

K-MEANS UNTUK MENENTUKAN CALON PENERIMA BEASISWA BIDIK MISI DI POLBENG

Jaroji¹, Danuri², Fajri Profesio Putra³
Politeknik Negeri Bengkalis, Jl. Bathin Alam, Sungai Alam, Bengkalis
*jaroji@polbeng.ac.id*¹, *danuri@polbeng.ac.id*², *fajri@polbeng.ac.id*³

Abstract – Distributing of Bidik Misi scholarships to students who come from poor families have to go through the selection involving certain criteria. Such criteria such as parental income, home ownership status, housing conditions, the amount of dependent parent, parental status and academic achievement. K-means algorithm can help classify the students were very decent, decent, with consideration and less deserving mission viewfinder. The purpose of this study is to determine clustering scholarship applicants so as to provide recommendations very worthy, worthy, with consideration and less worthy to receive a scholarship to the 6 criteria. The data set used by 129 instances are derived from the selection committee bidik misipoliteknik negeri bengkalis (polbeng). These data were calculated using the K-Means algorithm and the testing done using Weka tool. From the test results generated many as 17 people recommended by the consideration, 24 were recommended very decent, decent recommended 32 people and 56 people recommended less feasible.

Keywords–*Clustering, K-Means, Bidik Misi*

Intisari – Penyaluran beasiswa bidik misi kepada mahasiswa yang berasal dari keluarga kurang mampu harus melalui seleksi yang melibatkan kriteria-kriteria tertentu. Kriteria tersebut seperti penghasilan orang tua, status kepemilikan rumah, kondisi rumah, jumlah tanggungan orang tua, status orang tua dan prestasi akademik. Algoritma *k-means* dapat membantu mengklasifikasikan mahasiswa yang sangat layak, layak, dengan pertimbangan dan kurang layak menerima bidik misi. Adapun tujuan penelitian ini adalah menentukan *clustering* pelamar beasiswa sehingga dapat memberikan rekomendasi sangat layak, layak, dengan pertimbangan dan kurang layak untuk menerima beasiswa dengan 6 kriteria. Data set yang digunakan sebanyak 129 instance yang berasal dari panitia seleksi bidik misi politeknik negeri bengkalis (polbeng). Data-data tersebut dihitung dengan menggunakan algoritma K-Means dan pengujianya dilakukan dengan menggunakan tool Weka. Dari hasil pengujian dihasilkan sebanyak 17 orang direkomendasi dengan pertimbangan, 24 orang direkomendasikan sangat layak, 32 orang direkomendasikan layak dan 56 orang direkomendasikan kurang layak.

Kata kunci – Klasifikasi, K-Means, Bidik misi

I. PENDAHULUAN

Bidik misi merupakan beasiswa yang diperuntukan untuk mahasiswa yang berasal dari keluarga tidak mampu. Beasiswa tersebut disalurkan melalui perguruan tinggi di bawah kementerian riset teknologi dan perguruan tinggi (Ristek Dikti). Seleksi pelamar bidik misi dilakukan pada saat proses penerimaan mahasiswa baru oleh perguruan tinggi yang

ditunjuk oleh Ristek Dikti. Pemberian beasiswa ini tentunya juga harus memperhatikan kriteria - kriteria tertentu sebelum di berikan kepada mahasiswa yang bersangkutan [4]. Kriteria tersebut seperti penghasilan orang tua, status kepemilikan rumah, kondisi rumah, jumlah tanggungan orang tua, status orang tua dan prestasi akademik. Algoritma *k-means* dapat membantu mengklasifikasikan mahasiswa

yang sangat layak, layak, dengan pertimbangan dan kurang layak menerima bidik misi.

Pada penelitian yang dilakukan oleh Nurul Rohmawati W (2015), yaitu menerapkan Algoritma *k-means* dan *k-medoids* dari teknik *clustering* untuk membantu dalam mengklasifikasi mahasiswa untuk menentukan penerima beasiswa dengan membagi data mahasiswa menjadi 3 *cluster* dengan 4 kriteria. Bahwa tingkat akurasi *clustering* hasil *cluster* algoritma *k-means* berdasarkan nilai purity measure, dataset yang dikodifikasi keseluruhan lebih baik dari pada dataset yang di kodifikasi sebagian dan dataset data asli [4].

Adapun tujuan penelitian ini adalah menentukan *clustering* mahasiswa yang sangat layak, layak, dengan pertimbangan dan kurang layak untuk menerima beasiswa dengan 6 kriteria.

II. SIGNIFAKASI STUDI

A. Tinjauan Pustaka

Permasalahan klasifikasi banyak dibahas dalam penelitian. Suprawoto(2016), Pengelompokan data mahasiswa menggunakan metode K-Means, terdiri dari nilai rerata ujian nasional (UN) dan indeks prestasi kumulatif (IPK) mahasiswa. Penelitian menggunakan data mahasiswa angkatan 2014/2015. Kemudian diperoleh kesimpulan bahwa kelompok mahasiswa dengan nilai rerata UN yang rendah memiliki pengaruh terhadap prestasi akademik mahasiswa yang rendah pada jenjang diploma-3(D-3) dan strata-1 (S-1). Jika mahasiswa memiliki nilai UN yang tinggi maka prestasi akademik mahasiswa juga tinggi pada semua jenjang. Dari hasil pengelompokan berdasarkan daerah asal sekolah IPK rata-rata yang tertinggi berasal dari propinsi Daerah Istimewa Yogyakarta (DIY) dan Jawa Tengah (Jateng). [6]

Pengelompokan menggunakan *k-means* diterapkan untuk mengelompokkan dokumen web trunojoyo.ac.id. Dokumen

web semakin besar, belum tertata dengan baik, belum terstruktur atau terkategori mengikuti kaidah tertentu dan tersebar pada banyak sub-domain. Sejauh ini, tidak ada gerbang atau portal web yang menyediakan akses ke berbagai situs web yang dihosting oleh data center PTIK Universitas Trunojoyo. Salah satu masalah yang telah diselesaikan adalah pengelompokan informasi atau berita web tersebut secara otomatis menggunakan algoritma *clustering* K-Means. Search engine RISE yang telah berjalan menghimpun semua halaman web yang ditulis dalam bahasa Indonesia di bawah domain trunojoyo.ac.id menggunakan teknik crawling. Halaman-halaman tersebut kemudian dipre-processing menggunakan teknik standar dalam text mining (information retrieval). Proses utamanya adalah penerapan teknik *k-means* sehingga terbentuk kelompok-kelompok berita otonom. Pengujian yang telah dilakukan menunjukkan bahwa teknik *clustering* yang diterapkan mampu bekerja dengan baik dan memberikan akurasi yang memuaskan. Ada sekitar 300 halaman web yang dilibatkan dalam proses *clustering* dimana diperoleh ukuran rata-rata F-Measure sebesar 0.6129192 dan Purity bernilai 0.67294195. Faktor yang cukup berpengaruh dalam *clustering* dan klasifikasi teks bahasa Indonesia adalah fase pre-processing, terutama pada pendekatan stemming. Perbaikan terhadap teknik stemming diyakini akan meningkatkan akurasi pengelompokan dokumen. (Husni, dkk., 2015). [2]

Pengelompokan data pelamar beasiswa Bantuan Belajar Mahasiswa (BBM) dikelompokkan menjadi 3 kelompok yaitu berhak menerima, dipertimbangkan dan tidak berhak menerima beasiswa. Pengelompokan menjadi 3 kelompok ini berguna untuk memudahkan dalam menentukan penerima beasiswa BBM. Algoritma *k-means* merupakan algoritma dari teknik *clustering* yang berbasis partisi. Teknik ini dapat mengelompokkan data mahasiswa pelamar beasiswa. Tujuan dari

penelitian ini adalah untuk pengukuran kinerja algoritma, Pengukuran ini di lihat dari hasil *cluster* dengan menghitung nilai kemurnian (purity measure) dari masing – masing *cluster* yang di hasilkan. Data yang digunakan dalam penelitian ini adalah data mahasiswa yang mengajukan beasiswa kepada Fakultas Ilmu Komputer UNSIKA sebanyak 36 mahasiswa. Data akan diubah menjadi 3 dataset dengan format yang berbeda-beda, yakni data atribut kodifikasi sebagian, atribut kodifikasi keseluruhan dan atribut data asli. Nilai purity pada dataset data kodifikasi sebagian untuk hasil *cluster* algoritma kmeans sebesar 61.11%. Pada dataset kodifikasi keseluruhan nilai purity hasil *cluster* algoritma kmeans sebesar 80.56%. Dan untuk dataset data asli nilai purity hasil *cluster* algoritma k-means sebesar 75%. Maka dapat di simpulkan bahwa algoritma kmeans lebih cocok digunakan pada dataset dengan format atribut yang dikodifikasi keseluruhan (Rohmawati.W, 2015).[4

B. Bahan Penelitian

Penelitian ini menggunakan data calon mahasiswa yang mengikuti seleksi penerimaan beasiswa bidik misi tahun akademik 2015/2016. Sample data terdiri atas 30% dari total pendaftar beasiswa bidik misi yang sudah disurvei. Data diperoleh dari UPT. Kemahasiswaan Politeknik Negeri Bengkulu

C. Metodologi Penelitian

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 adalah sebagai berikut (Sarwono dalam Rohmawati,W., 2015):

1. Menentukan k sebagai jumlah *cluster* yang ingin di bentuk.
2. Membangkitkan nilai random untuk pusat *cluster* awal (centroid) sebanyak k.

3. Menghitung jarak setiap data input terhadap masing – masing centroid menggunakan rumus jarak Euclidean (Euclidean Distance) hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Berikut adalah persamaan Euclidian Distance:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \quad (1)$$

Dimana : x_i : data kriteria, μ_j : centroid pada *cluster* ke-j

4. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid (jarak terkecil).
5. Memperbaharui nilai centroid. Nilai centroid baru di peroleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan rumus:

$$\mu_j(t+1) = \frac{1}{N_{S_j}} \sum_{j \in S_j} x_j \quad (2)$$

Dimana: $\mu_j(t+1)$: centroid baru pada iterasi ke (t +1) N_{S_j} : banyak data pada *cluster* S_j .

6. Melakukan perulangan dari langkah 3 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah.

Jika langkah 6 telah terpenuhi, maka nilai pusat *cluster* (μ_j) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data

III. PEMBAHASAN

A. Data Pengujian

Data pengujian yang digunakan berupa data pelamar beasiswa bidik misi yang telah disurvei dengan komponen penyusun data sebagai berikut:

1. Memiliki 8 atribut yaitu nisn, nama, penghasilan (PH), status rumah (SR), kondisi rumah (KR), tanggungan (TG), status orang tua (SO) dan akademik (AK).
2. Jumlah instance adalah 129.

B. Perhitungan Clustering

Selanjutnya digunakan algoritma klasifikasi k-mean untuk mengelompokan

data yang ada. Perhitungan *clustering* mengikuti langkah-langkah berikut:

1. Menetapkan jumlah *cluster*. Jumlah *cluster* ditetapkan berdasarkan rekomendasi hasil seleksi beasiswa bidik misi yaitu Sangat Layak (SL), Layak (L), Dengan Pertimbangan (DP) dan kurang layak. Oleh karena itu jumlah *cluster* yang ditetapkan adalah 4 *cluster*.
2. Inisialisasi pusat *cluster* (*centroid*) secara acak berdasarkan data skor hasil survei lapangan. Nilai pusat *cluster* dapat dilihat pada tabel berikut:

TABEL I INISIALISASI CLUSTER

Cluster	PH	SR	KR	TG	SO	AK
0	60	20	100	20	25	75
1	80	60	60	40	25	45
2	60	20	100	40	25	75
3	60	20	30	20	25	60

3. Hitung jarak setiap data terhadap pusat cluster. Misalnya untuk menghitung

jarak instance pertama dengan pusat cluster pertama adalah [6]:

$$d_{11} = \frac{80 - 60^2 + 20 - 20^2 + 100 - 100^2}{+ (60 - 20)^2 + (25 - 25)^2 + (100 - 75)^2} = 54,9049$$

Jarak instance pertama dengan pusat cluster kedua adalah:

$$d_{12} = \frac{80 - 80^2 + 20 - 60^2 + 100 - 60^2}{+ (60 - 40)^2 + (25 - 25)^2 + (100 - 45)^2} = 81,3941$$

Jarak instance pertama dengan pusat cluster ketiga adalah:

$$d_{13} = \frac{80 - 60^2 + 20 - 20^2 + 100 - 100^2}{+ (60 - 40)^2 + (25 - 25)^2 + (100 - 75)^2} = 37,7492$$

Jarak instance pertama dengan pusat cluster keempat adalah:

$$d_{14} = \frac{80 - 60^2 + 20 - 20^2 + 100 - 30^2}{+ (60 - 20)^2 + (25 - 25)^2 + (100 - 60)^2} = 92,1954$$

Berikut adalah perhitungan lengkap 20 instance pertama dari 129 instance.

TABEL II HASIL PERHITUNGAN JARAK 20 INSTANCE PERTAMA

Instance	PH	SR	KR	TG	SO	AK	C0	C1	C2	C3
0	80	20	100	60	25	100	54,9049	81,3941	37,7492	92,1954
1	60	60	100	40	25	75	48,8830	53,8516	40,0000	84,4097
2	60	20	100	40	25	75	28,0988	67,0820	0,0000	74,3303
3	60	20	100	20	25	75	19,7368	70,0000	20,0000	71,5891
4	60	100	60	40	25	75	93,7526	53,8516	89,4427	89,0225
5	60	20	60	60	25	100	64,9195	73,6546	51,2348	64,0312
6	60	20	60	60	25	100	64,9195	73,6546	51,2348	64,0312
7	80	20	100	40	25	75	34,4898	64,0312	20,0000	76,9740
8	60	20	100	40	25	75	28,0988	67,0820	0,0000	74,3303
9	60	20	100	40	25	75	28,0988	67,0820	0,0000	74,3303
10	60	20	100	40	25	60	31,8519	61,8466	15,0000	72,8011
11	60	20	30	40	25	100	79,4641	76,9740	74,3303	44,7214
12	60	20	100	40	25	100	37,6104	81,3941	25,0000	83,0662
13	80	20	60	40	25	100	58,4341	68,0074	51,2348	57,4456
14	60	20	100	40	25	30	53,0523	61,8466	45,0000	78,7401
15	80	20	100	40	25	100	42,5975	78,8987	32,0156	85,4400
16	80	20	100	40	75	30	61,1626	76,9740	70,1783	95,3939
17	60	20	60	40	25	60	51,1326	47,1699	42,7200	36,0555
18	80	20	60	40	75	60	59,5051	65,7647	68,7386	64,8074
19	60	20	100	20	25	15	63,1628	70,0000	63,2456	83,2166

4. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid :

TABEL III POSISI CLUSTER PADA ITERASI PERTAMA

Instance	PH	SR	KR	TG	SO	AK	C0	C1	C2	C3	Cluster
0	80	20	100	60	25	100	54,9049	81,3941	37,7492	92,1954	C2
1	60	60	100	40	25	75	48,8830	53,8516	40,0000	84,4097	C2
2	60	20	100	40	25	75	28,0988	67,0820	0,0000	74,3303	C2
3	60	20	100	20	25	75	19,7368	70,0000	20,0000	71,5891	C0

Instance	PH	SR	KR	TG	SO	AK	C0	C1	C2	C3	Cluster
4	60	100	60	40	25	75	93,7526	53,8516	89,4427	89,0225	C1
5	60	20	60	60	25	100	64,9195	73,6546	51,2348	64,0312	C2
6	60	20	60	60	25	100	64,9195	73,6546	51,2348	64,0312	C2
7	80	20	100	40	25	75	34,4898	64,0312	20,0000	76,9740	C2
8	60	20	100	40	25	75	28,0988	67,0820	0,0000	74,3303	C2
9	60	20	100	40	25	75	28,0988	67,0820	0,0000	74,3303	C2
10	60	20	100	40	25	60	31,8519	61,8466	15,0000	72,8011	C2
11	60	20	30	40	25	100	79,4641	76,9740	74,3303	44,7214	C3
12	60	20	100	40	25	100	37,6104	81,3941	25,0000	83,0662	C2
13	80	20	60	40	25	100	58,4341	68,0074	51,2348	57,4456	C2
14	60	20	100	40	25	30	53,0523	61,8466	45,0000	78,7401	C2
15	80	20	100	40	25	100	42,5975	78,8987	32,0156	85,4400	C2
16	80	20	100	40	75	30	61,1626	76,9740	70,1783	95,3939	C0
17	60	20	60	40	25	60	51,1326	47,1699	42,7200	36,0555	C3
18	80	20	60	40	75	60	59,5051	65,7647	68,7386	64,8074	C0
19	60	20	100	20	25	15	63,1628	70,0000	63,2456	83,2166	C0

5. Memperbaharui nilai *centroid*. Sebagai contoh untuk *cluster* pertama (C0) terdapat 4 instance, yaitu instance ke 3, 16, 18 dan 19 dengan catatan jika semua (129) *instance* ditampilkan maka C0 akan memiliki jumlah *instance* (anggota cluster) yang berbeda. Berikut contoh perhitungan untuk memperbaharui *centroid* dari 20 *instance* [4].

$$C_{CPH} = \frac{60 + 80 + 80 + 60}{4} = 70,0$$

$$C_{LSR} = \frac{20 + 20 + 20 + 20}{4} = 20,0$$

$$C_{CKR} = \frac{100 + 100 + 60 + 100}{4} = 90,0$$

$$C_{CTG} = \frac{20 + 40 + 40 + 20}{4} = 30,0$$

$$C_{CSO} = \frac{25 + 75 + 75 + 25}{4} = 50,0$$

$$C_{CAK} = \frac{75 + 30 + 60 + 15}{4} = 45,0$$

Lakukan hal yang sama untuk C1, C2 dan C3. Berikut adalah hasil perhitungan lengkap untuk 4 *cluster* dari 20 *instance*.

TABEL IV PEMBAHARUAN CLUSTER PADA ITERASI PERTAMA

Cluster	PH	SR	KR	TG	SO	AK
0	70,0	20,0	90,0	30,0	50,0	45,0
1	60,0	100,0	60,0	40,0	25,0	75,0
2	66,1538	23,0769	90,7692	44,6154	25,0	81,9231
3	60,0	20,0	45,0	40,0	25,0	80,0

6. Ulangi langkah 3 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah. Pada penelitian ini, iterasi berhenti pada iterasi ke 8.

C. Pengujian Dengan Tool Weka

Software yang digunakan dalam penelitian ini adalah Weka. Tujuan dari penggunaan software ini adalah membandingkan hasil dengan perhitungan secara teoritis dengan hasil yang didapatkan dengan proses di Weka Interface ini. Alat penelitian Weka Interface adalah aplikasi data mining open source berbasis Java. Aplikasi ini dikembangkan pertama kali oleh Universitas Waikato di Selandia Baru. Weka memiliki banyak algoritma machine learning yang dapat digunakan untuk melakukan generalisasi atau formulasi dari sekumpulan data sampling. Salah satunya adalah clustering dengan menggunakan algoritma K-Means. (Asroni, 2015) [1].

Pengujian data instance dengan tool weka menghasilkan data berupa:

1. Nilai inisialisasi awal centroid, terlihat pada gambar 1.
2. Nilai akhir centroid pada iterasi ke tujuh, terlihat pada gambar 1.

```

Initial starting points (random):

Cluster 0: 60,20,100,20,25,75
Cluster 1: 60,60,60,40,25,45
Cluster 2: 60,20,100,40,25,75
Cluster 3: 60,20,30,20,25,60

Missing values globally replaced with mean/node

Final cluster centroids:

Attribute      Full Data      Cluster#
              (129.0)      (17.0)      (24.0)      (32.0)      (56.0)
-----
salary         64.8062      48.2353      65.8333      67.5      61.785
statdmh       34.2636      24.7059      91.6667      21.25      2
kondrmh       72.8682      100      67.5      100      51.428
tanggungan   36.7442      32.9412      35      37.5      38.214
statortu     36.8217      45.9882      42.7083      26.5625      37.
arad         67.7132      37.9412      69.5833      82.3125      67.321
    
```

Gambar 1 Output Clusterer Weka

Dari nilai akhir *centroid cluster*, dilakukan kodifikasi untuk menentukan rekomendasi SL, L, DP dan KL. *Cluster* dengan jumlah *centroid* yang paling besar ditetapkan sebagai SL.

TABEL V KODIFIKASI CLUSTER
KE REKOMENDASI

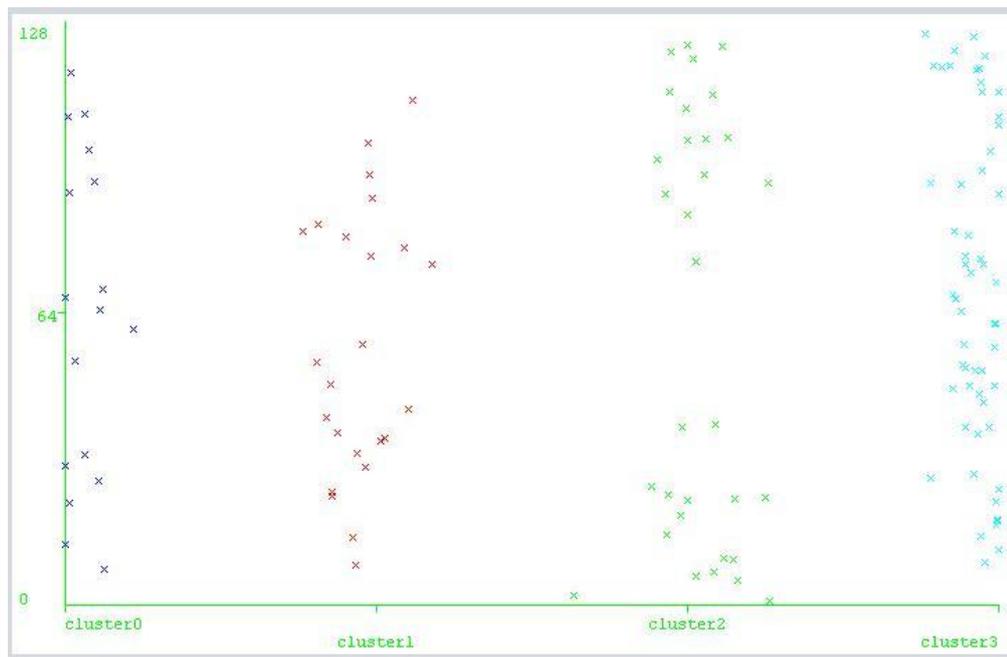
KATEGORI	C0	C1	C2	C3
PH	68,23 53	65,83 33	67,50 00	61,78 57
SR	24,70 59	91,66 67	21,25 00	20,00 00
KR	100,0 000	67,50 00	100,0 000	51,42 86
TG	32,94 12	35,00 00	37,50 00	38,21 43
SO	45,58 82	42,70 83	26,56 25	37,50 00
AK	37,94 12	69,58 33	82,81 25	67,32 14
Jumlah	309,4 118	372,2 916	335,6 25	276,2 5
Rekomendasi	DP	SL	L	KL

3. Clustering 129 data instance

TABEL VI HASIL CLUSTERING 129 DATA INSTANCE

No	NISN	Instance	PH	SR	KR	TG	SO	AK	Cluster
1	9962268578	0.0	80.0	20.0	100.0	60.0	25.0	100.0	cluster2
2	9971708275	1.0	60.0	60.0	100.0	40.0	25.0	75.0	cluster2
3	9971705880	2.0	60.0	20.0	100.0	40.0	25.0	75.0	cluster2
4	9962262774	3.0	60.0	20.0	100.0	20.0	25.0	75.0	cluster2
5	9962262686	4.0	60.0	100.0	60.0	40.0	25.0	75.0	cluster1
6	9971708298	5.0	60.0	20.0	60.0	60.0	25.0	100.0	cluster3
7	9962268596	6.0	60.0	20.0	60.0	60.0	25.0	100.0	cluster3
8	9971707526	7.0	80.0	20.0	100.0	40.0	25.0	75.0	cluster2
9	9962262693	8.0	60.0	20.0	100.0	40.0	25.0	75.0	cluster2
10	9971708272	9.0	60.0	20.0	100.0	40.0	25.0	75.0	cluster2
...									
120	9969800533	119.0	80.0	100.0	100.0	40.0	25.0	100.0	cluster1
121	9962585815	120.0	80.0	20.0	60.0	20.0	75.0	100.0	cluster3
122	9955201200	121.0	60.0	20.0	100.0	40.0	25.0	100.0	cluster2
123	9981645488	122.0	80.0	20.0	100.0	40.0	25.0	100.0	cluster2
124	9941202397	123.0	60.0	20.0	60.0	40.0	25.0	100.0	cluster3
125	9962444220	124.0	60.0	20.0	60.0	40.0	25.0	100.0	cluster3
126	9971846009	125.0	60.0	20.0	60.0	40.0	25.0	75.0	cluster3
127	9962444118	126.0	60.0	20.0	30.0	40.0	25.0	75.0	cluster3
128	9943381962	127.0	80.0	20.0	100.0	40.0	25.0	75.0	cluster2
129	9971845418	128.0	60.0	20.0	100.0	40.0	25.0	75.0	cluster2

4. Grafik *clustering*, terlihat pada gambar 2



Gambar 2 Grafik Hasil Clustering

D. Hasil Pengujian dengan Tool Weka

Dari data yang dilatih, didapatkan 4 cluster dengan hasil sebagai berikut:

1. Calon penerima beasiswa bidik misi yang berada di cluster 0 sebanyak 17 orang atau 13% dari 129 instance di rekomendasikan dengan pertimbangan untuk menerima beasiswa bidik misi.
2. Calon penerima beasiswa bidik misi yang berada di cluster 1 sebanyak 24 orang atau 19% dari 129 instance di rekomendasikan sangat layak untuk menerima beasiswa bidik misi.
3. Calon penerima beasiswa bidik misi yang berada di cluster 2 sebanyak 32 orang atau 25% dari 129 instance di rekomendasikan layak untuk menerima beasiswa bidik misi.
4. Calon penerima beasiswa bidik misi yang berada di cluster 3 sebanyak 56 orang atau 43% dari 129 instance di rekomendasikan kurang layak untuk menerima beasiswa bidik misi.

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, algoritma K-Means mampu mengelompokkan calon penerima beasiswa bidik misi ke dalam 4 cluster, yang mana cluster 0 berarti memberikan rekomendasi dengan pertimbangan, cluster 1 memberikan rekomendasi sangat layak, cluster 2 memberikan rekomendasi layak dan cluster 3 memberikan rekomendasi kurang layak. Dari hasil pengujian dihasilkan sebanyak 17 orang direkomendasi dengan pertimbangan, 24 orang direkomendasikan sangat layak, 32 orang direkomendasikan layak dan 56 orang direkomendasikan kurang layak. Hasil clustering hanya merupakan rekomendasi bagi pengambil keputusan dan bukan penentu akhir penerima beasiswa bidik misi.

REFERENSI

- [1] Asroni, Adrian, Ronald, 2015, Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang, *Jurnal Ilmiah Semesta Teknik*, 18 (1), 76 – 82
- [2] Husni, Negara, Y.D.P., Syarief, M., 2015, Clusterisasi Dokumen Web (Berita) Bahasa Indonesia Menggunakan Algoritma K-Means, *Jurnal Simantec*, 4(3), 159 – 166.
- [3] Prasetyo, Eko, 2012, Data Mining Konsep dan Aplikasi Menggunakan MATLAB, Penerbit Andi, Yogyakarta
- [4] Rismawan, Tedy, Kusumadewi, Sri, 2008, Aplikasi K-Means Untuk Pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index (Bmi) & Ukuran Kerangka, SNATI, Yogyakarta
- [5] Rohmawati, W.N., Defiyanti, S., Jajuli, M., 2015, Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa, *Jurnal Ilmiah Teknologi Informasi Terapan*, 1(2), 62-68
- [6] Suprawoto, T., (2016), Klasifikasi Data Mahasiswa Menggunakan Metode Kmeans Untuk Menunjang Pemilihan Strategi Pemasaran, *Jurnal Informatika dan Komputer (JIKO)*, 1(1), 12 – 18