

# Algoritma *K-Means Clustering* Untuk Rekomendasi Pemberian Beasiswa Bagi Siswa Berprestasi

Febry Sandrian Sagala <sup>1</sup>, Mugiarto <sup>1,\*</sup>, Wowon Priatna <sup>1</sup>

\* Korespondensi: e-mail: [mugiarto@dsn.ubharajaya.ac.id](mailto:mugiarto@dsn.ubharajaya.ac.id)

<sup>1</sup> Informatika; Fakultas Ilmu Komputer, Universitas Bhayangkara Jakarta Raya; Jl. Raya Perjuangan, Marga Mulya, Bekasi Utara, Jawa Barat 17121. Telp: 021-88955882, 889955883, e-mail: [febry.sandrian17@mhs.ubharajaya.ac.id](mailto:febry.sandrian17@mhs.ubharajaya.ac.id), [mugiarto@dsn.ubharajaya.ac.id](mailto:mugiarto@dsn.ubharajaya.ac.id), [wowon.priatna@dsn.ubharajaya.ac.id](mailto:wowon.priatna@dsn.ubharajaya.ac.id)

Submitted : **2 Agustus 2021**  
Revised : **7 September 2021**  
Accepted : **1 Oktober 2021**  
Published : **30 November 2021**

## Abstract

*Scholarships are given to underprivileged students or outstanding students through a selection involving certain criteria. The criteria include the average value of report cards, parents' income, distance from home to school, number of dependents of parents, condition of the house, and status of the house. This study aims to assist the selection team in determining the award of scholarships so that they can provide appropriate and inappropriate recommendations, taking into account 6 criteria. The problem is that the existing scholarships are only given to students who do not have a father. The K-Means Clustering Algorithm can help Cluster students who are not eligible and eligible to get scholarship recommendations. The dataset used was 145 instances from the MAS scholarship selection committee. Attaqwa 02 Babylon. The data is calculated and tested using the K-Means Clustering algorithm. The results of the test were 32 people were recommended as eligible and 113 people were not eligible. The K-Means Clustering Algorithm can help the selection team to determine the scholarship award.*

**Keywords:** *K-Means Clustering Algorithm, Scholarship Award, Students.*

## Abstrak

Beasiswa diberikan kepada siswa yang kurang mampu atau siswa berprestasi melalui seleksi yang melibatkan kriteria-kriteria tertentu. Kriterianya antara lain nilai rata-rata raport, penghasilan orang tua, jarak rumah ke sekolah, jumlah tanggungan orang tua, kondisi rumah, dan status rumah. Penelitian bertujuan untuk membantu tim penyeleksi dalam menentukan pemberian beasiswa sehingga dapat memberikan rekomendasi layak dan tidak layak, dengan pertimbangan 6 kriteria. Masalahnya beasiswa yang ada hanya diberikan kepada siswa yang tidak memiliki Ayah. Dengan Algoritma *K-Means Clustering* dapat membantu pembuatan *Clustering* siswa yang tidak layak dan layak untuk mendapatkan rekomendasi beasiswa. Dataset yang digunakan sebanyak 145 *instance* yang berasal dari panitia seleksi beasiswa MAS. Attaqwa 02 Babelan. Data tersebut dihitung serta pengujiannya menggunakan algoritma *K-Means Clustering*. Hasil pengujian sebanyak 32 orang direkomendasikan layak dan 113 orang tidak layak. Dengan Algoritma *K-Means Clustering* dapat membantu tim seleksi untuk menentukan pemberian beasiswa.

**Kata Kunci:** Algoritma *K-Means Clustering*, Pemberian Beasiswa, Siswa.

## 1. Pendahuluan

Setiap lembaga pendidikan manapun, seperti sekolah di negeri ataupun swasta umumnya memiliki suatu program pendidikan yaitu Beasiswa. Program pemberian Beasiswa merupakan suatu wujud kepedulian suatu lembaga terhadap pelajar yang ada di Indonesia, yang bertujuan untuk membantu siswa berprestasi maupun kepada siswa yang kurang mampu selama menempuh pendidikannya.

Siswa adalah orang yang datang ke suatu lembaga untuk memperoleh atau mempelajari beberapa tipe pendidikan (Khan, 2005). Siswa adalah pribadi yang “unik” yang mempunyai potensi dan mengalami proses berkembang (Daradjat, 1995). Dalam proses berkembang itu siswa membutuhkan bantuan yang sifat dan contohnya tidak ditentukan oleh guru tetapi oleh anak itu sendiri, dalam suatu kehidupan bersama dengan individu-individu yang lain.

Beasiswa adalah wujud pemberian individu untuk melanjutkan jenjang pendidikan yang lebih tinggi, bisa berupa masuk ke institusi melalui jalur tertentu, atau bisa berupa bantuan ekonomi. Beasiswa diberikan secara selektif, dan kemudian perlu ditetapkan kriteria tertentu untuk menentukan siapa yang akan dipilih untuk menerima beasiswa (Murniasih, 2009).

Namun ada masalah yang terjadi, yaitu hanya ada pemberian beasiswa (Subsidi Silang) kepada siswa yang sudah tidak memiliki ayah (Yatim) dan tidak memberikan beasiswa kepada siswa kurang mampu dan berprestasi. Subsidi silang hanya diberikan kepada siswa saat menjadi siswa baru dan pemberian subsidi ini adalah pembebasan biaya sekolah dari awal masuk sampai lulus sekolah. Tim seleksi beasiswa sendiri adalah seorang guru, sehingga terdapat kekurangan yang diperlukan yaitu pemberian beasiswa bagi siswa berprestasi.

Penelitian penentuan rekomendasi penerima beasiswa ini dilakukan di sebuah sekolah swasta, yaitu MAS. Attaqwa 02 Babelan yang terletak di Jl. Raya Babelan, Babelan Kota, Kec. Babelan, Bekasi, Jawa Barat 17610. Dalam kegiatan belajar mengajar saat ini, Sekolah MAS. Attaqwa 02 Babelan memiliki 6 kelas dengan 22 guru pengajar dan 1 tenaga kependidikan.

*K-Means Clustering* merupakan algoritma yang efektif dalam analisis data yang dapat melakukan proses pemodelan tanpa pengawasan (*unsupervised*), dan melakukan pengelompokan data dengan menggunakan sistem partisi. Alasan digunakannya *K-Means Clustering* adalah karena tingkat akurasi dalam penggunaan untuk mendapatkan hasil sangat tinggi, setelah mengetahui sebuah permasalahan yang ada (Aditya et al., 2020).

Algoritma *K-Means* adalah algoritma yang mempartisi data ke dalam *Cluster-Cluster* sehingga data yang memiliki kemiripan berada pada satu *Cluster* yang sama dan data yang memiliki ketidaksamaan berada pada *Cluster* yang lain. Sarwono mengemukakan secara lebih detail, Algoritma *K-Means Clustering* merupakan algoritma yang efektif untuk menentukan *Cluster* dalam sekumpulan data, dimana pada algoritma tersebut dilakukan analisis kelompok yang mengacu pada pemartisian N objek ke dalam K kelompok (*Cluster*) berdasarkan nilai rata-rata (means) terdekat. Adapun tahapan algoritmanya adalah sebagai berikut (a) Menentukan berapa banyak jumlah K (*Cluster*); (b) Acak, menentukan nilai *record* yang menjadi lokasi pusat

*Cluster* awal *centroid* sebanyak  $K$ ; (c) Menghitung Jarak setiap *record* terhadap masing-masing *centroid* menggunakan rumus Euclidean Distance hingga ditemukan jarak yang paling dekat dari setiap *record* dengan *centroid*; (d) Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil); (e) Memperbaharui nilai *centroid*. Nilai *centroid* baru di peroleh dari rata-rata *Cluster* yang bersangkutan; (f) Ulangi langkah 3 sampai 5 hingga tidak ada *record* yang berpindah *Cluster* atau konvergen. Jika langkah 6 telah terpenuhi, maka nilai pusat *Cluster* ( $\mu_j$ ) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi (Nurul Rohmawati, Sofi Defiyanti, 2015).

Berikut adalah istilah-istilah didalam algoritma *K-Means Clustering* : (a) *Cluster* adalah kelompok atau grup; (b) *Cendroid* adalah titik tengah atau pusat; (c) *Iterasi* adalah pengulangan (Priyatman et al., 2019).

## **2. Metode Penelitian**

Metode penelitian adalah cara ilmiah untuk mendapatkan data yang valid dengan tujuan data akan dapat ditemukan, dikembangkan, dan dibuktikan, dengan suatu pengetahuan tertentu sehingga dapat digunakan untuk memahami, memecahkan, dan mengantisipasi masalah (Sugiyono, 2014).

### **2.1. Pengumpulan Data**

Metode yang digunakan dalam penulisan laporan penelitian ini adalah metode pengumpulan data, dengan tujuan memperoleh data dan bahan penelitian yang sesuai harapan penelitian. Tiga jenis teknik pengumpulan data yang digunakan adalah sebagai berikut: (a) Observasi menggunakan metode penelitian atau metode pengumpulan data dilakukan dengan mengamati langsung objek penelitian *dataset* siswa MAS. Attaqwa 02 Babelan sebagai sumber data infomasinya yang akan diteliti; (b) Kepustakaan untuk menentukan metode yang tepat untuk menyelesaikan masalah penelitian, serta memperoleh teori dan informasi yang berkaitan dengan topik pembahasan; (c) Wawancara dengan metode ini dilakukan dengan cara berbincang atau melakukan wawancara langsung bersama guru pengajar dan para pihak terkait dari MAS. Attaqwa 02 Babelan.

### **2.2. Pengolahan Data**

Dalam pengolahan data diperlukan sebuah algoritma, *K-Means Clustering* merupakan bidang penelitian dalam analisis dan data mining. Pada algoritma ini teknik pengelompokannya berdasarkan kemiripan data yang tidak memiliki acuan apapun (*unsupervised*). Tetapi, akan membagi keseluruhan data yang akan menjadi kelompok ataumempunyai kemiripan yang sama.

Pada dasarnya algoritma ini menghitung jarak pada setiap data dengan pusat data (*centroid*) untuk mengukur kemiripan data (Rohmah et al., 2021). Untuk menghitung antara jarak titik data objek ke titik data pusat (*centroid*), pada tahap perhitungan jarak ini menggunakan rumus dengan *Euclidean Distance* pada rumus 1.

$$d_{Euclidean}(x,y) = \sqrt{\sum_i (x_i - y_i)^2} \tag{1}$$

Keterangan:

d (x,y) : jarak antara data pada titik x dan y

x : titik data objek

y : titik data *centroid*

i : jumlah atribut data

Saat pengelompokan data objek untuk menentukan anggota *Cluster* berdasarkan jarak minimum, pada proses ini setelah menghitung jarak diambil nilai minimum yang diberi nilai 1 dan lainnya 0, dimana nilai 1 untuk data yang ditempatkan pada ke *Cluster* dan nilai 0 untuk data yang tempatkan ke *Cluster* yang lain hingga akan membentuk sebuah matriks. Kembali pada tahap ke 2 secara acak untuk menentukan nilai *record* yang menjadi lokasi pusat *Cluster* awal *centroid* sebanyak K, Lakukan pengulangan hingga poin 4 hingga nilai pada tiap *Cluster* tidak berubah tempat (Rohmah et al., 2021).

### 3. Hasil dan Pembahasan

Pada pemberian beasiswa bagi siswa yang kurang mampu dan siswa berprestasi dengan menggunakan algoritma *K-Means Clustering*, dimana hasil sebuah sistemnya sesuai dengan analisis dan perancangan yang dibuat. Dengan adanya sistem ini, maka akan mempermudah tim seleksi untuk melakukan penyeleksian pemberian beasiswa sesuai data yang ada.

Data pengujian yang digunakan berupa data pelamar beasiswa yang telah disurvei dengan komponen penyusun data dengan Memiliki 8 atribut yaitu nama, nilai rata-rata, penghasilan orang tua dalam juta (PH), jarak rumah ke sekolah (JRK), tanggungan orang tua (TG), kondisi Rumah (KR) dan Status Rumah (SR), serta jumlah *instance* adalah 145.

Selanjutnya melakukan perhitungan manual *Clustering*, menggunakan algoritma *K-Means Clustering* untuk mengelompokan data yang ada. Perhitungan *Clustering* mengikuti langkah-langkah berikut (a) Menetapkan jumlah cluster. Jumlah cluster yang ditetapkan berdasarkan rekomendasi hasil seleksi beasiswa yaitu tidak layak pada C1 dan Layak pada C2. Oleh karena itu jumlah cluster yang ditetapkan adalah 2 cluster; (b) Inisialisasi pusat cluster (*centroid*) secara acak berdasarkan data skor hasil data siswa. Nilai pusat cluster pada tabel 1.

Tabel 1. Tabel Nilai Pusat

| Cluster | Nilai | PH | JRK | TGO | KR | SR |
|---------|-------|----|-----|-----|----|----|
| 1       | 75    | 2  | 5   | 4   | 70 | 75 |
| 2       | 84    | 2  | 3   | 4   | 60 | 70 |

Sumber: Hasil Penelitian (2021)

- a. Hitung jarak setiap data terhadap pusat cluster. Misalnya untuk menghitung jarak *instance* pertama dengan pusat cluster pertama adalah:

$$D1 = \sqrt{(77 - 75)^2 + (2-2)^2 + (5-5)^2 + (4-4)^2 + (70-70)^2 + (75-75)^2} = 2$$

Menghitung jarak *instance* pertama dengan pusat *Cluster* kedua:

$$D2 = \sqrt{(77 - 84)^2 + (2-2)^2 + (5-3)^2 + (4-4)^2 + (70-60)^2 + (75-70)^2} = 13,34166406$$

Menghitung jarak *instance* kedua dengan pusat *Cluster* pertama:

$$D3 = \sqrt{(77 - 75)^2 + (2-2)^2 + (3-5)^2 + (4-4)^2 + (65-70)^2 + (85-75)^2} = 11,532563$$

Menghitung jarak *instance* kedua dengan pusat *Cluster* kedua:

$$D4 = \sqrt{(77 - 84)^2 + (2-2)^2 + (3-3)^2 + (4-4)^2 + (65-60)^2 + (85-70)^2} = 17,29161647$$

Cara hitung manual akan seperti itu dari iterasi yang pertama sampai hasil akhir nanti, dan Tabel 2 adalah perhitungan lengkap 20 *instance* pertama dari 145 *instance*.

Tabel 2. Tabel 20 *Instance* Pertama

| <i>Instance</i> | Nilai | PH | JRK | TGO | KR | SR | C1         | C2            |
|-----------------|-------|----|-----|-----|----|----|------------|---------------|
| 1               | 77    | 2  | 5   | 4   | 70 | 75 | 2          | 1,334,166,406 |
| 2               | 77    | 2  | 3   | 4   | 65 | 85 | 11,532,563 | 1,729,161,647 |
| 3               | 76    | 2  | 2   | 3   | 65 | 75 | 6          | 1,077,032,961 |
| 4               | 75    | 2  | 5   | 4   | 70 | 75 | 0          | 1,449,137,675 |
| 5               | 77    | 2  | 1   | 3   | 75 | 70 | 84,261,498 | 1,670,329,309 |
| 6               | 76    | 3  | 2   | 3   | 85 | 75 | 15,394,804 | 2,677,685,568 |
| 7               | 84    | 2  | 3   | 4   | 60 | 70 | 14,491,377 | 0             |
| 8               | 77    | 2  | 5   | 3   | 75 | 70 | 74,161,985 | 1,670,329,309 |
| 9               | 76    | 2  | 1   | 5   | 75 | 75 | 65,574,385 | 178,605,711   |
| 10              | 75    | 3  | 1   | 4   | 80 | 75 | 10,816,654 | 2,260,530,911 |
| 11              | 79    | 2  | 4   | 4   | 75 | 75 | 64,807,407 | 1,661,324,773 |
| 12              | 75    | 2  | 3   | 3   | 65 | 75 | 54,772,256 | 1,148,912,529 |
| 13              | 79    | 2  | 2   | 3   | 60 | 70 | 12,288,206 | 5,196,152,423 |
| 14              | 76    | 2  | 1   | 4   | 60 | 70 | 11,916,375 | 8,246,211,251 |
| 15              | 75    | 2  | 1   | 3   | 70 | 75 | 41,231,056 | 1,452,583,905 |
| 16              | 77    | 2  | 3   | 4   | 75 | 75 | 57,445,626 | 1,729,161,647 |
| 17              | 77    | 2  | 3   | 5   | 75 | 75 | 58,309,519 | 1,732,050,808 |
| 18              | 84    | 2  | 4   | 4   | 70 | 75 | 90,553,851 | 1,122,497,216 |
| 19              | 78    | 2  | 3   | 4   | 70 | 75 | 36,055,513 | 1,268,857,754 |
| 20              | 78    | 3  | 2   | 3   | 80 | 80 | 12,041,595 | 2,321,637,353 |

Sumber: Hasil Penelitian (2021)

- b. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* pada iterasi pertama, Tabel 3 tabel hasil pada posisi iterasi pertama.

Tabel 3. Tabel Hasil Jarak Terdekat Iterasi Pertama

| <i>Instance</i> | Nilai | PH | JRK | TGO | KR | SR | C1         | C2            | <i>Cluster</i> |
|-----------------|-------|----|-----|-----|----|----|------------|---------------|----------------|
| 1               | 77    | 2  | 5   | 4   | 70 | 75 | 2          | 1,334,166,406 | C1             |
| 2               | 77    | 2  | 3   | 4   | 65 | 85 | 11,532,563 | 1,729,161,647 | C1             |
| 3               | 76    | 2  | 2   | 3   | 65 | 75 | 6          | 1,077,032,961 | C1             |
| 4               | 75    | 2  | 5   | 4   | 70 | 75 | 0          | 1,449,137,675 | C1             |
| 5               | 77    | 2  | 1   | 3   | 75 | 70 | 84,261,498 | 1,670,329,309 | C1             |
| 6               | 76    | 3  | 2   | 3   | 85 | 75 | 15,394,804 | 2,677,685,568 | C1             |
| 7               | 84    | 2  | 3   | 4   | 60 | 70 | 14,491,377 | 0             | C2             |
| 8               | 77    | 2  | 5   | 3   | 75 | 70 | 74,161,985 | 1,670,329,309 | C1             |
| 9               | 76    | 2  | 1   | 5   | 75 | 75 | 65,574,385 | 178,605,711   | C1             |
| 10              | 75    | 3  | 1   | 4   | 80 | 75 | 10,816,654 | 2,260,530,911 | C1             |
| 11              | 79    | 2  | 4   | 4   | 75 | 75 | 64,807,407 | 1,661,324,773 | C1             |
| 12              | 75    | 2  | 3   | 3   | 65 | 75 | 54,772,256 | 1,148,912,529 | C1             |
| 13              | 79    | 2  | 2   | 3   | 60 | 70 | 12,288,206 | 5,196,152,423 | C2             |
| 14              | 76    | 2  | 1   | 4   | 60 | 70 | 11,916,375 | 8,246,211,251 | C2             |

| Instance | Nilai | PH | JRK | TGO | KR | SR | C1         | C2            | Cluster |
|----------|-------|----|-----|-----|----|----|------------|---------------|---------|
| 15       | 75    | 2  | 1   | 3   | 70 | 75 | 41,231,056 | 1,452,583,905 | C1      |
| 16       | 77    | 2  | 3   | 4   | 75 | 75 | 57,445,626 | 1,729,161,647 | C1      |
| 17       | 77    | 2  | 3   | 5   | 75 | 75 | 58,309,519 | 1,732,050,808 | C1      |
| 18       | 84    | 2  | 4   | 4   | 70 | 75 | 90,553,851 | 1,122,497,216 | C1      |
| 19       | 78    | 2  | 3   | 4   | 70 | 75 | 36,055,513 | 1,268,857,754 | C1      |
| 20       | 78    | 3  | 2   | 3   | 80 | 80 | 12,041,595 | 2,321,637,353 | C1      |

Sumber: Hasil Penelitian (2021)

- c. Memperbaharui nilai *centroid* kedua dengan menghitung nilai rata-rata dari perhitungan rata-rata dari C1 dan C2 yang sudah terpilih dari iterasi pertama.

Tabel 4. Tabel Pembaruan Nilai *Centroid* Kedua

| Cluster | Nilai   | PH        | JRK       | TGO       | KR        | SR        |
|---------|---------|-----------|-----------|-----------|-----------|-----------|
| 1       | 77      | 2,176,470 | 2,823,529 | 3,705,882 | 7,294,118 | 7,529,412 |
| 2       | 796,667 | 2         | 2         | 366,667   | 60        | 70        |

Sumber: Hasil Penelitian (2021)

- d. Hitung kembali jarak setiap data terhadap Nilai *Centroid* kedua. Menghitung jarak *instance* pertama dengan pusat *Cluster* 1 pada Nilai *Centroid* kedua sebagai berikut:

$$D1 = \sqrt{(77-77)^2 + (2-2,176470588)^2 + (5-2,823529)^2 + (4-3,705882)^2 + (70-72,94118)^2 + (75-75,29412)^2} = 3,68669$$

Menghitung jarak *instance* pertama dengan pusat *Cluster* kedua:

$$D2 = \sqrt{(77-79,6667)^2 + (2-2)^2 + (5-2)^2 + (4-3,66667)^2 + (70-60)^2 + (75-70)^2} = 11,88369555$$

Menghitung jarak *instance* kedua dengan pusat *Cluster* pertama:

$$D3 = \sqrt{(77 - 77)^2 + (2-2,176470588)^2 + (5-2,823529)^2 + (4-3,705882)^2 + (65-72,94118)^2 + (85-75,29412)^2} = 12,54652242$$

Menghitung jarak *instance* kedua dengan pusat *Cluster* kedua:

$$D4 = \sqrt{(77-79,6667)^2 + (2-2)^2 + (3-2)^2 + (4-3,66667)^2 + (65-60)^2 + (85-70)^2} = 16,06929432$$

Pada Tabel 5 menunjukkan 20 *instance* yang kedua dari nilai *centroid* kedua terbaru.

Tabel 5. Tabel 20 *Instance* dari Nilai *Centroid* Kedua

| Instance | Nilai | PH | JRK | TGO | KR | SR | C1        | C2            |
|----------|-------|----|-----|-----|----|----|-----------|---------------|
| 1        | 77    | 2  | 5   | 4   | 70 | 75 | 3,686,692 | 1,188,369,555 |
| 2        | 77    | 2  | 3   | 4   | 65 | 85 | 1,254,652 | 1,606,929,432 |
| 3        | 76    | 2  | 2   | 3   | 65 | 75 | 8,084,331 | 7,993,052,818 |
| 4        | 75    | 2  | 5   | 4   | 70 | 75 | 4,194,246 | 1,248,554,711 |
| 5        | 77    | 2  | 1   | 3   | 75 | 70 | 6,010,084 | 1,528,252,466 |
| 6        | 76    | 3  | 2   | 3   | 85 | 75 | 1,218,016 | 2,578,543,956 |
| 7        | 84    | 2  | 3   | 4   | 60 | 70 | 1,564,131 | 4,459,695,801 |
| 8        | 77    | 2  | 5   | 3   | 75 | 70 | 6,126,408 | 1,554,205,778 |
| 9        | 76    | 2  | 1   | 5   | 75 | 75 | 3,218,136 | 163,163,174   |
| 10       | 75    | 3  | 1   | 4   | 80 | 75 | 7,616     | 2,118,699,806 |
| 11       | 79    | 2  | 4   | 4   | 75 | 75 | 3,134,803 | 1,595,479,719 |
| 12       | 75    | 2  | 3   | 3   | 65 | 75 | 8,228,569 | 8,556,998,697 |
| 13       | 79    | 2  | 2   | 3   | 60 | 70 | 1,416,719 | 0,942811401   |
| 14       | 76    | 2  | 1   | 4   | 60 | 70 | 1,414,018 | 3,815,174,093 |
| 15       | 75    | 2  | 1   | 3   | 70 | 75 | 4,073,291 | 121,746,551   |
| 16       | 77    | 2  | 3   | 4   | 75 | 75 | 2,115,195 | 1,606,929,432 |
| 17       | 77    | 2  | 3   | 5   | 75 | 75 | 2,462,171 | 1,612,106,944 |
| 18       | 84    | 2  | 4   | 4   | 70 | 75 | 7,696,672 | 1,216,095,747 |
| 19       | 78    | 2  | 3   | 4   | 70 | 75 | 3,144,171 | 1,135,292,415 |
| 20       | 78    | 3  | 2   | 3   | 80 | 80 | 865,026   | 2,245,489,316 |

Sumber: Hasil Penelitian (2021)

- e. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* pada iterasi kedua, berikut tabel 6 hasil pada posisi iterasi kedua.

Tabel 6. Tabel Hasil Jarak Terdekat Iterasi Kedua

| Instance | Nilai | PH | JRK | TGO | KR | SR | C1        | C2            | Cluster |
|----------|-------|----|-----|-----|----|----|-----------|---------------|---------|
| 1        | 77    | 2  | 5   | 4   | 70 | 75 | 3,686,692 | 1,188,369,555 | C1      |
| 2        | 77    | 2  | 3   | 4   | 65 | 85 | 1,254,652 | 1,606,929,432 | C1      |
| 3        | 76    | 2  | 2   | 3   | 65 | 75 | 8,084,331 | 7,993,052,818 | C2      |
| 4        | 75    | 2  | 5   | 4   | 70 | 75 | 4,194,246 | 1,248,554,711 | C1      |
| 5        | 77    | 2  | 1   | 3   | 75 | 70 | 6,010,084 | 1,528,252,466 | C1      |
| 6        | 76    | 3  | 2   | 3   | 85 | 75 | 1,218,016 | 2,578,543,956 | C1      |
| 7        | 84    | 2  | 3   | 4   | 60 | 70 | 1,564,131 | 4,459,695,801 | C2      |
| 8        | 77    | 2  | 5   | 3   | 75 | 70 | 6,126,408 | 1,554,205,778 | C1      |
| 9        | 76    | 2  | 1   | 5   | 75 | 75 | 3,218,136 | 163,163,174   | C1      |
| 10       | 75    | 3  | 1   | 4   | 80 | 75 | 7,616     | 2,118,699,806 | C1      |
| 11       | 79    | 2  | 4   | 4   | 75 | 75 | 3,134,803 | 1,595,479,719 | C1      |
| 12       | 75    | 2  | 3   | 3   | 65 | 75 | 8,228,569 | 8,556,998,697 | C1      |
| 13       | 79    | 2  | 2   | 3   | 60 | 70 | 1,416,719 | 0,942811401   | C2      |
| 14       | 76    | 2  | 1   | 4   | 60 | 70 | 1,414,018 | 3,815,174,093 | C2      |
| 15       | 75    | 2  | 1   | 3   | 70 | 75 | 4,073,291 | 121,746,551   | C1      |
| 16       | 77    | 2  | 3   | 4   | 75 | 75 | 2,115,195 | 1,606,929,432 | C1      |
| 17       | 77    | 2  | 3   | 5   | 75 | 75 | 2,462,171 | 1,612,106,944 | C1      |
| 18       | 84    | 2  | 4   | 4   | 70 | 75 | 7,696,672 | 1,216,095,747 | C1      |
| 19       | 78    | 2  | 3   | 4   | 70 | 75 | 3,144,171 | 1,135,292,415 | C1      |
| 20       | 78    | 3  | 2   | 3   | 80 | 80 | 865,026   | 2,245,489,316 | C1      |

Sumber: Hasil Penelitian (2021)

- f. Karena belum adanya *Cluster* yang tetap karena masih ada yang berubah, maka akan berlanjut ke iterasi ketiga. Lalu nilai *centroid* berubah dengan menghitung nilai rata-rata dari perhitungan rata-rata dari C1 dan C2 yang ada pada iterasi kedua.

Tabel 7. Tabel Pembaruan Nilai *Centroid* Ketiga

| Cluster | Nilai  | PH     | JRK    | TGO    | KR     | SR      |
|---------|--------|--------|--------|--------|--------|---------|
| 1       | 77,125 | 21,875 | 26,875 | 36,875 | 73,125 | 753,125 |
| 2       | 78,75  | 2      | 2      | 3,5    | 61,25  | 71,25   |

Sumber: Hasil Penelitian (2021)

- g. Hitung kembali jarak setiap data terhadap Nilai *Centroid* ketiga. Menghitung jarak *instance* pertama dengan pusat *Cluster* 1 pada Nilai *Centroid* ketiga sebagai berikut:

$$D1 = \sqrt{(77-77,125)^2 + (2-2,1875)^2 + (5-2,6875)^2 + (4-3,6875)^2 + (70-73,125)^2 + (75-75,3125)^2} = 3,919104$$

Menghitung jarak *instance* pertama dengan pusat *Cluster* kedua:

$$D2 = \sqrt{(77-78,75)^2 + (2-2)^2 + (5-2)^2 + (4-3,5)^2 + (70-61,25)^2 + (75-71,25)^2} = 10,14581$$

Menghitung jarak *instance* kedua dengan pusat *Cluster* pertama:

$$D3 = \sqrt{(77 - 77,125)^2 + (2-2,1875)^2 + (5-2,6875)^2 + (4-3,6875)^2 + (65-73,125)^2 + (85-75,3125)^2} = 12,65343$$

Menghitung jarak *instance* kedua dengan pusat *Cluster* kedua:

$$D4 = \sqrt{(77-78,75)^2 + (2-2)^2 + (3-2)^2 + (4-3,5)^2 + (65-61,25)^2 + (85-71,25)^2} = 14,40269$$

Pada Tabel 8 menunjukkan *instance* 20 yang ketiga dari nilai *centroid* terbaru.

Tabel 8. Tabel 20 *Instance* dari Nilai *Centroid* Ketiga

| Instance | Nilai | PH | JRK | TGO | KR | SR | C1        | C2        |
|----------|-------|----|-----|-----|----|----|-----------|-----------|
| 1        | 77    | 2  | 5   | 4   | 70 | 75 | 3,919,104 | 1,014,581 |
| 2        | 77    | 2  | 3   | 4   | 65 | 85 | 1,265,343 | 1,440,269 |

| Instance | Nilai | PH | JRK | TGO | KR | SR | C1        | C2        |
|----------|-------|----|-----|-----|----|----|-----------|-----------|
| 3        | 76    | 2  | 2   | 3   | 65 | 75 | 8,267,973 | 5,994,789 |
| 4        | 75    | 2  | 5   | 4   | 70 | 75 | 4,456,386 | 1,067,415 |
| 5        | 77    | 2  | 1   | 3   | 75 | 70 | 5,925,316 | 13,962    |
| 6        | 76    | 3  | 2   | 3   | 85 | 75 | 1,199,935 | 2,422,679 |
| 7        | 84    | 2  | 3   | 4   | 60 | 70 | 1,574,752 | 5,651,327 |
| 8        | 77    | 2  | 5   | 3   | 75 | 70 | 6,132,648 | 1,424,561 |
| 9        | 76    | 2  | 1   | 5   | 75 | 75 | 3,079,671 | 146,266   |
| 10       | 75    | 3  | 1   | 4   | 80 | 75 | 7,448,783 | 1,954,322 |
| 11       | 79    | 2  | 4   | 4   | 75 | 75 | 2,997,395 | 1,440,269 |
| 12       | 75    | 2  | 3   | 3   | 65 | 75 | 8,440,046 | 6,590,713 |
| 13       | 79    | 2  | 2   | 3   | 60 | 70 | 1,431,728 | 185,405   |
| 14       | 76    | 2  | 1   | 4   | 60 | 70 | 1,430,854 | 3,455,069 |
| 15       | 75    | 2  | 1   | 3   | 70 | 75 | 421,122   | 1,029,259 |
| 16       | 77    | 2  | 3   | 4   | 75 | 75 | 1,964,529 | 1,440,269 |
| 17       | 77    | 2  | 3   | 5   | 75 | 75 | 2,341,874 | 1,447,196 |
| 18       | 84    | 2  | 4   | 4   | 70 | 75 | 7,680,129 | 1,106,515 |
| 19       | 78    | 2  | 3   | 4   | 70 | 75 | 3,295,357 | 9,614,442 |
| 20       | 78    | 3  | 2   | 3   | 80 | 80 | 8,462,232 | 2,073,493 |

Sumber: Hasil Penelitian (2021)

- h. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* pada iterasi ketiga, tabel 9 hasil pada posisi iterasi ketiga.

Tabel 9. Tabel Hasil Jarak Terdekat Iterasi Ketiga

| Instance | Nilai | PH | JRK | TGO | KR | SR | C1        | C2        | Cluster |
|----------|-------|----|-----|-----|----|----|-----------|-----------|---------|
| 1        | 77    | 2  | 5   | 4   | 70 | 75 | 3,919,104 | 1,014,581 | C1      |
| 2        | 77    | 2  | 3   | 4   | 65 | 85 | 1,265,343 | 1,440,269 | C1      |
| 3        | 76    | 2  | 2   | 3   | 65 | 75 | 8,267,973 | 5,994,789 | C2      |
| 4        | 75    | 2  | 5   | 4   | 70 | 75 | 4,456,386 | 1,067,415 | C1      |
| 5        | 77    | 2  | 1   | 3   | 75 | 70 | 5,925,316 | 13,962    | C1      |
| 6        | 76    | 3  | 2   | 3   | 85 | 75 | 1,199,935 | 2,422,679 | C1      |
| 7        | 84    | 2  | 3   | 4   | 60 | 70 | 1,574,752 | 5,651,327 | C2      |
| 8        | 77    | 2  | 5   | 3   | 75 | 70 | 6,132,648 | 1,424,561 | C1      |
| 9        | 76    | 2  | 1   | 5   | 75 | 75 | 3,079,671 | 146,266   | C1      |
| 10       | 75    | 3  | 1   | 4   | 80 | 75 | 7,448,783 | 1,954,322 | C1      |
| 11       | 79    | 2  | 4   | 4   | 75 | 75 | 2,997,395 | 1,440,269 | C1      |
| 12       | 75    | 2  | 3   | 3   | 65 | 75 | 8,440,046 | 6,590,713 | C2      |
| 13       | 79    | 2  | 2   | 3   | 60 | 70 | 1,431,728 | 185,405   | C2      |
| 14       | 76    | 2  | 1   | 4   | 60 | 70 | 1,430,854 | 3,455,069 | C2      |
| 15       | 75    | 2  | 1   | 3   | 70 | 75 | 421,122   | 1,029,259 | C1      |
| 16       | 77    | 2  | 3   | 4   | 75 | 75 | 1,964,529 | 1,440,269 | C1      |
| 17       | 77    | 2  | 3   | 5   | 75 | 75 | 2,341,874 | 1,447,196 | C1      |
| 18       | 84    | 2  | 4   | 4   | 70 | 75 | 7,680,129 | 1,106,515 | C1      |
| 19       | 78    | 2  | 3   | 4   | 70 | 75 | 3,295,357 | 9,614,442 | C1      |
| 20       | 78    | 3  | 2   | 3   | 80 | 80 | 8,462,232 | 2,073,493 | C1      |

Sumber: Hasil Penelitian (2021)

- i. Karena belum adanya *Cluster* yang tetap karena masih ada yang berubah, maka akan berlanjut ke iterasi keempat. Lalu nilai *centroid* berubah dengan menghitung nilai rata-rata dari perhitungan rata-rata dari C1 dan C2 yang ada pada iterasi ketiga.

Tabel 10. Tabel Pembaruan Nilai *Centroid* Keempat

| Cluster | Nilai | PH  | JRK     | TGO | KR | SR            |
|---------|-------|-----|---------|-----|----|---------------|
| 1       | 77,2  | 2,2 | 286,667 | 3,8 | 74 | 7,533,333,333 |
| 2       | 78    | 2   | 2,2     | 3,4 | 62 | 72            |

Sumber: Hasil Penelitian (2021)

- j. Hitung kembali jarak setiap data terhadap Nilai *Centroid* keempat. Menghitung jarak *instance* pertama dengan pusat *Cluster* 1 pada Nilai *Centroid* keempat sebagai berikut:

$$D1 = \sqrt{(77-77,2)^2 + (2-2,2)^2 + (5-2,86667)^2 + (4-3,8)^2 + (70-74)^2 + (75-75,33333333)^2} = 4,558750706$$

Menghitung jarak *instance* pertama dengan pusat *Cluster* kedua:

$$D2 = \sqrt{(77-78)^2 + (2-2)^2 + (5-2,2)^2 + (4-3,4)^2 + (70-62)^2 + (75-72)^2} = 9,066421565$$

Menghitung jarak *instance* kedua dengan pusat *Cluster* pertama:

$$D3 = \sqrt{(77 - 77,2)^2 + (2-2,2)^2 + (5-2,86667)^2 + (4-3,8)^2 + (65-74)^2 + (85-75,33333333)^2} = 13,21295657$$

Menghitung jarak *instance* kedua dengan pusat *Cluster* kedua:

$$D4 = \sqrt{(77-78)^2 + (2-2)^2 + (3-2,2)^2 + (4-3,4)^2 + (65-62)^2 + (85-72)^2} = 13,41640786$$

Tabel 11 menunjukkan *instance* 20 yang keempat dari nilai *centroid* terbaru.

Tabel 11. Tabel 20 *Instance* dari Nilai *Centroid* Keempat

| <i>Instance</i> | Nilai | PH | JRK | TGO | KR | SR | C1            | C2            |
|-----------------|-------|----|-----|-----|----|----|---------------|---------------|
| 1               | 77    | 2  | 5   | 4   | 70 | 75 | 4,558,750,706 | 9,066,421,565 |
| 2               | 77    | 2  | 3   | 4   | 65 | 85 | 1,321,295,657 | 1,341,640,786 |
| 3               | 76    | 2  | 2   | 3   | 65 | 75 | 9,164,181,796 | 4,711,687,596 |
| 4               | 75    | 2  | 5   | 4   | 70 | 75 | 5,057,885,724 | 9,497,368,056 |
| 5               | 77    | 2  | 1   | 3   | 75 | 70 | 5,800,767,303 | 1,325,141,502 |
| 6               | 76    | 3  | 2   | 3   | 85 | 75 | 1,116,164,092 | 2,330,665,141 |
| 7               | 84    | 2  | 3   | 4   | 60 | 70 | 1,645,546,175 | 6,708,203,932 |
| 8               | 77    | 2  | 5   | 3   | 75 | 70 | 5,891,989,587 | 1,349,073,756 |
| 9               | 76    | 2  | 1   | 5   | 75 | 75 | 2,741,453,629 | 136,381,817   |
| 10              | 75    | 3  | 1   | 4   | 80 | 75 | 6,716,812,339 | 1,856,879,102 |
| 11              | 79    | 2  | 4   | 4   | 75 | 75 | 239,072,123   | 1,351,295,675 |
| 12              | 75    | 2  | 3   | 3   | 65 | 75 | 9,308,538,446 | 5,272,570,531 |
| 13              | 79    | 2  | 2   | 3   | 60 | 70 | 1,513,656,372 | 3,033,150,178 |
| 14              | 76    | 2  | 1   | 4   | 60 | 70 | 1,514,757,081 | 3,714,835,124 |
| 15              | 75    | 2  | 1   | 3   | 70 | 75 | 5,011,543,475 | 9,143,303,561 |
| 16              | 77    | 2  | 3   | 4   | 75 | 75 | 1,117,536,576 | 1,341,640,786 |
| 17              | 77    | 2  | 3   | 5   | 75 | 75 | 1,627,540,475 | 1,349,814,802 |
| 18              | 84    | 2  | 4   | 4   | 70 | 75 | 7,982,201,952 | 1,061,131,472 |
| 19              | 78    | 2  | 3   | 4   | 70 | 75 | 4,104,739,699 | 8,602,325,267 |
| 20              | 78    | 3  | 2   | 3   | 80 | 80 | 7,774,888,726 | 1,972,815,247 |

Sumber: Hasil Penelitian (2021)

- k. Mengklasifikasikan kembali setiap data berdasarkan kedekatannya dengan *centroid* pada iterasi keempat, berikut tabel 12 hasil pada posisi iterasi keempat.

Tabel 12. Tabel Hasil Akhir Jarak Terdekat Iterasi Keempat

| <i>Instance</i> | Nilai | PH | JRK | TGO | KR | SR | C1            | C2            | <i>Cluster</i> |
|-----------------|-------|----|-----|-----|----|----|---------------|---------------|----------------|
| 1               | 77    | 2  | 5   | 4   | 70 | 75 | 4,558,750,706 | 9,066,421,565 | C1             |
| 2               | 77    | 2  | 3   | 4   | 65 | 85 | 1,321,295,657 | 1,341,640,786 | C1             |
| 3               | 76    | 2  | 2   | 3   | 65 | 75 | 9,164,181,796 | 4,711,687,596 | C2             |
| 4               | 75    | 2  | 5   | 4   | 70 | 75 | 5,057,885,724 | 9,497,368,056 | C1             |
| 5               | 77    | 2  | 1   | 3   | 75 | 70 | 5,800,767,303 | 1,325,141,502 | C1             |
| 6               | 76    | 3  | 2   | 3   | 85 | 75 | 1,116,164,092 | 2,330,665,141 | C1             |
| 7               | 84    | 2  | 3   | 4   | 60 | 70 | 1,645,546,175 | 6,708,203,932 | C2             |
| 8               | 77    | 2  | 5   | 3   | 75 | 70 | 5,891,989,587 | 1,349,073,756 | C1             |
| 9               | 76    | 2  | 1   | 5   | 75 | 75 | 2,741,453,629 | 136,381,817   | C1             |
| 10              | 75    | 3  | 1   | 4   | 80 | 75 | 6,716,812,339 | 1,856,879,102 | C1             |
| 11              | 79    | 2  | 4   | 4   | 75 | 75 | 239,072,123   | 1,351,295,675 | C1             |
| 12              | 75    | 2  | 3   | 3   | 65 | 75 | 9,308,538,446 | 5,272,570,531 | C2             |
| 13              | 79    | 2  | 2   | 3   | 60 | 70 | 1,513,656,372 | 3,033,150,178 | C2             |
| 14              | 76    | 2  | 1   | 4   | 60 | 70 | 1,514,757,081 | 3,714,835,124 | C2             |
| 15              | 75    | 2  | 1   | 3   | 70 | 75 | 5,011,543,475 | 9,143,303,561 | C1             |
| 16              | 77    | 2  | 3   | 4   | 75 | 75 | 1,117,536,576 | 1,341,640,786 | C1             |

|    |    |   |   |   |    |    |               |               |    |
|----|----|---|---|---|----|----|---------------|---------------|----|
| 17 | 77 | 2 | 3 | 5 | 75 | 75 | 1,627,540,475 | 1,349,814,802 | C1 |
| 18 | 84 | 2 | 4 | 4 | 70 | 75 | 7,982,201,952 | 1,061,131,472 | C1 |
| 19 | 78 | 2 | 3 | 4 | 70 | 75 | 4,104,739,699 | 8,602,325,267 | C1 |
| 20 | 78 | 3 | 2 | 3 | 80 | 80 | 7,774,888,726 | 1,972,815,247 | C1 |

Sumber: Hasil Penelitian (2021)

Hasil telah ditemukan, bahwa tidak ada perubahan lagi dari hasil iterasi ketiga sampai iterasi keempat. Pada penelitian ini, Iterasi berhenti pada iterasi ke-4, dimana ada 15 siswa yang berada di C1 dan 5 siswa pada C2, sehingga yang dapat direkomendasikan ke Kepala Sekolah untuk menerima beasiswa dari 20 orang dalam percobaan 145 data sebanyak 5 siswa pada C2.

#### 4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, algoritma *K-Means Clustering* mampu mengelompokkan calon penerima beasiswa bagi siswa kurang mampu dan berprestasi ke dalam 2 *Cluster*, yang mana *Cluster 1* memberikan rekomendasi tidak layak dan *Cluster 2* memberikan rekomendasi layak. Berdasarkan hasil pengujian 145 Siswa calon penerima beasiswa dihasilkan 32 orang direkomendasikan layak dari *Cluster 2* dan 113 orang direkomendasikan tidak layak pada *Cluster 1*. Hasil *Clustering* hanya merupakan rekomendasi bagi pengambil keputusan dan bukan penentu akhir penerima beasiswa.

#### Daftar Pustaka

- Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(1), 51–58. <https://doi.org/10.30865/mib.v4i1.1784>
- Daradjat, Z. (1995). *Remaja Harapan dan Tantangan*. Ruhama.
- Khan, P. D. A. S. (2005). *Filsafat Pendidikan Al-Ghazali*. Pustaka Setia.
- Murniasih, E. (2009). *Buku Pintar Beasiswa*. Gagas Media.
- Nurul Rohmawati, Sofi Defiyanti, M. J. (2015). Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa. *Jitter 2015*, 1(2), 62–68.
- Priyatman, H., Sajid, F., & Haldivany, D. (2019). *Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan*. 5(1), 62–66.
- Rohmah, A., Sembiring, F., & Erfina, A. (2021). Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: SMK Yaspim Gegerbitung). *SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)*. 1(01), 290-298.
- Sugiyono. (2014). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, kualitatif, dan R&D*. Alfabeta.