

METODE K-MEANS UNTUK CLUSTERING MAHASISWA BERDASARKAN NILAI AKADEMIK

K-MEANS METHOD FOR STUDENT CLUSTERING BASED ON ACADEMIC VALUE

Yassirli Amri

Universitas Muhammadiyah Sumatera Barat

Yassirliamri09@gmail.com

ABSTRAK : Proses seleksi mahasiswa calon peserta beasiswa berprestasi Universitas Muhammadiyah Sumatera Barat masih dilakukan secara manual. Sehingga, proses penyeleksian belum obyektif. Salah satu cara agar proses penyeleksian penerimaan beasiswa mahasiswa berprestasi berjalan objektif dengan meng-Cluster mahasiswa berdasarkan nilai IP, Penghasilan orang tua dan jumlah tanggungan orang tua. Tujuan penelitian ini membantu dalam penentuan calon mahasiswa penerima beasiswa berprestasi menggunakan metode K-Means. Data yang digunakan pada penelitian ini adalah data nilai mahasiswa fakultas ekonomi angkatan 2016 Prodi Manajemen semester 4 Tahun Akademik 2018/2019 yang digunakan dalam penelitian ini. Hasil dari penelitian ini adalah cluster layak sebanyak 7 orang dan cluster tidak layak sebanyak 13 orang sehingga penelitian ini dapat dijadikan rekomendasi pemilihan penerima beasiswa untuk masa yang akan datang.

Kata kunci: K-means Clustering, Beasiswa, Nilai Akademik.

ABSTRACT : *The selection process of outstanding scholarship prospective student Muhammadiyah University of West Sumatra is still done manually. So, the selection process is not objective. One way for the selection process of student scholarship to achieve is objective by clustering students based on IP scores, parents' income and the number of dependents of parents. The purpose of this study is to assist in the determination of prospective scholarship recipients to achieve using the K-Means method. The data used in this study is the data value of the students of the economic faculty in class 2016 Management Study Program 4th semester of Academic Year 2018/2019 used in this study. The results of this study are 7 decent clusters and 13 unlucky clusters so that this research can be used as a recommendation for future scholarship recipients.*

Keywords: *K-means Clustering, Scholarship, Academic Value*

A. PENDAHULUAN

Setiap universitas memiliki banyak skema bantuan pendidikan bagi mahasiswa yang dapat menunjukkan prestasi dalam bidang akademik. Banyaknya dana pendidikan yang diberikan pemerintah dan swasta kepada para mahasiswa membuat staf kemahasiswaan kewalahan dan kebingungan dalam menyalurkan beasiswa kepada yang berhak (Dina & Marjianto, 2018). Selama ini proses penyeleksian mahasiswa calon penerima beasiswa berprestasi di Universitas Muhammadiyah Sumatera Barat masih dilakukan secara manual sehingga proses penyeleksian belum obyektif. Salah satu cara agar proses penyeleksian penerimaan beasiswa mahasiswa berprestasi berjalan objektif dengan meng-Cluster mahasiswa berdasarkan nilai IP, Penghasilan orang tua dan jumlah tanggungan orang tua. Dimana nilai afektif diambil

Salah satu cara agar proses penyeleksian penerimaan beasiswa mahasiswa berprestasi berjalan objektif dengan meng-Cluster mahasiswa berdasarkan nilai IP, Penghasilan orang tua dan jumlah tanggungan orang tua. Nilai IP diambil dari KHS Semester IV, sedangkan data penghasilan orang tua dan jumlah tanggungan orang tua didata langsung kepada mahasiswa. Cluster atau Clustering adalah cara untuk menemukan

kelompok objek yang memiliki kemiripan dan dapat menemukan pola penyebaran dan pola hubungan dalam kumpulan data yang besar. Dalam proses clustering yang terpenting adalah mengumpulkan pola ke kelompok yang sesuai untuk menemukan persamaan dan perbedaan agar menghasilkan kesimpulan yang berharga (Asroni, Fitri, H dan Prasetyo, E, 2018). Clustering bermanfaat untuk melakukan analisis pola-pola data, mengelompokkan, dan membuat keputusan (Eko, 2018).

Dalam proses cluster mahasiswa peneliti menggunakan metode K-Means . Dalam proses cluster mahasiswa peneliti menggunakan metode K-Means. K-Means merupakan algoritma clustering yang berulang-ulang algoritma K-Means dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya cluster yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari cluster atau biasa disebut dengan centroid, mean atau “means”. Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus Euclidian hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. (Putra, R, R dan Wadisman, 2018).

Metode k-means memudahkan peneliti dalam menentukan mahasiswa penerima beasiswa mencari hasil dalam penentuan penerimaan beasiswa berprestasi karna proses clustering didasari dengan nilai IP, penghasilan orang tua dan jumlah tanggungan orang tua. Penelitian ini bertujuan dalam membantu pihak universitas dalam menentukan mahasiswa calon penerima beasiswa. Metode yang digunakan dalam penelitian ini adalah metode K-Means.

B. METODE PENELITIAN

Data yang digunakan pada penelitian ini adalah berupa nilai IP, nilai afektif, penghasilan orang tua dan jumlah tanggungan. Data penghasilan orang tua sudah di transformasi oleh peneliti kedalam scope yang di inginkan. Semua data tersebut di robah kedalam format .xls agar memudahkan peneliti dalam melakukan pengolahan data dengan menggunakan metode K-means. Sesuai dengan tabel 2.1 yaitu

2.1 Tabel Data

No	Nama	Nilai IP	Nilai Afektif	Penghasilan Ortu	Jlm tanggungan
1	ayu putriani	2,88	3,00	1.000.000	1
2	cici oktavia	3,17	4,00	2.000.000	4
3	delvia	3,33	3,00	3.000.000	2
4	devi indriani	3,00	3,00	1.000.000	4
5	dian mustika mega	1,86	0,00	2.000.000	3
6	eka putri yanti	3,92	4,00	1.000.000	2
7	ella gusnita	2,86	3,00	2.000.000	3
8	gusrinaldi	2,70	2,00	2.000.000	2

9	indah febrina	3,33	3,00	1.000.000	3
10	jelli	3,79	3,00	2.000.000	2
11	nini astika	3,42	4,00	≤ 1.000.000	4
12	pratiwi eka putri	3,21	3,00	2.000.000	2
13	resa fitri yanti	3,08	3,00	1.000.000	3
14	rio yulio candra	3,63	4,00	≤ 1.000.000	4
15	selvia putri	3,46	3,00	1.000.000	4
16	tiara isti rana	4,00	4,00	3.000.000	4
17	tria septria richi	3,33	3,00	3.000.000	2
18	umi jumyati	3,58	3,00	3.000.000	2
19	wahyu irawan	2,95	3,00	≤ 1.000.000	4
20	wandi kurniawan	3,14	3,00	2.000.000	2

2.1. Tahap Pengolahan Data

Data yang telah diperoleh akan diolah terlebih dahulu untuk dapat diclustering. Dalam tahap sebelumnya, data akan dijumlah setiap aspeknya sehingga pada tahapan ini sudah diperoleh perhitungan nilai berdasarkan nilai akademik.

2.2 Tahap Analisis

Pada tahapan ini dilakukan analisis data mahasiswa berdasarkan nilai akademik mahasiswa menggunakan aplikasi tool Weka. Weka merupakan perangkat lunak yang memiliki banyak algoritma machine learning untuk keperluan data mining . Data yang diperoleh diolah dengan menggunakan perhitungan bobot dari tiap indeks kepuasan. Pada tahapan sebelumnya, telah ditentukan akan dicluster ke dalam 2 cluster yakni cluster_layak dan cluster_tidak layak. Pada tahapan inilah akan dianalisis hasilnya.

C. HASIL DAN PEMBAHASAN

Dalam melakukan clusterisasi data yang akan dihitung terlebih dahulu tentukan bobot yang akan ditentukan. Data penghasilan orang tua di transformasi terlebih dahulu agar memudahkan peneliti dalam pengolahan sesuai tabel 3.1

3.1 Data Tranformasi

Penghasilan orang tua	Frekuensi	Inisial

$\leq 1.000.000$		3	4
1.000.000 2.000.000	-	6	2
2.000.000 3.000.000	-	7	1
3.000.000 4.000.000	-	4	3
$\geq 4.000.000$		0	5

3.2 Algoritma K-means

Algoritma k-means adalah metode clustering yang paling sederhana dan umum di gunakan dalam berbagai penyelesaian masalah dalam kehidupan sehari hari. Hal ini dikarenakan K-means mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien. Secara garis besar tahapan algoritma *K-means* adalah sebagai berikut:

3.2.1 Iterasi Pertama

Iterasi pertama merupakan tahapan yang dilakukan untuk mendapatkan hasil pengelompokkan terhadap nilai pusat *cluster*. Sebagaimana dijelaskan sebelumnya mengenai tahapan algoritma *K-means*, pada tahap *iterasi* dilakukan proses inialisasi pusat *cluster*, pemilihan pusat klaster awal, menghitung jarak, dan pengelompokan.

3.2.1.1 Menentukan Jumlah Cluster

Dalam penelitian ini dilakukan pengelompokan berdasarkan data mahasiswa berdasarkan nilai akademik, penghasilan orang tua dan jumlah tanggungan orang tua.

Adapun jumlah cluster atau kelompok di bagi menjadi dua cluster yaitu:

1. C0 menggambarkan cluster mahasiswa yang layak menerima beasiswa berprestasi
2. C1 menggambarkan cluster mahasiswa yang tidak layak menerima beasiswa berprestasi.

Inialisasi pusat klaster (*cendroid*) merupakan tahap awal pada algoritma *K-means*. Pada penelitian ini, dalam menentukan pusat klaster awal diperoleh secara acak. Bilangan acak dipilih berdasarkan dari data sampel yang telah dipilih untuk menentukan pusat klaster yang dibutuhkan oleh masing-masing kelompok. Untuk perhitungan manual, maka untuk pusat klaster awal dari penelitian ini adalah C0 diambil dari data sampel ke-2, C1

diambil data sampel ke-13. Berikut ini adalah pusat kluster yang telah penulis buat dalam bentuk tabel 3.2

3.2 Nilai Centroid Awal

Pusat kluster awal				
	Nilai IP	Nilai Afektif	Penghasilan Orang Tua	Jumlah Tanggungan
C0	3,17	4,00	1	4
C1	2,70	2,00	1	2

3.2.1.2 Menghitung Jarak Iterasi Pertama

Pada penelitian ini, penulis menggunakan rumus *Euclidean* untuk menghitung jarak antar objek terhadap pusat kluster dengan rumus berikut:

$$d_{(x,y)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

D = Titik Dokumen

X = Data Record

Y = Data Centroid

$$X_1 = (2,88 : 3,00 : 2 : 1) \quad C_0 = (3,17 : 4,00 : 1 : 4)$$

$$D_{11} = \sqrt{(2,88 - 3,17)^2 + (3,00 - 4,00)^2 + (2 - 1)^2 + (1 - 4)^2} \\ = 3,329$$

$$X_2 = (3,17 : 4,00 : 1 : 4) \quad C_0 = (3,17 : 4,00 : 1 : 4)$$

$$D_{11} = \sqrt{(3,17 - 3,17)^2 + (4,00 - 4,00)^2 + (1 - 1)^2 + (4 - 4)^2} \\ = 0$$

$$X_3 = (3,33 : 3,00 : 3 : 2) \quad C_0 = (3,17 : 4,00 : 1 : 4)$$

$$D_{11} = \sqrt{(3,33 - 3,17)^2 + (3,00 - 4,00)^2 + (3 - 1)^2 + (2 - 4)^2} \\ = 3,004$$

Selanjutnya Perhitungan C1 seperti berikut ini :

$$X_1 = (2,88 : 3,00 : 2 : 1) \quad C_1 = (2,70 : 2,00 : 1 : 2)$$

$$D_{11} = \sqrt{(2,88 - 2,70)^2 + (3,00 - 2,00)^2 + (2 - 1)^2 + (1 - 2)^2} = 1,7349$$

$$X_2 = (3,17 : 4,00 : 1 : 4) \quad C_1 = (2,70 : 2,00 : 1 : 2)$$

$$D_{11} = \sqrt{(3,17 - 2,70)^2 + (4,00 - 2,00)^2 + (1 - 1)^2 + (4 - 2)^2} = 2,8551$$

Pada kasus ini peneliti melakukan 4 kali iterasi agar hasilnya sudah tidak perlu dilakukan iterasi lagi. Dimana hasil akhirnya terbentuk 2 cluster dengan cluster 0 sebanyak 7 orang dan cluster 1 sebanyak 13 orang. Hasil diatas dapat dilihat pada tabel 3.3

Tabel 3.3 hasil

	Nilai		Nama Mahasiwa	Layak / Tidak Layak
	C0	C1		
Cluster 1	3,210	1,299	Ayu Putriani	Tidak Layak
Cluster 0	1,917	2,295	Cici Oktavia	Layak
Cluster 1	2,085	1,265	Delvia	Tidak Layak
Cluster 0	1,096	1,804	Devi Indriani	Layak
Cluster 1	4,416	3,251	Dian Mustika Mega	Tidak Layak
Cluster	2,283	1,479	Eka Putri Yanti	Tidak Layak
Cluster 1	2,245	1,165	Ella Gusnita	Tidak Layak
Cluster 1	3,205	1,187	Gusrinaldi	Tidak Layak
Cluster 1	1,436	0,848	Indah Febrina	Tidak Layak
Cluster 1	2,818	1,060	Jelli	Tidak Layak
Cluster 0	1,221	3,099	Nini Astika	Layak
Cluster 1	2,793	0,855	Pratiwi Eka Putri	Tidak Layak
Cluster 1	1,465	0,834	Resa Fitri Yanti	Tidak Layak

Cluster 0	1,246	3,123	Rio Yulio Candra	Layak
Cluster 0	1,033	1,822	Selvia Putri	Layak
Cluster 0	0,769	2,611	Tiara Isti Rana	Layak
Cluster 1	2,085	1,265	Tria Septria Richi	Tidak Layak
Cluster 1	2,094	1,322	Umi Jumyati	Tidak Layak
Cluster 0	1,344	2,848	Wahyu Irawan	Layak
Cluster 1	2,798	0,853	Wandi Kurniawan	Tidak Layak

D. KESIMPULAN

Berdasarkan penelitian dan pembahasan maka dapat disimpulkan yaitu penelitian ini terdiri dari 20 data, terdapat 2 cluster yaitu cluster 0 dan cluster 1 dimana cluster 0 sebanyak 3 orang mahasiswa dan cluster 1 sebanyak 13 orang mahasiswa. Penelitian ini dapat membantu pihak fakultas dalam penentuan calon mahasiswa penerima beasiswa berprestasi, agar proses penseleksian lebih selektif dan tepat sasaran. Kemudian penelitian ini bermanfaat bagi penulis untuk mengetahui tingkat akurasi data yang diuji dengan menggunakan Software Weka.

E. DAFTAR PUSTAKA

- Dina, N, Z ., Marjianto, R, S. (2018)**, “*Prediksi Penentuan Penerima Beasiswa Dengan Metode Knearest Neighbours*”, Jurnal Nasional Informatika dan Teknologi Jaringan, Vol. 2 No.2, <https://doi.org/10.30743/infotekjar.v2i2.269>
- Asroni, Fitri, H dan Prasetyo, E. (2018)**, “ *Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)*”, Semesta Teknika, Vol. 21. No. 1, 60-64 , Mei 2018, <https://10.18196/st.211211>
- Eko Yulian. (2018)**, “ *Text Mining dengan K-Means Clustering pada Tema LGBT dalam Arsip Tweet Masyarakat Kota Bandung*, Vol. 04 No. 01, <https://doi.org/10.15642/mantik.2018.4.1.53-58>

- Sulastrri, H. dan Gufroni, A. I. (2017)**, “Penerapan Data Mining Dalam Pengelompokan Penderita *thalassaemia*”, Jurnal Teknologi dan Sistem Informasi VOL. 03 NO. 02 (2017) 299-305, <https://dx.doi.org/10.25077/TEKNOSI.v3i2.2017.299-305>
- Jiang, Y., Rosenweig, E, Q dan Gaspard, H. (2018)** “An expectancy-value-cost approach in predicting adolescent students academic motivation and achievement”, Contemporary Educational Psychology 54 (2018) 139–152, <https://doi.org/10.1016/j.cedpsych.2018.06.005>
- Li, Z., Shang, C dan Shen, Q. (2016)** “Fuzzy-clustering embedded regression for predicting student academic performance”, Institute of Mathematics, Physics and Computer Science, DOI: [10.1109/FUZZ-IEEE.2016.7737707](https://doi.org/10.1109/FUZZ-IEEE.2016.7737707)
- Zanuardi, A dan Supriyanto, H. (2018)**, “Analisa Karakteristik Kecelakaan Lalu Lintas di Jalan Ahmad Yani Surabaya melalui Pendekatan Knowledge Discovery in Database”, Jurnal Manajemen Aset Infrastruktur & Fasilitas, Vol. 2, No. 1, Maret, <http://www.iptek.its.ac.id/index.php/jmaif/article/view/3767>
- Chandra, A. (2017)**, “Penerapan Data Mining Menggunakan Pohon Keputusan Dengan Algoritma c4.5 dalam Menentukan Kecelakaan Penerbangan”, STMIK AMIKOM Yogyakarta, 4 Februari, <https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/view/1768>
- Verano, D., A (2016)**, “Assosiasi Rules Dan Moving Average Untuk Memprediksi Persediaan Bahan Baku Produksi”, ANNUAL RESEARCH SEMINAR 2016, <http://www.seminar.ilkom.unsri.ac.id/index.php/ars/article/view/958>
- Widayu H, dkk. (2017)**, “Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma c4.5”, media informatika budidarma, vol 1, no 2, juni, <https://ejurnal.stmikbudidarma.ac.id/index.php/mib/article/view/323>
- Herasmus, H. (2017)**, “Analisa Customer Service System Menggunakan Metode Data Mining Dengan Algoritma FP-Growth (Studi Kasus Di PT Batamindo Investment Cakrawala)”, Jurnal Teknik Ibnu Sina JT-IBSI, Vol 2, No 2 <http://ojs.stt-ibnusina.ac.id/index.php/JT-IBSI/article/view/57>
- Rerung, R., R. (2018)**, “Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk” JTERA - Jurnal Teknologi Rekayasa, Vol. 3, No. 1, Juni, Hal. 89-98, DOI: [10.31544/jtera.v3.i1.2018.89-98](https://doi.org/10.31544/jtera.v3.i1.2018.89-98)
- Febriana, M., D, Sarita, M., I dan Tajidun L, M(2017)**, “Aplikasi Clustering Data Serapan Alumni Di Dunia Kerja Menggunakan Metode Hierarchical Clustering”, semanTIK, Vol.3, No.2, Jul-Des, pp. 175-180,
- Goreti1, M., Novia, Y dan Wahyuningsih, S (2016)**, “Perbandingan Hasil Analisis Cluster dengan Menggunakan Metode Single Linkage dan Metode C-Means

(Studi Kasus: Data Tingkat Kualitas Udara Ambien pada Perusahaan Perkebunan di Kabupaten Kutai Barat Tahun 2014)” Jurnal EKSPONENSIAL Volume 7, Nomor 1, Mei, <http://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/15>

Jumailah, F.S. dan Pratama. D., (2018), “*Identifikasi Cluster Penduduk Usia Kerja Pada Provinsi Sumatera Selatan Menggunakan K-Modes*”, Jurnal Komputer Terapan Vol 4 No 1 Mei, <https://jurnal.pcr.ac.id/index.php/jkt/article/view/1629>

Nurzahputra, A., Muslim, M.,A dan Khusniati, M (2017), “*Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa*” Techno.COM, Vol. 16, No. 1, Februari: 17-24, <http://publikasi.dinus.ac.id/index.php/technoc/article/view/1284>