Jurnal Sistem Informasi dan Teknologi Peradaban (JSITP)
Vol. 4, No. 1, 2023
e-ISSN 2797-1570

Pengelompokan Data Calon Siswa Baru Di Sekolah Menengah Kejuruan menggunakan Algoritma K-Means

Ningsi Multi Purnamasari[1], Achmad Syauqi[2], Danar Ardian Pramana[3]

[1]Universitas Peradaban
[2]Universitas Peradaban
[3]Universitas Peradaban
Email: [1]ningsimultipurnama@gmail.com, [2]okysyauqi@peradaban.ac.id, [3]danarmath@gmail.com

Abstrak

Pemilihan jurusan di sekolah menengah kejuruan (SMK) merupakan salah satu hal yang menentukan keberhasilan belajar peserta didik. SMK Muhammadiyah Paguyangan merupakan salah satu sekolah yang menjadi pilihan untuk para calon siswa baru dalam melanjutkan pendidikannya, dengan tujuan untuk dapat meningkatkan sumber daya manusia yang sedang dibutuhkan. Seringkali dijumpai siswa yang salah masuk jurusan dikarenakan memilih jurusan hanya berdasarkan informasi dari teman. Dalam prosesnya selama ini belum diterapkan metode sebuah metode klasterisasi yang akurat sehingga membuat tingkat akurasi yang dihasilkan dari klasterisasi jurusan belum diketahui. Salah satu langkah untuk mencapai hasil maksimal dalam pengelompokan jurusan adalah dengan pengolahan data. Salah satunya dengan pengklasteran menggunakan algoritma *K-Means*. Algoritma *K-Means* merupakan algoritma klasterisasi pengelompokan data berdasarkan titik pusat *cluster (centroid)* terdekat dengan data. Pada penelitian ini peneliti menerapkan algoritma *K-Means Clustering* untuk mengelompokan data siswa baru di SMK Muhammadiyah Paguyangan. Penelitian ini bertujuan untuk mengelompokan data calon siswa baru berdasarkan jurusan dengan menggunakan variabel nilai dan minat siswa, dengan menerapkan metode *K-Means*. Hasil dari penelitian ini diharapkan dapat membantu dalam pengolahan data untuk penentuan jurusan.

Kata Kunci : *Pemilihan Jurusan, Klasterisasi, K-Means*

I. PENDAHULUAN

Perkembangan teknologi yang semakin berkembang belum sepenuhnya di ikuti oleh banyak sekolah di Indonesia dalam menyelenggarakan Penerimaan Siswa Baru berbasis komputer. Dengan manfaat dan kemudahan yang ada, sudah seharusnya sistem ini dikembangkan oleh tiap tiap sekolah. Hal ini sejalan dengan kemajuan teknologi informasi dan komunikasi seperti teknologi yang mampu mendukung proses *input* dan *output* data secara cepat dan akurat, khususnya dalam penerimaan siswa baru. SMK Muhammadiyah Paguyangan merupakan salah satu Sekolah Menengah Kejuruan di Kabupaten Brebes. Adapun pelajaran yang diberikan disesuaikan dengan jurusan yang diambil, antara lain: Teknik Kendaraan Ringan, Teknik Ototronik, Teknik Sepeda Motor, Rekayasa Perangkat Lunak, Produksi Film dan Program Televisi. Setiap tahun pelajaran baru SMK Muhammadiyah Paguyangan mengadakan seleksi bagi calon siswa baru dengan membentuk Panitia Penerimaan yang tugasnya adalah mengelompokkan para calon siswa baru perjurusan berdasarkan kriteria masing - masing jurusan yang sudah ditentukan oleh panitia agar menghasilkan calon siswa - siswi yang sesuai atau cocok dengan jurusannya.

Metode yang akan penulis terapkan pada penerimaan siswa baru, yang sebelumnya menggunakan sistem pembobotan berdasarkan keinginan siswa baru menjadi seleksi Nilai dari SMP asal dan minat siswa. Maka pengolahan data merupakan langkah yang sangat penting untuk memperoleh hasil maksimal pengelompokan jurusan. Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, algoritma dalam data mining sangat bervariasi. Pemilihan metode yang tepat sangat bergantung pada tujuan dan proses KDD(*Knowledge Discovery in Database*) secara keseluruhan [1]. Salah satu contoh pengolahan data adalah dengan metode *Clustering*, dimana algoritma pengelompokan data yang digunakan dalam penelitian ini adalah algoritma *K-Means.*

K-Means banyak digunakan karena sifatnya yang cukup sederhana dan mudah dipahami bahkan bagi orang yang tidak memiliki latar belakang ilmu statistik, namun efektif bisa menemukan *cluster* di dalam data secara cepat. Prinsip dasar dari *K-Means* adalah melakukan proses *iteratif* (berulang) untuk menggeser *centroid*, yaitu suatu titik imajiner di dalam setiap *cluster* agar letaknya tepat ada di titik tengah *cluster.* Di titik tengah dalam konteks *clustering* artinya angka yang secara aritmetis menunjukan jarak rata-rata dari *centroid* keseluruh titik data yang ada di dalam *cluster* bernilai sama [2].

Berdasarkan latar belakang diatas peneliti tertarik untuk mengadakan penelitian tentang "Penerapan Algoritma K-*Means* Pada Pengelompokan Data Calon Siswa Baru di Sekolah Menengah Kejuruan (Studi Kasus : SMK Muhammadiyah Paguyangan)".

II. TINJAUAN PUSTAKA

A. Studi Literatur

Beberapa penelitian terkait dengan penerapan metode *clustering* dengan menggunakan algoritma *K-Means* untuk dapat mengelompokan data calon siswa baru, diantaranya adalah:

pertama, penelitian yang dilakukan oleh Firza dan Sarjono pada tahun 2020 mengenai "Penerapan Algoritma *K-Means* dalam Metode *Clustering* untuk Peminatan Jurusan Bagi Siswa Swasta Pelita Raya Kota Jambi"[3]. Penelitian ini bertujuan untuk menentukan jurusan bagi siswa di Sekolah Menengah Pertama (SMK) Swasta Pelita Raya Kota Jambi. Peminatan jurusan dilakukan pada saat siswa kelas X (Sepuluh) yang akan naik ke kelas XI (Sebelas). Pada penelitian ini hanya sebatas akan menerapkan Algoritma *K-Means Clustering* dengan menggunakan *tools RapidMiner Studio.* Hasil penilitian ini adalah rancangan *prototype* Sistem Informasi Pengelompokan peminatan jurusan pada Sekolah Menengah Kejurusan(SMK) Swasta Pelita Raya Kota Jambi yang dapat menampilkan *output* laporan pengelompokan peminat jurusan bagi kelas X(sepuluh) jurusan multimedia dan akuntansi, serta dapat menjadi salah satu solusi atau referensi pembagian tempat magang jurusan sesuai dengan minat dan bakat pada siswa dikelas sebelas multimedia dan akuntansi.

Kedua, penelitian yang dilakukan oleh Muh. Sulkifly Said dan Yusti pada tahun 2020 mengenai "Penerapan Algoritma *K-Means* dalam Penentuan Jurusan Siswa SMAN 05 Bombana"[4]. SMA 05 Bombana merupakan salah satu sekolah yang ada di Kel. Dongkala, Kec. Kabaena Timur Kab. Bombana Sulawesi Tenggara yang mengadakan penjurusan siswa ke dalam 3 jurusan yaitu IPA, IPS dan Bahasa. Data yang digunakan dalam penelitian ini merupakan data kuantitatif yaitu diambil dari data nilai siswa tahun ajaran 2018/2019. Penelitian ini bertujuan untuk mendapatkan gambaran yang jelas mengenai suatu keadaan di SMAN 05 Bombana berdasarkan data yang diperoleh dengan cara menyajikan, mengumpulkan dan menganalisis data tersebut sehingga data akan menjadi informasi baru yang dapat digunakan untuk menganalisa data yang sedang diteliti. Berdasarkan hasil implementasi terhadap algoritma *K-Means* untuk sistem pendukung keputusan untuk penentuan penjurusan siswa SMA, maka hasilnya adalah algoritma *K-Means* dapat diterapkan dalam sistem pendukung keputusan untuk penentuan penjurusan siswa di SMAN 05 Bombana.

Ketiga, penelitian yang dilakukan oleh Ika Purnama Sari dan Rika Harman pada tahun 2020 mengenai "*Decission Tree Technique* dalam Menentukan Penjurusan Siswa Menengah Kejuruan"[5]. Penelitian ini dilakukan di sekolah menengah kejuruan (SMK) Putra Jaya School yang mempunyai tiga jurusan yang paling diminati oleh calon siswa baru, yaitu keperawatan, farmasi dan teknik komputer jaringan. Penelitian ini menggunakan teknnik data mining dengan metode *Decission Tree Technique* (klasifikasi) yaitu algoritma C4.5. Parameter pemilihan jurusan pada penelitian ini adalah test buta warna, kesehatan dan wawancara. Hasil pengujian dan evaluasi menunjukan bahwa algoritma *Decission Tree* C4.5 akurat diterapkan untuk penentuan atau pemilihan jurusan dan rekomendasi siswa pada sekolah menengah kejuruan (SMK).

B. *Clustering*

Clustering merupakan salah satu teknik yang dikenal dalam data mining. Pengertian *clustering* dalam data mining adalah suatu pengelompokan sejumlah data atau objek ke dalam *cluster* sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin dan berbeda dengan objek dalam *cluster* yang lainnya. Saat ini, para ilmuwan masih terus melakukan berbagai usaha untuk melakukan perbaikan model *cluster* dan menghitung jumlah *cluster* yang optimal sehingga dapat dihasilkan *cluster* yang paling baik. Terdapat dua metode *clustering* yang dikenal, yaitu *hierarchical clustering* dan *partitioning*. Metode *hierarchical clustering* sendiri terdiri dari *complete linkage clustering*, *single linkage clustering*, *average linkage clustering* dan *centroid linkage clustering*. Sedangkan metode *partitioning* sendiri terdiri dari *K-Means* dan *Fuzzy K-Means* [10].

C. *K-Means*

Sedangkan menurut Pratama (2017) Algoritma *K-Means* adalah metode *clustering nonhierarchical* berbasis jarak yang membagi data ke dalam *cluster* dan algoritma ini bekerja pada atribut numerik. Algoritma *K-Means* termasuk dalam *partitioning clustering* yang memisahkan data ke daerah bagian yang terpisah. Algoritma *K-Means* sangat terkenal karena kemudahannya dan kemampuannya untuk meng-*cluster* data besar dan *outlier* dengan sangat cepat[15].

Secara umum algoritma *K-Means* memiliki langkah-langkah dalam pengelompokan, diantaranya[16]:

1. Pilih jumlah *cluster k*.
2. Inisialisasi *k* pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka random.

3. Alokasikan semua data atau objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap tiitk pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$d(x,y) = \| x - y \| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad ; i = 1,2,3, \dots n$$

Dimana:
d(*xy*) = jarak data ke *i* ke pusat *cluster j*
x_i = data ke *i* pada atribut data ke *k*
y_i = titik pusat ke *i* pada atribut ke *k*

4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan *median* dari cluster tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang bisa dipakai. Berikut ini adalah rumus untuk menentukan jumlah *cluster:*

$$\frac{nilai\ hasil}{banyak\ hasil}$$

5. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

III. METODE PENELITIAN

A. Identifikasi Masalah

Penjelasan terkait dengan penentuan masalah sebelumnya sudah dibahas di BAB I, yaitu dapat menghasilkan cluster atau mengelompokan data calon siswa baru untuk penentuan jurusan di SMK Muhammadiyah Paguyangan.

B. Studi literatur

Studi literatur yaitu mengumpulkan data dan informasi dari sumber bacaan. Seperti buku, jurnal, paper dan bacaan-bacaan yang ada kaitannya dengan metode *K-Means Clustering* untuk pengelompokan data calon siswa baru.

C. Pengumpulan Data

Jenis data yang digunakan adalah data kuantitatif. Data kuantitatif adalah jenis data yang dapat dihitung, berupa angka atau minimal. Sumber data yang digunakan adalah data primer dan data sekunder. Sumber data primer merupakan sumber data yang diperoleh secara langsung dari sumber asli dan tidak melalui media perantara, data nilai yang digunakan diperoleh secara langsung dari objek melalui wawancara. Sedangkan data sekunder merupakan sumber data penelitian yang diperoleh dan dicatat oleh pihak lain, data sekunder pada umumnya berupa bukti catatan atau laporan yang dipublikasikan.

D. Menganalisa data

Tahap ini dilakukan untuk menganalisa data yang sudah diperoleh dari wawancara yang dilakukan kepada panitia PPDB (Penerimaan Peserta Didik Baru) SMK Muhammadiyah Paguyangan. Tahap analisa data ini akan dilakukan beberapa langkah seperti *preprocessing* data dan *K-Means Clustering*, Tahap pertama yaitu input data mentah yang berupa data nilai siswa yang berbentuk file *excel* akan mengalami representasi dari data.

1. Pembersihan data (*cleaning*) dilakukan untuk membuang data yang tidak konsisten.
2. Setelah pembersihan data (*cleaning*) maka akan dihasilkan data yang sudah diproses atau data matang siap untuk ketahap selanjutnya.
3. Selanjutnya akan dilakukan tahap transformasi data, yaitu mengubah data yang berjenis alfabet seperti jurusan harus dilakukan proses inisialisasi data terlebih dahulu ke dalam bentuk angka/numerikal.
4. Tahap akhir yaitu *K-Means clustering* disini tahap *K-Means clustering* adalah mengelompokan data yang sudah ada kedalam beberapa kelompok.

IV. HASIL PENELITIAN DAN PEMBAHASAN

A. Persiapan Data

Data dalam peneliltian ini merupakan data nilai dari SMP asal yang diperoleh dari SMK Muhammadiyah Paguyangan dan data minat siswa yang di dapat dari hasil wawancara terhadap siswa. Dimana data nilai ini hanya diambil beberapa mata pelajaran yang dapat digunakan untuk *clutering*. Mata pelajaran yang diambil dari nilai diantaranya Matematika(MTK), Ilmu Pengetahuan Alam(IPA), dan Bahasa Inggris(B.ING). Data tersebut merupakan data yang akan digunakan untuk analisis. Untuk mengetahui informasi data, dapat dilihat pada tabel 4.1 berikut:

Tabel 4. 1 Data Nilai

NO	NAMA	MTK	IPA	B.ING
1	Abhista Nibras	82	83	86
2	Ade Setiawan	82	82	85
3	Amilatul	82	83	82
4	Bagas Bakti Pambudi	80	83	81
5	Desta Abdi Pratama	83	83	82
6	Dimas Arya Saputra	85	83	87
7	Fatimatuzzahro	85	84	81
8	Hawa Dita Saputri	83	84	88
9	Juminah	83	84	88
10	Lely Ramadani	85	82	80
…	…	…	…	…
163	Ningsih Lestari	83	84	79

Berdasarkan tabel tersebut, dapat diketahui bahwa data yang penulis miliki terdiri dari 163 siswa dan nilai mata pelajaran dari masing-masing siswa. Selain data nilai penulis juga memiliki data minta siswa. Data minat siswa dapat dilihat pada tabel 4.2 sebagai berikut:

Tabel 4. 2 Data Minat Siswa

MINAT	TKR	TSM	RPL
Abhista Nibras	3	2	1
Ade Setiawan	3	1	2
Amilatul	2	1	3
Bagas Bakti Pambudi	3	2	1
Desta Abdi Pratama	1	3	2
Dimas Arya Saputra	2	1	3
Fatimatuzzahro	1	2	3
Hawa Dita Saputri	3	2	1
Juminah	1	3	2
Lely Ramadani	3	1	2
…	…	…	…
Ningsih Lestari	1	3	2

Data minat siswa digunakan untuk mengkonversi nilai mata pelajaran sebelum melakukan *clustering*. Data minat kemudian dilakukan proses konversi data.

Tabel 4. 3 Pengkategorian Minat

Pengkategorian Minat	Kodifikasi
Pilihan Pertama	1.0
Pilihan Kedua	0.7
Pilihan Ketiga	0.3

B. *Transformasi* Data

Tahap *transformasi* dilakukan sebagai tahap awal dan tahap penting dalam penelitian. Tahapan *transformasi* data merupakan proses merubah data ke dalam bentuk yang sesuai untuk di mining. Perubahan awal yang dilakukan adalah dengan mengkonversi data nilai mata pelajaran dengan data minat, dapat di lihat dalam tabel 4.4 berikut:

Tabel 4. 4 Konversi Minat

KONVERSI MINAT	TKR	TSM	RPL
Abhista Nibras	0.3	0.7	1.0
Ade Setiawan	0.3	1.0	0.7
Amilatul	0.7	1.0	0.3
Bagas Bakti Pambudi	0.3	0.7	1.0
Desta Abdi Pratama	1.0	0.3	0.7
Dimas Arya Saputra	0.7	1.0	0.3
Fatimatuzzahro	1.0	0.7	0.3
Hawa Dita Saputri	0.3	0.7	1.0
Juminah	1.0	0.3	0.7
Lely Ramadani	0.3	1.0	0.7
…	…	…	…
Ningsih Lestari	1.0	0.3	0.7

Berdasarkan data konversi minat selanjutnya dikalikan dengan data nilai mata pelajaran yang akan digunakan untuk *clustering*. Rumus Konversi Jurusan dapat dilihat sebagai berikut.

Nilai Konversi Jurusan = Nilai * Minat Jurusan

Perhitungan Jurusan TKR:

- Konversi Jurusan TKR (Abhista Nibras)
  = Nilai MTK * Minat (TKR)
  = 82 * 0.3 = 24,6
- Konversi Jurusan TKR (Abhista Nibras)
  = Nilai IPA * Minat (TKR)
  = 83 * 0.3 = 24,9
- Konversi Jurusan TKR (Abhista Nibras)
  = Nilai B. ING * Minat (TKR)
  = 86 * 0.3 = 25.8

Tabel 4. 5 Konversi Jurusan TKR

KONVERSI TKR	MTK	IPA	B.ING
ABHISTA NIBRAS	24,6	24,9	25,8
ADE SETIAWAN	24,6	24,6	25,5
AMILATUL	57,4	58,1	57,4
BAGAS BAKTI PAMBUDI	24	24,9	24,3
DESTA ABDI PRATAMA	83	83	82
DIMAS ARYA SAPUTRA	59,5	58,1	60,9
FATIMATUZZAHRO	85	84	81
HAWA DITA SAPUTRI	24,9	25,2	26,4
JUMINAH	83	84	88
LELY RAMADANI	25,5	24,6	24
…	…	…	…
NINGSIH LESTARI	83	84	79

C. Implementasi *pada Jupyter Notebook*

Proses *clustering* akan dilakukan pada tiap-tiap Jurusan, berdasarkan hasil dari data konversi. Proses *Clustering* menggunakan *Jupyter Notebook* dan algoritma yang digunakan adalah *K-Means.*

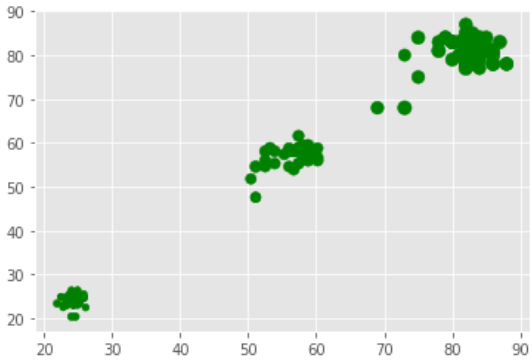
a. Memuat Data

Melakukan *loading* data pada file dataset yang sudah di transformasi untuk dimasukan ke dalam *software jupyter nootebok.*

	KONVERSI TKR	MTK	IPA	B.ING
0	ABHISTA NIBRAS	24.6	24.9	25.8
1	ADE SETIAWAN	24.6	24.6	25.5
2	AMILATUL	57.4	58.1	57.4
3	BAGAS BAKTI PAMBUDI	24.0	24.9	24.3
4	DESTA ABDI PRATAMA	83.0	83.0	82.0
5	DIMAS ARYA SAPUTRA	59.5	58.1	60.9
6	FATIMATUZZAHRO	85.0	84.0	81.0
7	HAWA DITA SAPUTRI	24.9	25.2	26.4
8	JUMINAH	83.0	84.0	88.0
9	LELY RAMADANI	25.5	24.6	24.0

Gambar 4. 1 Dataset Jurusan TKR

Terdapat 3 kolom yang akan kita gunakan, maka data dari kedua kolom di pisahkan terlebih dahulu dan ditampung ke dalam tiga variabel yang berbeda. Ketiga variabel tersebut, akan di plot kedalam sebuah *scatter* dengan semua warna data berwarna hijau yang artinya data tersebut masih merupakan seluruh data dibagian yang sama.



Gambar 4. 2 Visualisai Dataset Jurusan TKR

b. *Clustering* Data

- Fungsi *Ecluidean Distance*

  Buat Fungsi *Euclidean Distance* untuk menghitung jarak dari data ke *centroid* dan juga berguna untuk menghitung jarak antara *centroid*.
- Nilai *Cluster* Dan *Centroid*

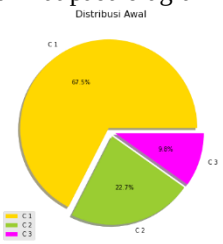
  Tentukan terlebih dahulu nilai k sebagai banyaknya *cluster* yang akan dihitung dan menentukan nilai *centroid* awal pada masing-masing varibel (MTK, IPA, dan B. ING). Pada penelitian ini penulis menentukan nilai k= 3 dan *centroid* awal = MTK (60, 80, 90), IPA (55, 75, 85), B. ING (55, 79, 85).
- Pengelompokan Data Pada Tiap *Cluster*
  - Buat variabel penampung yang akan menampung data pada tiap *cluster* yang berbeda. Variabel tersebut sejumlah dengan jumlah *cluster* yang digunakan.
  - Lakukan pengelompokan selama panjang data.
  - Tampilkan indeks dari tiap data yang ada pada suatu *cluster* dengan banyaknya data pada *cluster* tersebut.

Gambar 4. 3 *Cluster* awal

Berdasarkan hasil *cluster* awal dapat diketahui jumlah data pada tiap-tiap *cluster*. *Cluster* 1 berjumlah 110 data, *Cluster* 2 berjumlah 37 data, dan *Cluster* 3 berjumlah 16 Data

- Distribusi *Cluster*

  Distribusi *cluster* digunakan untuk menunjukan kontribusi relatif dari tiap *cluster* terhadap total keseluruhan data. Distribusi *cluster* dapat dilihat pada diagram berikut:



Gambar 4. 4 Distribusi Awal

Berdasarkan diagram diatas distribusi *cluster* terbanyak didapati pada C1 dimana terdapat distribusi *cluster* sebanyak 67,5%, sedangkan C2 dan C3 mendapati distribusi *cluster* yang lebih kecil yaitu 22.7% dan 9,8%.

- *Update Centroid*

  Update centroid digunakan untuk menetapkan *centroid* terbaik. *Update centroid* dapat dilakukan sebagi berikut:
  - Buat variabel untuk menampung *centroid* lama, yaitu *centroid* yang belum di *update*. *Centroid* lama tersebut akan digunakan untuk menghitung jarak antara *centroid* lama dengan *centroid* baru (yang sudah di-*update*).
  - Buat variabel untuk menampung jenis *cluster* dari tiap data, sehingga data dengan indeks ke i setelah proses perhitungan jarak akan memiliki nilai *cluster*-nya.
  - Buat variabel untuk menampung jarak antara *centroid* lama dengan *centroid* baru (yang sudah di-*update*).
  - Lakukan perulangan selama jarak antara *centroid* lama dengan *centroid* baru (yang sudah di*update*) sudah bernilai 0. Artinya

letak *centroid* lama dan *centroid* baru sudah tidak berpindah pindah lagi.

– Tampilkan nilai *centroid* baru yang sudah tidak berpindah pindah lagi

Hasil *centroid* yang sudah di*update* sebagai berikut:

```
Centroid I2:  [[24 24 24]
 [56 56 56]
 [81 80 80]]
```

Gambar 4. 5 *Update Centroid*

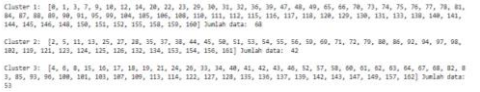
- Pengelompokan Menggunakan *Centroid* yang sudah di *update*

Setelah mendapatkan *centroid* yang sudah di-*update* kemudian lakukan pengelompokan kembali untuk mendapatkan hasil *cluster* terbaik. Pengelompokan dilakukan sebagai berikut:

– Buat variabel penampung yang akan menampung data pada tiap *cluster* yang berbeda. Variabel tersebut sejumlah dengan jumlah *cluster* yang digunakan.

– Lakukan pengelompokan selama panjang data.

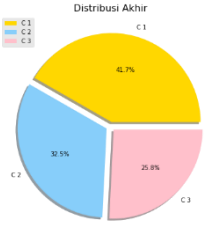
– Tampilkan indeks dari tiap data yang ada pada suatu *cluster* dengan banyaknya data pada *cluster* tersebut.



Gambar 4. 6 Hasil *Cluster* Akhir

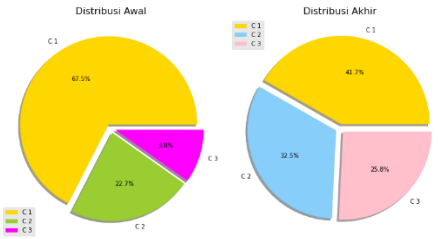
Hasil *clustering* menggunakan *centroid* yang sudah di*update* data pada tiap-tiap *cluster* sudah mengalami perubahan akhir dimana hasil *cluster* 1 berjumlah 68 data, *cluster* 2 berjumlah 42 data, dan *cluster* 3 berjumlah 53 data.

- Distribusi *Cluster* Akhir



Gambar 4. 7 Distribusi Akhir

Berdasarkan diagram diatas distribusi *cluster* terbanyak dimiliki pada C1 dimana terdapat distribusi *cluster* sebanyak 41,7%, sedangkan C2 dan C3 mendapati distribusi *cluster* yang lebih kecil yaitu 32,5% dan 25,8%. Perbandingan distribusi *cluster* dapat dilihat pada gambar berikut:



Gambar 4. 8 Distribusi *Cluter*

D. Hasil *Cluster*

Berdasarkan hasil *clusterisasi* yang telah dilakukan, terdapat perubahan rekomendasi jurusan dari data minat siswa yang telah didapat dari hasil wawancara. Hasil rekomendasi jurusan ini berdasarkan data nilai mata pelajaran siswa yang telah dikonversi menggunakan minat siswa. Hasil ini bisa menjadi acuan rekomendasi terhadap pemilihan jurusan yang dilakukan oleh siswa. Rekomendasi dari hasil clustering dapat dilihat dari tabel 4. 20 berikut :

Tabel 4. 6 Hasil Rekomendasi

NAMA	Rekomendasi		
	TKR	TSM	RPL
Abdul Azis Prasetyo	2	3	1
Abhista Nibras	3	2	1
Abid Muhilal	1	2	3
Abil Bagus Setyawan	3	2	1
Abizar Aditya	1	2	3
Adam Manda Fikia	3	2	1
Ade Setiawan	2	3	1
Adi Khoerul Rifki	3	2	1
Adrian Purniawan	1	2	3
Afiyan Muhajir Saputra	2	3	1

V. KESIMPULAN

Berdasarkan penerapan Algoritma *K-Means* dalam metode Clustering untuk pengelompokan jurusan bagi calon siswa baru SMK Muhammadiyah Paguyangan. Hasil penelitian ini antara lain: yang pertama pada jurusan Teknik Kendaraan Ringan menghasilkan distribusi *cluster* akhir untuk cluster 1 sebanyak 41,7% cluster 2 sebanyak 32,5% dan cluster 3 sebanyak 25,8%, untuk yang kedua jurusan Teknik Sepeda Motor menghasilkan distribusi *cluster* akhir untuk cluster 1 65,6% cluster 2 sebanyak 17,8% dan cluster 3 16,6%, untuk yang ketiga jurusan Rekayasa Perangkat Lunak menghasi17,2% dan cluster 3 sebanyak 16,0%. Dari hasil tersebut maka dengan menggunakan algoritma *K-Means* untuk pengelompokan data siswa baru lebih cepat, tepat dan akurat dalam penentuan jurusan atau rekomendasi jurusan siswa baru. Maka dengan adanya penerapan algoritma *K-Means* ini diharapkan mampu memberikan solusi bagi siswa dan dapat membantu pihak Sekolah

Menengah Kejuruan (SMK) Muhammadiyah Paguyangan dalam menentukan jurusan yang sesuai yang akan ditempuh oleh siswa selama bersekolah di SMK sehingga peluang untuk sukses dan meningkatnya kemampuan disekolah tersebut semakin besar.

VI. DAFTAR PUSTAKA

[1] F. Yunita, "Penerapan Data Mining Menggunkan Algoritma K-Means Clustring Pada Penerimaan Mahasiswa Baru," Sistemasi, vol. 7, no. 3, p. 238, 2018.

[2] D. Kurniawan, Pengenalan Machine Learning dengan Python. Jakarta: PT Elex Media Komputindo, 2020.

[3] F. Firza and S. Sarjono, "Penerapan Algoritma K-Means Dalam Metode Clustering Untuk Peminatan Jurusan Bagi Siswa Swasta Pelita Raya Kota Jambi," J. Manaj. Sist. Inf., vol. 5, no. 3, pp. 371–382, 2020, [Online]. Available: http://ejournal.stikom-db.ac.id/index.php/manajemensisteminformasi/article/view/907.

[4] M. S. Said and Y. Yusti, "Penerapan Algoritma K-Means Dalam Penentuan Jurusan Siswa Sman 05 Bombana," Simtek J. Sist. Inf. dan Tek. Komput., vol. 5, no. 2, pp. 114–122, 2020, doi: 10.51876/simtek.v5i2.87.

[5] I. P. Sari and R. Harman, "Decission Tree Technique Dalam Menentukan Penjurusan Siswa Menengah Kejuruan," J. Inf. Syst. Res., vol. 1, no. 4, pp. 296–304, 2020.

[6] S. Susanto and D. Suryadi, Pengantar Data Mining Menggali pengetahuan dari bongkahan data. Yogyakarta: Andi, 2010.

[7] Suyanto, Data Mining Untuk Klasifikasi dan Klasterisasi Data. Bandung, 2017.

[8] M. Bramer, Principles of data mining. London: Springer, 2007.

[9] R. T. Wulandari, Data Mining. Yogyakarta: Gava Media, 2017.

[10] T. Alfina, B. Santosa, and A. R. Barakbah, "Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabugan Keduanya dalam Membentuk Cluster Data (Studi Kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS)," Junal Tek. ITS, vol. 1, no. 1, pp. 1–5, 2012.

[11] T. Khotimah, "Pengelompokan Surat Dalam Al Qur'an Menggunakan Algoritma K-Means," Simetris J. Tek. Mesin, Elektro dan Ilmu Komput., vol. 5, no. 1, pp. 83–88, 2014, doi: 10.24176/simet.v5i1.141.

[12] R. Munir, Matematika Diskrit. Bandung: Informatika, 2012.

[13] Barakbah, A. Ridho, T. Karlita, and A. S. Ahsan, Logika dan Algoritma. Surabaya: Politeknik Elektronika Negeri Surabaya, 2013.

[14] R. A. Asroni, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," Ilm. Semesta Tek., vol. 18, no. 1, pp. 76–82, 2015.

[15] N. H. Pratama, "Implementasi Metode K-Means Pada Penerimaan Siswa," 2018.

[16] B. Santoso, Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu, 2007.

[17] J. Han, K. Micheline, and J. Pei, Data mining: concepts and techniques. Waltham: Elsevier, 2012.

[18] I. Anaconda, "Anaconda Navigator," Docs.Anaconda.Com, 2020. https://docs.anaconda.com/anaconda/navigator/#anaconda-navigator (accessed Jul. 10, 2021).

[19] Jupyter, "Jupyter," 2020. https://jupyter.org/ (accessed Jul. 10, 2021).

[20] Indrajani, Database Design (Case Study All in One). Jakarta: PT. Elex Media Komputindo, 2015.

[21] Scikit-Learn, "Sci-kit Learn," 2007. http://scikit-learn.github.io/stable (accessed Jul. 10, 2021).

[22] Sugiyono, Statisktika Untuk Penelitian. Bandung: Alfabeta, 2006.

[23] S. Margono, Metodologi penelitian pendidikan. Bandung: Rineka Cipta, 2010.