6	1	4	15	75	75	79	77	79	69	77	78	77	74	C2
Tedy Syahputra	4	2	2	1	4	1	2	2	1	4	8	61	72	77	76	76	62	80	78	74	75	C2
wahyu sahputra	4	1	2	2	2	1	1	2	1	5	20	68	64	77	76	79	63	75	81	82	82	C2
mario ardianto	5	2	3	2	2	1	2	3	2	1	20	67	61	76	72	76	66	75	77	78	76	C2
Mustaqim	5	1	1	5	3	1	1	1	5	2	25	59	75	77	77	70	78	76	81	80	78	C2
sukma juwita	5	1	7	1	2	1	1	7	1	4	14	65	69	79	78	76	67	78	85	78	81	C2
Rata - Rata	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	
Cluster 3																						
WHEA NOVIRA VA	1	4	7	2	1	1	4	4	5	1	5	85	81	85	78	79	84	89	79	45	66	C3
Said M Nouval Al Ha	4	6	6	1	3	1	4	7	1	4	15	86	85	88	83	84	86	88	78	39	71	C3
M.Ananta neutuah	5	2	6	1	1	1	2	6	1	5	12	76	76	73	74	78	87	84	83	39	65	C3
Prayoga	5	1	1	1	3	1	1	1	1	3	15	79	72	80	80	89	81	85	80	39	79	C3
Rata - Rata	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	
Cluster 4																						
FEBRIA LUFITA SA	2	1	3	1	1	2	1	3	1	4	9	57	72	73	59	69	71	81	89	90	84	C4
Farhan yahya parilli	4	4	1	3	4	2	4	1	3	1	7	68	71	72	58	76	67	76	78	85	82	C4
Fhirda nurul hidayah	4	3	9	2	2	1	3	9	2	5	8	72	65	78	74	78	66	82	81	90	85	C4
M. Padly	4	1	1	2	2	1	4	1	2	5	7	59	75	77	73	79	79	76	79	82	84	C4
Rahmat ramadhan	4	1	4	1	4	3	4	4	1	4	10	74	72	71	73	79	59	77	80	85	84	C4
Rata - Rata	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	

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

Tabel 11. Centroid Final

Centroid 1	2,5	2,6	6,3	2,2	1,8	1,5	2,5	5,9	2,2	3,3	21,0	83,2	80,9	83,2	81,1	81,6	82,8	88,3	83,3	87,8	84,7
Centroid 2	4,6	3,0	2,8	1,6	3,1	1,2	2,6	3,4	1,5	3,7	18,4	64,8	70,8	78,2	76,0	74,7	69,3	77,6	79,3	77,7	77,7
Centroid 3	3,6	2,8	5,4	1,4	2,0	1,0	3,2	5,0	2,4	3,4	12,0	81,0	78,2	81,8	78,6	81,2	83,8	85,4	79,8	42,8	72,4
Centroid 4	3,5	2,1	3,5	1,6	2,7	1,6	2,7	3,5	1,6	4,1	6,3	65,7	71,9	76,0	72,3	76,0	73,0	78,2	80,5	85,0	83,1
Centroid 5	3,0	2,6	5,3	2,0	1,5	1,4	2,5	5,4	2,1	3,3	4,8	82,6	80,8	82,4	81,0	81,0	82,6	87,9	82,5	85,2	82,7
Centroid 6	4,2	2,7	6,3	1,6	2,0	1,4	2,3	6,8	1,7	3,5	17,5	75,1	76,3	79,2	77,0	76,8	80,1	79,3	83,3	81,6	81,3
Centroid 7	3,2	2,5	5,9	2,1	2,0	1,7	2,4	5,8	2,4	3,4	20,1	83,7	81,5	84,1	80,6	80,6	83,1	88,2	80,5	75,9	77,7
Centroid 8	4,5	3,2	4,0	1,5	2,2	1,6	2,7	4,0	1,7	3,7	7,8	76,7	76,5	80,3	77,4	78,0	80,7	80,7	80,6	76,6	77,2
Centroid 9	5,0	3,2	3,3	2,0	3,0	1,3	3,0	3,0	1,8	3,3	11,2	70,7	51,4	80,1	74,0	70,6	78,9	78,3	80,3	76,7	78,3
Centroid 10	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

Tabel 12. Hasil Akhir

No	Nama Lengkap	C										Jarak Centroid Terdekat	Bidang Studi										
		1	2	3	4	5	6	7	8	9	10												
1	Agus Purnama											✓									13,36665	Sains dan Teknologi	
2	Amanda Putri																					9,661496	Seni dan Budaya
3	Andi Irene Febrianty																					12,08308	Sains dan Teknologi
4	Angga Wahyudi											✓										11,84242	Ekonomi dan Bisnis
5	Apika Lestari																					12,00003	Sains dan Teknologi
"...."	"...."	"...."	"...."
232	Yogi Syaputra																					13,71999	Pendidikan
233	Yohanes Roma Harapan																					13,68743	Teknik
	Jumlah	32	12	5	10	45	44	36	39	9	1												

3.2 Hasil Perhitungan Manual Dan Perhitungan Menggunakan Tools SPSS

Pada perhitungan manual dan perhitungan menggunakan tools SPSS sama – sama mengolah data sebanyak 233 data siswa SMKN Negeri 1 Kota Jambi dan terdapat 21 atribut. Pada perhitungan manual dan perhitungan tools SPSS penulis sama – sama merekomendasikan 10 cluster. Pada perhitungan manual pengambilan pusat cluster diambil secara random oleh penulis, pada perhitungan tools SPSS diambil secara random otomatis oleh sistem, keduanya mengambil data yang sama sebagai pusat cluster. Jumlah Centroid pada perhitungan manual dan SPSS adalah 10. Pada perhitungan manual terdapat jumlah iterasi sebanyak 15 kali iterasi dan perhitungan SPSS terdapat jumlah iterasi sebanyak 15 kali dilihat dari Iteration History hasil perhitungan SPSS. Pada Cluster C1 sebanyak 32 siswa direkomendasikan masuk dalam Bidang Keagamaan, C2 sebanyak 12 siswa direkomendasikan masuk dalam Bidang Hukum, C3 sebanyak 5 siswa direkomendasikan masuk dalam Bidang Sosial dan Politik, C4 sebanyak 10 siswa direkomendasikan masuk dalam Bidang Komunikasi, C5 sebanyak 45 siswa direkomendasikan masuk dalam Ekonomi dan Bisnis, C6 sebanyak 44 siswa direkomendasikan masuk dalam Bidang Pendidikan, C7 sebanyak 36 siswa direkomendasikan masuk dalam Bidang Sains dan Teknologi, C8 sebanyak 39 siswa direkomendasikan masuk dalam Bidang Seni dan Budaya, C9 sebanyak 9 siswa direkomendasikan masuk dalam Bidang Teknik, C10 sebanyak 1 siswa direkomendasikan masuk dalam Bidang Kedokteran dan Kesehatan.

4. KESIMPULAN

Berdasarkan pembahasan dan perhitungan pada penelitian mengenai Penerapan Data Mining menggunakan Algoritma K-Means ini, maka dapat di ambil kesimpulan pada penelitian ini menggunakan data siswa-siswi yang diambil di SMK Negeri 1 Kota Jambi dengan jumlah data yang digunakan 233 dan mempunyai 21 atribut 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. Metode yang digunakan dalam penelitian ini adalah K-Means Clustering, dari perhitungan manual dan perhitungan menggunakan tools SPSS (Statistical Product for Service Solutions) yang telah dilakukan, maka penulis merekomendasikan 10 cluster. Pada perhitungan manual dan perhitungan menggunakan tools SPSS terdapat jumlah iterasi sebanyak 15 kali iterasi. Jumlah Centroid pada perhitungan manual dan tools SPSS adalah 10. Jumlah cluster pada perhitungan manual dan perhitungan menggunakan tools SPSS adalah 10 Cluster. Pada Cluster C1 sebanyak 32 siswa direkomendasikan masuk Bidang Keagamaan, C2 sebanyak 12 siswa direkomendasikan masuk Bidang Hukum, C3 sebanyak 5 siswa direkomendasikan masuk Bidang Sosial dan

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.