



KLASTERISASI SELEKSI MAHASISWA CALON PENERIMA BEASISWA YAYASAN MENGGUNAKAN *K-MEANS* *CLUSTERING*

Sabrina Aulia Rahmah¹, Jovi Antares²

^{1,2}Universitas Dharmawangsa

Jl. K.L.Yos Sudarso No. 224, Medan, 20115

e-mail : sabrinaaulia@dharmawangsa.ac.id

ABSTRAK

Perekomendasi untuk penerima beasiswa Yayasan dikelompokkan menjadi 3 cluster yaitu, diterima, dipertimbangkan dan ditolak sebagai penerima beasiswa yayasan. Algoritma K-Means Clustering merupakan salah satu teknik unsupervised learning yang digunakan untuk merekomendasi penerima beasiswa yayasan. Adapun tujuan dari penelitian ini adalah untuk merekomendasikan calon penerima beasiswa yayasan dengan menggunakan algoritma K-Means Clustering, rekomendasi menghasilkan penempatan data pendaftar beasiswa ke masing-masing kelompok cluster yang dihasilkan. Data pendaftar yang digunakan sebanyak 80 pendaftar. Melalui penyeleksian atribut k-means melakukan perhitungan untuk menempatkan setiap data ke cluster yang sudah ditentukan. Hasil dari perhitungan yang telah diolah sebanyak 16% diterima, 61% dipertimbangkan dan 23% ditolak.

Kata kunci : *Beasiswa, Clustering, K-Means Clustering, Hypertext Preprocessor (PHP)*

ABSTRACT

Recommendations for Foundation scholarship recipients are grouped into 3 clusters, namely, acceptance, consideration and rejection as foundation scholarship recipients. The K-Means Clustering algorithm is one of the unsupervised learning techniques used to recommend foundation scholarship recipients. The purpose of this study is to recommend prospective foundation scholarship recipients using the K-Means Clustering algorithm, the recommendations result in the placement of scholarship applicant data into each of the resulting cluster groups. Registrant data used as many as 80 registrants. Through selecting the k-means attribute, it performs calculations to place each data into a predetermined cluster. The results of the calculations that have been processed are 16.3% accepted, 61.3% considered and 22.5% rejected.

Keywords : *Scholarship, Clustering, K-Means Clustering, Hypertext Preprocessor (PHP)*

1. PENDAHULUAN

Banyak jenis beasiswa yang ditawarkan untuk seseorang melanjutkan pendidikan ke tingkat perguruan tinggi diantaranya adalah beasiswa Yayasan. Beasiswa Yayasan merupakan beasiswa yang diberikan oleh Yayasan bagi calon mahasiswa baru untuk masuk perguruan tinggi swasta. Salah satunya adalah Yayasan Pendidikan Dharmawangsa, yayasan memberikan beasiswa berupa gratis uang kuliah selama 4 tahun dengan kata lain selama awal mula kuliah hingga akhir kuliah. Beasiswa diberikan kepada

mahasiswa yang kurang mampu dari segi finansial ekonomi tetapi memiliki keinginan untuk melanjutkan pendidikan ke jenjang perguruan tinggi. Beasiswa ini biasanya disalurkan langsung oleh perguruan tinggi yang bersangkutan. Beasiswa sendiri diberikan kepada siswa sekolah tingkat SMA/ SMK yang baru lulus atau paling lama 2 tahun setelah lulus SMA/SMK.

Untuk seleksi pendaftaran beasiswa yayasan memiliki persyaratan-persyaratan yang harus dipenuhi calon mahasiswa penerima sebelum



menerimanya. Adapun syarat yang harus dipenuhi diantaranya adalah Nilai Ujian Akhir Sekolah (UAS), Nilai Rapor, Status Rumah, Kondisi Rumah, Penghasilan Orang tua. Dalam penyeleksian penerima beasiswa yayasan digunakan algoritma *K-Means Clustering*.

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Data mining mulai ada sejak 1990 -an sebagai cara yang benar dan tepat untuk mengambil pola dan informasi yang digunakan untuk menemukan hubungan antara data untuk melakukan pengelompokan ke dalam satu atau lebih *Cluster* sehingga objek - objek yang berada dalam satu *Cluster* akan mempunyai kesamaan yang tinggi antara satu dengan lainnya (Putri et al., 2021).

Clustering bisa diartikan sebagai identifikasi kelas objek yang memiliki kemiripan dengan data yang lain. Dengan menggunakan teknik *Clustering* dapat mengidentifikasi kepadatan dan jarak daerah dalam objek ruang dan dapat menemukan secara keseluruhan pola distribusi dan kolerasi antara atribut (Ginting & Riandari, 2020).

Algoritma K-Means merupakan algoritma pada *unsupervised learning* pada proses *Clustering* yang mengelompokkan data berdasarkan kemiripan / kesamaan. Data dengan karakter sama akan dikelompokkan pada satu *Cluster* sedangkan data dengan karakter berbeda akan dikelompokkan pada *Cluster* yang lain (Salam et al., 2020).

Algoritma ini digunakan untuk mengelompokkan calon penerima beasiswa dalam 3 kategori yaitu diterima, dipertimbangkan dan ditolak. Dengan menggunakan Algoritma *K-Means Clustering* dapat membantu klasterisasi pengusulan calon mahasiswa penerima beasiswa kedalam kategori yang layak dan tidak layak untuk menerima beasiswa sesuai dengan kriteria-kriteria yang sudah ditentukan. Penggunaan Algoritma *K-means* dilakukan karena sudah banyak penelitian terhadap pengelompokkan suatu hal yang menggunakan *Clustering*.

Pada penelitian sebelumnya Algoritma *K-means* dilakukan dalam hal menentukan beasiswa yayasan dengan menghasilkan empat golongan berdasarkan jumlah dari kriteria, hasil pengelompokan memperlihatkan jumlah dan keputusan yayasan terhadap pemberian beasiswa terhadap mahasiswanya. (Sudarsono & Lestari, 2021). Pada penelitian yang lain Algoritma *K-means* dapat membantu untuk mengklasifikasi

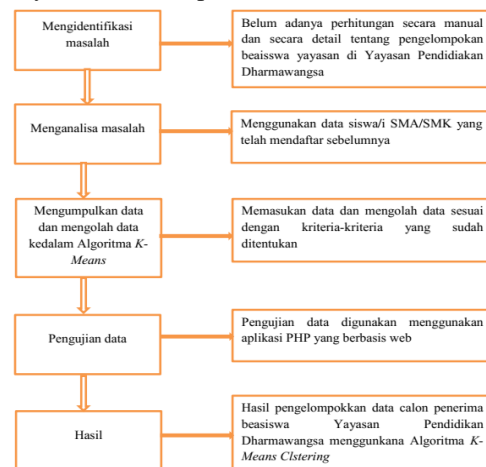
siswa-siswi yang sangat layak dalam mendapatkan bantuan berupa beasiswa, dari 128 data set yang di gunakan menghasilkan perhitungan yang diperoleh sebesar C1:73 item, C2: 30 item dann C3: 25 Item (Manihuruk et al., 2020). Algoritma *K-Means Clustering* juga digunakan dalam mengelompokkan provinsi lampung yang terkena kasus *covid-19* dengan membagi kedalam 4 *Cluster* yaitu zona merah, zona orange, zona kuning dan zona hijau (Teknologi et al., 2021).

Algoritma *K-means Clustering* adalah salah satu cara untuk mengkluster sebuah kumpulan data dan membagi data tersebut untuk membentuk sebuah *Cluster* yang baru (Rahayu et al., 2019). *K-Means Clustering* merupakan salah satu metode data *Clustering* non hirarki yang berusaha mempartisi data yang ada kedalam bentuk satu atau lebih *Cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan dalam satu *Cluster* (Rahmah, 2020).

Dalam penelitian ini pengujian data menggunakan aplikasi PHP yang berbasis web yang dapat diakses menggunakan *browser*, yang mana sampel yang digunakan adalah data-data dari calon mahasiswa penerima beasiswa yayasan yang telah mendaftar sebelumnya. PHP adalah akronim dari *Hypertext Preprocessor*, yaitu suatu bahasa pemrograman berbasiskan kode-kode (*script*) yang digunakan untuk mengolah suatu data dan mengirimkannya kembali ke web *browser* menjadi kode HTML (Hasan & Muhammad, 2020).

2. METODOLOGI PENELITIAN

Kerangka penelitian adalah sebuah alur yang dibuat untuk memudahkan peneliti dalam menyelesaikan tahapan masalah.



Gambar 1. Kerangka Penelitian



1. Mengidentifikasi masalah
Identifikasi masalah dilakukan sebelum adanya perhitungan data secara manual dan secara detail tentang pengelompokan beasiswa yaysana di Yayasan Pendidikan Dharmawangsa.
2. Menganalisa masalah
Setelah masalah diidentifikasi tahapan selanjutnya adalah menganalisa masalah dengan menggunakan data siswa/i SMA/SMK yang telah mendaftar untuk calon mahasiswa penerima beasiswa yaysana.
3. Mengumpulkan data dan Mengolah data kedalam Algoritma *K-Means Clustering*
Data yang telah dikumpulkan akan langsung diolah dengan menggunakan rumus Algoritma *K-Means* yang sudah ada, dan sesuai dengan kriteria yang sudah ditentukan.
4. Pengujian data
Data diuji menggunakan aplikasi PHP yang berbasis web untuk memudahkannya dalam membaca hasil pengelompokan.
5. Hasil
Hasil yang didapat berupa grafik visualisasi persantase hasil data yang telah diolah.

3. HASIL DAN PEMBAHASAN

Dalam pemanfaatan algoritma *K-Means Clustering* dalam mengelompokkan calon penerima beasiswa dibutuhkan data yang terkait. Sumber data penelitian didapat dari data yang dikumpulkan sebelumnya. Data yang digunakan dalam penelitian ini adalah Nilai Ujian Akhir Sekolah (UAS), Nilai Rapor, Status Rumah, Kondisi Rumah dan Penghasilan Orang Tua yang terdiri dari 30 calon mahasiswa, data yang diperoleh diolah dengan melakukan pembagian terhadap 3 *Cluster* yaitu Diterima, Dipertimbangkan dan Ditolak.

Algoritma *K-Means Clustering* merupakan algoritma akan mengelompok data kedalam *Cluster - Cluster* sehingga ketika ada data yang memiliki karkter yang sama maka akan berada pada *Cluster* yang sama dan data yang memiliki ketidak samaan akan berada di dalam *Cluster* yang lain, berikut ini merupakan tahapan perhitungan algoritma *Clustering*, yaitu: (Sudarsono & Lestari, 2021)

1. Menentukan banyaknya yang akan di bentuk ada 3 *Cluster* (k=3). Untuk penentuan *Cluster* harus lebih kecil dari pada banyak nya data (k<n).

2. Menentukan nilai secara manual atau random untuk pusat *Cluster* awal sebanyak *Cluster* yang ditentukan.

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \tag{1}$$

Keterangan:

Vij : Centroid rata-rata *Cluster* ke-1 untuk variabel ke j

Ni : Jumlah anggota *Cluster* ke-i

i,k : Imdeks dari *Cluster*

j : Indeks dari Variabel

Xkj: Nilai data ke-k variabel ke-j untuk *Cluster* tersebut

3. Untuk menghitung jarak data dengan centroid menggunakan rumus Euclidian Distance.

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

Keterangan:

De : Euclidian Distance

i : Banyaknya objek

(x,y) : Koordinat objek

(s,t) : Koordinat centroid

4. Mengecek setiap kedekatan data dengan jarak terkecil.
5. Centroid baru akan dihitung dengan melakukan perhitungan nilai rata-rata data pada setiap *Cluster*.
6. Melakukan perulangan jika perhitungan iterasi yang baru berbeda dengan iterasi sebelumnya, maka proses akan di lanjutkan ke langkah perulangan selanjutnya. Namun jika iterasi yang baru di hitung memiliki hasil yang sama dengan iterasi sebelumnya, maka proses *Clustering* selesai. Maka nilai pusat *Cluster* pada iterasi terakhir akan di gunakan sebagai parameter yang dimana untuk menentukan kasifikasi data

Berikut adalah tampilan hasil yang digunakan dengan menerapkan algoritma *K-Means Clustering* dengan menggunakan pengujian aplikasi PHP:

a) Tampilan Awal Pada Aplikasi



Gambar 2. Tampilan Awal

Tampilan awal berisi tentang penjelasan dari Data Mining Metode K-Means.



b) Tampilan Atribut Atribut

Kode	Nama Atribut
A01	Nilai UAS
A02	Nilai Rapor
A03	Status Rumah
A04	Kondisi Rumah
A05	Penghasilan Orang Tua

Gambar 3. Tampilan Atribut

Tampilan atribut menjelaskan atribut yang digunakan untuk perhitungan algoritma *K-Means Clustering* berisi kode dan nama atribut. Kode atribut terdiri dari A01, A02, A03, A04, A05 dan nama atributnya terdiri dari Nilai UAS, Nilai Rapor, Status Rumah, Kondisi Rumah, Penghasilan Orang Tua.

c) Tampilan Nilai Atribut Nilai Atribut

No	Kode	Nama Atribut	Nama Nilai Atribut	Nilai
1	A03	Status Rumah	Mongotrak	0
2	A03	Status Rumah	Milik Pribadi	1
3	A04	Kondisi Rumah	Ragus	1
4	A04	Kondisi Rumah	Sederhana	0
5	A05	Penghasilan Orang Tua	< 2.000.000	0
6	A05	Penghasilan Orang Tua	>= 2.000.000	1

Gambar 4. Tampilan Nilai Atribut

Tampilan nilai atribut berisi tentang nilai dari masing-masing dari atribut. Nilai yang diberikan untuk nilai atribut adalah 0 dan 1.

d) Tampilan Alternatif

No	Kode	Nama Alternatif	Keterangan	Nilai UAS	Nilai Rapor	Status Rumah	Kondisi Rumah	Penghasilan Orang Tua
1	002061444	BAGUS ESATRA ADIRA	S1 HUKUM	78	96	Milik Pribadi	Ragus	>= 2.000.000
2	002061448	DARA FADULLAH NIST	DIII SISTEM INFORMASI	79	92	Milik Pribadi	Sederhana	< 2.000.000
3	002061452	HEZTA NAKORA SOFYAN	S1 ADMINISTRASI PUBLIK	87	96	Milik Pribadi	Sederhana	>= 2.000.000
4	002061441	ANDINA NATALIA	S1 EKONOMI BISNIS	90	93	Milik Pribadi	Sederhana	>= 2.000.000
5	002061468	PUTRI ANGGRANI	S1 EKONOMI MANAJEMEN	88	90	Milik Pribadi	Sederhana	>= 2.000.000
6	002069449	EPRIKA SHOTANG	S1 ADMINISTRASI PUBLIK	90	90	Milik Pribadi	Sederhana	>= 2.000.000
7	002061384	SAM DEVA NAGRA	S1 TEKNOLOGI INFORMASI	93	95	Milik Pribadi	Ragus	>= 2.000.000
8	002061449	ADELIA RAJAGUNGIK	S1 EKONOMI MANAJEMEN	88,5	92	Milik Pribadi	Ragus	>= 2.000.000

Gambar 5. Tampilan Alternatif

Tampilan ini berisi data calon mahasiswa penerima beasiswa yayasan yang akan diolah dan dikelompokkan.

e) Tampilan Perhitungan

Dalam tampilan perhitungan ada beberapa tampilan yang dijalankan sekaligus, diantaranya:

- Hasil Analisa

Perhitungan

Kode	Nama	Nilai UAS	Nilai Rapor	Status Rumah	Kondisi Rumah	Penghasilan Orang Tua
002061444	BAGUS ESATRA ADIRA	78	96	1	0	1
002061448	DARA FADULLAH NIST	79	92	1	0	0
002061452	HEZTA NAKORA SOFYAN	87	96	1	0	1
002061441	ANDINA NATALIA	90	93	1	0	1
002061468	PUTRI ANGGRANI	88	90	1	0	1
002069449	EPRIKA SHOTANG	90	90	1	0	1
002061384	SAM DEVA NAGRA	93	95	1	1	1
002061449	ADELIA RAJAGUNGIK	88,5	92	1	1	1
002061447	SAPTA DEVI SARTIKA	80	90	1	1	1
002061466	NONI ANDRIYANI	79	87	1	0	1
002061471	PUTRI NABILA DWI RANITU	94	96	0	0	1

Gambar 6. Hasil Analisa

- Maksimum iterasi yang akan diolah

Maksimum Iterasi *



Gambar 7. Iterasi

- Perhitungan

Nama	Nilai UAS	Nilai Rapor	Status Rumah	Kondisi Rumah	Penghasilan Orang Tua
CO1	78	89	0	0	0
CO2	89	94	0	1	0
CO3	97	98	1	1	1

Gambar 8 . Hasil Iterasi 1

Nama	CO1	CO2	CO3
002061444	2,2111	11,2094	18,1025
002061448	3,3166	10,2956	19,0263
002061452	11,4891	3,3166	10,2427
002061441	12,7279	2,2361	9,5603
002061468	10,1489	4,4721	12,083
002069449	12,1244	4,4721	10,6171
002061384	16,3401	4,2309	5
002061449	11,0567	2,5	10,4043

Gambar 9. Jarak Terhadap Pusat Centroid

Nama	Nilai UAS	Nilai Rapor	Status Rumah	Kondisi Rumah	Penghasilan Orang Tua
CO1	79,8	91,0901	0,45453	0,72723	0,72727
CO2	88,3069	92,45317	0,5	0,3069	0,67241
CO3	95,13636	95,90909	0,45453	0,18182	0,81818

Gambar 10 . Hasil Iterasi 2

Nama	CO1	CO2	CO3
002061444	5,3141	10,8913	17,1658
002061448	1,2266	6,3098	16,0232
002061452	6,7099	3,6464	6,1192
002061441	10,2085	1,84	5,5237
002061468	6,0962	2,7467	6,2646
002069449	10,7999	3,212	7,8526
002061384	13,7983	5,3995	2,5279

Gambar 11. Jarak Terhadap Pusat Centroid Baru

Nama	Nilai UAS	Nilai Rapor	Status Rumah	Kondisi Rumah	Penghasilan Orang Tua
CO1	80,98154	92,38402	0,53846	0,23077	0,76523
CO2	87,806	92,42	0,44	0,2	0,66
CO3	94,76471	94,64708	0,38824	0,22529	0,76471

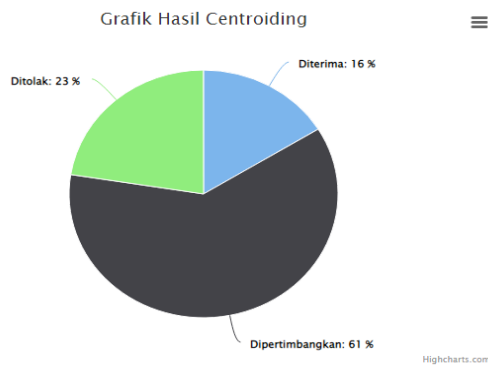
Gambar 12 . Hasil Iterasi 2



Nama	CO1	CO2	CO3
0002041444	4.2027	10.565	16.8403
0002041446	1.4604	6.9402	16.0107
0002041452	7.8843	3.7511	7.8995
0002041441	9.9735	2.2966	5.0691
0002041468	8.3081	2.5177	8.2342
0002041449	10.2362	3.3855	6.5767
0002041484	13.2126	5.8035	2.0121
0002041440	6.4979	1.3738	6.9603
0002041447	2.5589	8.3132	15.5049

Gambar 13. Jarak Terhadap Pusat Centroid Baru

Karena iterasi (3) sudah mencapai maksimum iterasi, maka iterasi dihentikan.



Gambar 14. Grafik Hasil Centroid

Berdasarkan hasil perhitungan yang telah dilakukan, maka didapatkan hasil yang diolah menjadi grafik. Adapun hasil Grafik *centroiding* adalah 16% Diterima, 61%, Dipertimbangkan dan 23% Ditolak. Dengan hasil diatas menunjukkan bahwa untuk penerima beasiswa banyak calon mahasiswa yang dipertimbangkan.

4. KESIMPULAN

Dari hasil pengolahan yang sudah dilakukan maka mendapatkan hasil bahwa banyak calon mahasiswa yang dipertimbangkan untuk penerima beasiswa yayasan, dengan adanya aplikasi ini membuat pekerjaan pihak panitia menjadi lebih ringan hanya dengan memasukkan data calon penerima kedalam aplikasi sesuai dengan kriteria-kriteria yang sudah ditentukan.

5. REFERENSI

Ginting, B., & Riandari, F. (2020). Implementasi Metode *K-Means Clustering* Dalam Pengelompokan Bibit Tanaman Kopi Arabika. *Journal of Innovation Information Technology and Application (JINITA)*, 2(02), 146–151. <https://doi.org/10.35970/jinita.v2i02.394>

Hasan, S., & Muhammad, N. (2020). Sistem Informasi Pembayaran Biaya Studi Berbasis Web Pada Politeknik Sains Dan Teknologi Wiratama Maluku Utara. *IJIS - Indonesian Journal On Information System*, 5(1), 44–55. <https://doi.org/10.36549/ijis.v5i1.66>

Manihuruk, N. A., Zarlis, M., Irawan, E., & Tambunan, H. S. (2020). Penerapan Data Mining Dalam Mengelompokkan Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 4(1), 29–34. <https://doi.org/10.30865/komik.v4i1.2575>

Putri, sannu ucha, Irawan, E., & Rizky, F. (2021). Implementasi Data Mining Untuk Prediksi Penyakit Diabetes. *KESATRIA(Jurnal Penerapan Sistem Informasi Dan Manajemen)*, 2(1), 39–46.

Putri, E. E. (2020). Analisis tingkat kejahatan jalanan menggunakan metode fp-growth (studi kasus di polresta padang). *Syntax: Journal of Software Engineering, Computer Science and Information Technology*, 1(1), 10-15.

Rahayu, A. E., Hikmah, K., Yustia, N., & Fauzan, A. C. (2019). Penerapan *K-Means Clustering* Untuk Penentuan Klasterisasi Beasiswa Bidikmisi Mahasiswa. *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, 1(2), 82–86. <https://doi.org/10.28926/ilkomnika.v1i2.23>

Rahmah, S. A. (2020). *KLASTERISASI POLA PENJUALAN PESTISIDA MENGGUNAKAN METODE K-MEANS CLUSTERING (STUDI KASUS DI TOKO JUANDA TANI KECAMATAN HUTABAYU RAJA)*. 1(1), 1–5. <http://jurnal.dharmawangsa.ac.id/index.php/djtechno/article/view/964>

Salam, A., Adiatma, D., & Zeniarja, J. (2020). Implementasi Algoritma K-Means Dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS. *JOINS (Journal of Information System)*, 5(1), 62–68. <https://doi.org/10.33633/joins.v5i1.3350>



Sudarsono, B. G., & Lestari, S. P. (2021). *Clustering Penerima Beasiswa Yayasan Untuk Mahasiswa*. 5, 258–263. <https://doi.org/10.30865/mib.v5i1.2670>

Teknologi, J., Jtsi, I., Provinsi, D. I., Dengan, L., & Means, A. K.-. (2021). *ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19*. 2(2), 100–108.