

PENGELOMPOKAN SISWA DALAM MENENTUKAN PENERIMA BEASISWA BERDASARKAN PRESTASI AKADEMIK DENGAN ALGORITMA K-MEANS

Zulham¹, Buyung Solihin Hasugian²

Fakultas Teknik dan Ilmu Komputer

Universitas Dharmawangsa

Email : zulham@dharmawangsa.ac.id¹, buyung@dharmawangsa.ac.id²

RINGKASAN - Beasiswa merupakan bantuan yang diberikan kepada seseorang dengan tujuan dipergunakan untuk menjalani Pendidikan sehingga dengan begitu pendidikan orang tersebut bisa selesai sesuai waktu yang telah ditentukan. Pemberian beasiswa biasanya dilaksanakan berbagai Lembaga baik negeri maupun swasta berdasarkan prestasi yang diraih oleh orang tersebut. Untuk mempermudah dalam proses penentuan penerima beasiswa yang berprestasi maka dibuatlah sebuah program Data mining. Data mining merupakan sistem yang berguna dalam membantu user dalam menentukan sebuah penentuan berdasarkan data dengan proses yang sistematis. Data mining biasanya digunakan untuk menentukan suatu hal yang memiliki banyak data. Dalam penyelesaian masalah terkait penentuan penerima beasiswa siswa berprestasi, metode yang digunakan adalah metode K-Means. Metode K-Means ini juga memiliki perhitungan yang mudah dipahami.

Kata Kunci: *Data Mining, K-Means, Algoritma, Beasiswa.*

PENDAHULUAN

Beasiswa merupakan bantuan yang diberikan kepada seseorang dengan tujuan dipergunakan untuk menjalani Pendidikan sehingga dengan begitu pendidikan orang tersebut bisa selesai sesuai waktu yang telah ditentukan. Pembagian beasiswa dilakukan oleh beberapa lembaga untuk membantu seseorang yang kurang mampu ataupun berprestasi selama menempuh studinya. Beasiswa ini diberikan bertujuan agar meringankan beban yang dijani oleh orang tersebut. Pemberian beasiswa dapat dikategorikan pada pemberian berdasarkan prestasi ataupun pemberian karena keadaan ekonomi.

SMK Negeri 1 Lubuk Pakam merupakan sekolah yang berada di Sumatera Utara yang mana sekolah tersebut selalu memberikan beasiswa

kepada calon siswanya baik yang berprestasi maupun karena alasan ekonomi. Pada prosesnya sekolah tersebut mengalami kesulitan dalam menentukan calon penerima beasiswa setiap tahunnya. Penelitian ini bertujuan untuk memberikan solusi agar kedepannya SMK Negeri 1 Lubuk Pakam bisa menentukan dan mengelompokkan calon siswa mana saja yang akan diberikan beasiswa sesuai dengan kriteria yang telah diinginkan.

Dengan teknik data mining dimana menggunakan algoritma K-Means dianggap paling sesuai untuk mengatasi permasalahan tersebut. Adapun algoritma K-Means merupakan algoritma yang dapat membuat *cluster* sesuai dengan kategori yang diinginkan khususnya jika terdapat kemiripan data yang diperoleh.

METODE PENELITIAN

Metode penelitian merupakan cara atau langkah yang harus dilakukan untuk mengumpulkan suatu informasi yang berisikan data yang kita peroleh dari seorang pakar atau ahli dalam bidangnya.

1. Data Collecting

Dalam teknik pengumpulan data terdapat beberapa yang dilakukan di antaranya yaitu sebagai berikut:

- a. Observasi adalah teknik pengumpulan data dengan melakukan tinjauan langsung ditempat dimana kita melakukan penelitian.
- b. Wawancara merupakan cara dimana kita dapat memperoleh sebuah informasi secara rinci, langsung, mendalam, tidak terstruktur, dan individu untuk menghasilkan sebuah informasi yang akurat.

Tabel 1
Data Primer

No	Nama Siswa	Nilai Rata-rata	Jumlah nilai
1	Adam Syahputra	90,35	2349
2	Agnes Rotua Uliana Purba	85,58	2225
3	Alfa Rozi	90,31	2348
4	Angga Dwifangga	84,50	2197
5	Annika Tobing	86,73	2255
6	Cantika Wike Habsari	89,46	2326
7	Cindy Ervina	85,58	2225
8	Dewi Wahyuni Friend Siska Pinem	85,12	2213
9	Dikky Mulanta Ginting	82,38	2142
10	Eva Shelonia Sihombing	86,12	2239
11	Fajar Mustika Dewi	84,81	2205
12	Fisca Julianti Sibarani	84,77	2204
13	Friska Olivia Purba	86,81	2257
14	Ibnu Hadi	87,23	2268
15	Irsad Fahrudin	85,58	2225
16	Khaironi	85,62	2226
17	Mhd. Duta Riansyah	85,81	2231
18	Muhammad Fachru Haemy	85,50	2223
19	Muhammad Fahryza	85,58	2225
20	Muhammad Ridho Siregar	90,31	2348
21	Mutiara Niken Ayu	84,50	2197
22	Nico Putra Damani	86,73	2255
23	Nita Wahyu Kanya	82,38	2142
24	Nur Zannah	85,58	2225
25	Nurhalizah Siregar	85,12	2213
26	Putri Ananda	82,38	2142
27	Rena Randu	86,12	2239
28	Rio Frans Nainggolan	84,81	2205
29	Serli Anggraini	85,58	2229
30	Sitria Ningsih	86,81	2257
31	Suherawati	87,23	2268

2. Studi Literatur

Dalam studi literatur, penulis mendapatkan referensi dari berbagai jurnal yang ada baik nasional maupun internasional.

HASIL PENELITIAN DAN PEMBAHASAN

1. Metode Pengembangan Sistem

Dalam merancang sistem ini khususnya *software* penulis mengadopsi beberapa metode diantaranya algoritma *waterfall* atau algoritma air terjun.

2. Algoritma Sistem

Algoritma sistem merupakan salah satu urutan maupun langkah-langkah cara pembuatan sistem sehingga memberikan intruksi luaran yang diinginkan berdasarkan ide atau masukan yang diberikan.

3. Penerapan Metode K-Means

Berikut ini adalah perhitungan metode K-Means untuk mendapatkan penerima beasiswa. Tahap ini dilakukan penerapan algoritma K-means dengan rumus:

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

Penerapan jumlah cluster (K) yaitu 3 *cluster*, Setelah menetapkan jumlah *cluster*, Tentukan titik pusat awal *cluster* (Centroid), Berikut ini titik Centroid yang telah dipilih:

Tabel 2
Centroid Awal

Centroid	Inisialisasi	Nilai Rata-rata	Jumlah Nilai	Jumlah Nilai Praktek	Keadaan
Centroid 1	S1	90,35	2349	2349	Tinggi
Centroid 2	S29	85,58	2225	2225	Sedang
Centroid 3	S23	82,38	2142	2142	Rendah

Hitung jarak data ke *Centroid* menggunakan rumus *Euclidean*, data tersebut akan ditetapkan sebagai anggota dari *cluster* terdekatnya. Menghitung *distance* (jarak) antara variabel dari setiap sampel data dengan *Centroid* yaitu:

Dengan Centroid S1 (90,35;2349;2349) Jarak antara S1 dengan titik S1

$$\begin{aligned} &= \sqrt{(\sum_{i=1}^n (X_i - Y_i)^2)} \\ &= \sqrt{((90,35-90,35)^2 + (2349-2349)^2 + (2349-2349)^2)} \\ &= 0,00 \end{aligned}$$

Dengan Centroid S29 (85,58;2225;2225) Jarak antara S1 dengan titik S29

$$\begin{aligned} &= \sqrt{(\sum_{i=1}^n (X_i - Y_i)^2)} \\ &= \sqrt{((90,35-85,58)^2 + (2349-2225)^2 + (2349-2225)^2)} \\ &= 15.398,75 \end{aligned}$$

Dengan Centroid S23 (82,38;2142;2142) Jarak antara S1 dengan titik S23

$$\begin{aligned} &= \sqrt{(\sum_{i=1}^n (X_i - Y_i)^2)} \\ &= \sqrt{((90,35-82,38)^2 + (2349-2142)^2 + (2349-2142)^2)} \\ &= 42.912,52 \end{aligned}$$

Tabel 3
Literasi 1

No	Nama	C1	C2	C3	Cluster	WCV
1	S1	0,00	15398,72	42912,46	C1	0,00
2	S2	16925,04	36,05	5937,80	C2	36,05
3	S3	8661,82	962,41	13015,27	C2	962,41
4	S4	32088,45	3029,49	785,17	C3	785,17
5	S5	6250,26	2027,98	16408,28	C2	2027,98
6	S6	15648,15	1,00	6733,98	C2	1,00
7	S7	13709,28	49,07	8112,01	C2	49,07
8	S8	15648,15	1,00	6733,98	C2	1,00
9	S9	39659,64	5633,34	64,10	C3	64,10
10	S10	17715,21	81,12	5484,13	C2	81,12
11	S11	9038,38	842,24	12562,60	C2	842,24
12	S12	17715,21	81,12	5484,13	C2	81,12
13	S13	22233,89	625,93	3369,00	C2	625,93
14	S14	13244,60	81,12	8476,55	C2	81,12
15	S15	15899,52	4,01	6570,73	C2	4,01
16	S16	15398,78	0,00	6899,22	C2	0,00

Dari tabel 3 di dapat keanggotaan sebagai berikut:

- C1 = {S1,S17,S34,S50}
- C2 = {S2,S3,S5, S6,S7,S8, S9,S10,S11, S12,S13,S14, S15,S16,S18,S19,S20,S21,S22,S24,S25,S26,S27,S28,S30,S31,S32,S33,S35 ,S36,S38,S39,S40,S41,S42,S43,S44,S45,S46,S47,S48,S49}
- C3 = {S4,S29,S23,S37}

Keterangan:

BCV: *Between Cluster Variation*

WCV: *Within Cluster Variation*

Pada langkah ini dihitung pula rasio besaran BCV

dan WCV:

Karena Centroid $m_1 = (90,35;2349;2349)$, $m_2 = (85,58;2225;2225)$,

$m_3 = (82,38;2142;2142)$

$$d(m_1, m_2) = \sqrt{((90,35-85,58)^2 + (2349-2225)^2 + (2349-2225)^2)} = 15.398,75$$

$$d(m_1, m_3) = \sqrt{((90,35-82,38)^2 + (2349-2142)^2 + (2349-2142)^2)} = 42.912,52$$

$$d(m_2, m_3) = \sqrt{((85,58-82,38)^2 + (2225-2142)^2 + (2349-2142)^2)} = 6.899,24$$

$$BCV = d(m1,m2) + d(m1,m3) + d(m2,m3) = 65.210,51$$

WCV = adalah memilih jumlah jarak terkecil di pangkat dua antara data dengan *centroid* pada masing-masing *cluster* = 6.899,24

$$\text{Sehingga Besar Rasio} = BCV/WCV = 65.210,51/6.899,24 = 9,45$$

Lakukan pembaruan *Centroid* dari hasil *cluster* seperti berikut:

$$- C1 = \text{rata-rata } \{S1,S17,S34,S50\} = (89,90; 2337)$$

$$- C2 = \text{rata-rata } \{S2,S3,S5, S6,S7,S8, S9,S10,S11, S12,S13,S14,S15,S16,S18,S19,S20,S21,S22,S24,S25,S26,S27,S28,S30,S31, S32,S33,S35,S36,S38,S39,S40,S41,S42,S43,S44,S45,S46,S47,S48,S49\} = (85,84;2232)$$

$$- C3 = \text{rata-rata } \{S4,S29,S23,S37\} = (83,80;2179)$$

Nilai *Centroid* akan berubah dari nilai *centroid* sebelumnya, maka selanjutnya algoritma dilanjutkan ketahap selanjutnya.

Kemudian hitunglah literasi 2 sama seperti pada literasi 1 sehingga mendapatkan nilai rasio yang sama dengan nilai rasio yang sebelumnya.

Dibawah ini adalah perhitungannya literasi kedua. Rumusnya adalah

$$\text{Euclidian} = \sqrt{(\sum_{i=1}^n (X_i - Y_i)^2)}$$

Objek V1

Dengan *Centroid* M1 (89,90; 2337;2337)

- Jarak antara S1 dengan titik M1

$$= \sqrt{(\sum_{i=1}^n (X_i - Y_i)^2)}$$

$$= \sqrt{((90,35-89,90)^2 + (2349-2337)^2 + (2349-2337)^2)} = 132,45$$

Dengan *Centroid* M2 (85,84; 2232;2232)

- Jarak antara S1 dengan titik M2

$$= \sqrt{(\sum_{i=1}^n (X_i - Y_i)^2)}$$

$$= \sqrt{((90,35-85,84)^2 + (2349-2232)^2 + (2349-2232)^2)} = 13.700,88$$

Dengan *Centroid* M3 (83,80; 2179;2179)

- Jarak antara S1 dengan titik M3

$$= \sqrt{(\sum_{i=1}^n (X_i - Y_i)^2)}$$

$$= \sqrt{((90,35-83,80)^2 + (2349-2179)^2 + (2349-2179)^2)}$$

$$= 28,942,75$$

Lakukan proses perhitungan yang sama sampai dengan objek ke 50 adapun hasil dari perhitungan literasi 2 dapat dilihat pada tabel dibawah ini:

Tabel 4
Literasi 2

No	Nama	C1	C2	C3	Cluster	WCV
1	S1	132,45	13700,88	28942,75	C1	132,45
2	S2	14063,02	170,18	1602,37	C2	170,18
3	S3	6652,08	575,14	5937,77	C2	575,14
4	S4	28097,75	3854,12	81,12	C3	81,12
5	S5	4562,99	1443,42	8293,25	C2	1443,42
6	S6	12901,31	64,67	2028,00	C2	64,67
7	S7	11146,71	0,00	2813,16	C2	0,00
8	S8	12901,31	64,67	2028,00	C2	64,67
9	S9	35208,26	6739,81	842,24	C3	842,24
10	S10	14784,09	257,52	1371,03	C2	257,52
11	S11	6982,56	483,14	5633,32	C2	483,14
12	S12	14784,09	257,52	1371,03	C2	257,52
13	S13	18934,22	1027,81	441,65	C3	441,65
14	S14	10728,10	3,86	3029,47	C2	3,86
15	S15	13129,64	81,76	1938,86	C2	81,76
16	S16	12674,97	49,57	2119,13	C2	49,57
17	S17	110,41	13467,61	28603,25	C1	110,41
18	S18	19769,45	1229,32	324,48	C3	324,48
19	S19	6816,32	528,14	5784,54	C2	528,14
20	S20	132,45	8842,35	21640,97	C1	132,45
21	S21	12674,97	49,57	2119,13	C2	49,57
22	S22	15523,18	362,89	1157,71	C2	362,89
23	S23	38276,79	8118,42	1371,03	C3	1371,03
24	S24	9716,60	48,57	3605,33	C2	48,57

Dari tabel 4 didapat keanggotaan sebagai berikut:

- C1 = {S1,S17, S20,S34,S50}

- C2 =

{S2,S3,S5,S6,S7,S8,S10,S11,S12,S14,S15,S16,S19,S21,S22,S24,S27,S28,

S30,S31,S32,S33,S35,S36,S38,S39,S40,S41,S43,S44,S45,S47,S48,S49}

- C3 = {S4,S9,S13,S18,S23,S25,S26,37,42,46}

Tabel 5
Literasi 3

No	Nama	C1	C2	C3	Cluster	WCV
1	S1	132,45	13700,88	28942,73	C1	132,45
2	S2	14063,02	170,18	1602,37	C2	170,18
3	S3	6652,08	575,14	5937,77	C2	575,14
4	S4	28097,75	3854,12	81,12	C3	81,12
5	S5	4562,99	1443,42	8293,29	C2	1443,42
6	S6	12901,31	64,67	2028,00	C2	64,67
7	S7	11146,71	0,00	2813,16	C2	0,00
8	S8	12901,31	64,67	2028,00	C2	64,67
9	S9	35208,26	6739,81	842,24	C3	842,24
10	S10	14784,09	257,52	1371,03	C2	257,52
11	S11	6982,56	483,14	5633,32	C2	483,14
12	S12	14784,09	257,52	1371,03	C2	257,52

Dari tabel 5 di dapat keanggotaan sebagai berikut:

- C1 = {S1,S17, S20, S34, S50}
- C2 = {S2,S3,S5,S6,S7,S8,S10,S11,S12,S14,S15,S16, S19,S21,S22,S24,S27,S28,S30, S31,S32,S33,S35,S36,S38.S39.S40.S41,S43,S44,S45,S47,S48,S49}
- C3 = {S4,S9,S13,S18,S23,S25,S26,S37,S42,S46}

Setelah dilakukan sebanyak 3 iterasi dan nilai *Centroid* nya sama dari nilai *centroid* sebelumnya yaitu *centroid* 2, maka hasil akhirnya adalah sebagai berikut:

Tabel 6
Pengelompokan Hasil Cluster

CLUSTER	Nama
CLUSTER 1	S1,S17, S20,S34,S50
CLUSTER 2	S2,S3,S5, S6,S7,S8, S10,S11,S12, S14,S15,S16, , S19,S21, S22, S24, S27, S28,S30,S31,S32, S33,S35, S36,S38,S39.S40,S41, S43, S44, S45,S47,S48,S49

Tabel 7
Hasil Kategori Tinggi

No	Nama	Cluster	Keterangan
1	Adam Syahputra	1	Tinggi
2	Mhd. Duta Riansyah	1	Tinggi
3	Muhammad Ridho Siregar	1	Tinggi
4	Billy Hendrawan	1	Tinggi
5	Udin Darwansyah	1	Tinggi

Tabel 8
Hasil Kategori Sedang

No	Nama	Cluster	Keterangan
1	Agnes Rotua Ulana Purba	2	Sedang
2	Alfa Rozi	2	Sedang
3	Annika Tobing	2	Sedang
4	Cantika Wika Habsari	2	Sedang
5	Cindy Ervina	2	Sedang
6	Dewi Wahyuni Friend Siska Pinam	2	Sedang
7	Eva Shelonia Sihombing	2	Sedang
8	Fajar Mustika Dewi	2	Sedang
9	Firsca Juliam Sibarami	2	Sedang
10	Ibnu Hadi	2	Sedang
11	Irsad Fahrudin	2	Sedang

Tabel 9
Hasil Kategori Rendah

No	Nama	Cluster	Keterangan
1	Angga Dwifangga	3	Rendah
2	Dikky Mulanta Ginting	3	Rendah
3	Friska Olivia Purba	3	Rendah
4	Muhammad Fachu Hasmy	3	Rendah
5	Nita Wahyu Kanya	3	Rendah
6	Nurhalizah Siregar	3	Rendah
7	Putri Ananda	3	Rendah
8	Angga Dwifangga	3	Rendah
9	Budi Santoro	3	Rendah
10	Cahaya Setiani	3	Rendah

Maka dari hasil *clustering* diatas didapatkan kesimpulan yaitu yang berhak untuk mendapatkan beasiswa adalah siswa yang memiliki *clustering*

tertinggi yaitu Adam Syahputra, Mhd Duta Riansyah, Muhammad Ridho Siregar, Billy Hendrawan, dan Udin Darwansyah.

SIMPULAN

Dari hasil penelitian ini maka dapat disimpulkan sebagai berikut:

1. Untuk mendesain sistem pada penelitian ini, didapatkan bahwasannya system yang dirancang sesuai dengan kebutuhan dalam menentukan kelayakan penerima beasiswa.
2. Data yang digunakan dalam sistem pakar ini adalah data yang didapat dari hasil observasi di Sekolah yang bersangkutan.
3. Untuk membangun system yang baik, digunakan sebuah metode yaitu metode K-Means dalam penyelesaian masalah dalam menentukan kelayakan penerima beasiswa.
4. Sistem yang dibangun sudah layak digunakan pada kantor sekolah yang bersangkutan.

DAFTAR PUSTAKA

- Asroni and R. Adrian. *Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang*. J. Ilm. Semesta Tek., vol. 18, no. 1, pp. 76–82, 2015.
- G. Gustientiedina, M. H. Adiya, and Y. Desnelita. *Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan*. J. Nas. Teknol. dan Sist. Inf., vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- H. Irsyad and M. R. Pribadi. *Implementasi Text Mining Dalam Pengelompokan Data Tweet Pertanian Indonesia Dengan K-Means*. KURAWAL J. Teknol. Inf. dan Ind., vol. 3, no. 2, pp. 164–172, 2020, [Online]. Available: <https://t.co/FXtzMcbdHp>.
- H. Priyatman, F. Sajid, and D. Haldivany. *Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa*. J. Edukasi dan Penelit. Inform., vol. 5, no. 1, p. 62, 2019, doi: 10.26418/jp.v5i1.29611.

- A. H. Nasyuha, Zulham, I. Jang Cik, M. Amin, S. Candra Setia, and D. Siregar. *An Integrated Multi Criteria Decision Making Method for Fashion Selection*. J. Phys. Conf. Ser., vol. 1424, no. 1, 2019, doi: 10.1088/1742-6596/1424/1/012030.
- Sugiono, S. Nurdiani, S. Linawati, R. A. Safitri, and E. P. Saputra. “*Pengelompokan Perilaku Mahasiswa Pada Perkuliahan E-Learning dengan K-Means Clustering*.” J. Kaji. Ilm., vol. 19, no. 2, pp. 126–133, 2019.
- D. Triyansyah and D. Fitriana. *Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing*. J. Telekomun. dan Komput., vol. 8, no. 3, p. 163, 2018, doi: 10.22441/incomtech.v8i3.4174.
- R. S. and N. Tes. *Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru*. J. Lentera Ict, vol. 3, no. 1, pp. 76–92, 2016.
- R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah. *Analisis K-Means Clustering pada Data Sepeda Motor*. INFORMAL Informatics J., vol. 5, no. 1, p. 10, 2020, doi: 10.19184/isj.v5i1.17071.
- S. Handoko, F. Fauziah, and E. T. E. Handayani. *Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering*. J. Ilm. Teknol. dan Rekayasa, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- R. A. Indraputra and R. Fitriana. *K-Means Clustering Data COVID-19*. J. Tek. Ind., vol. 10, no. 3, p. 3, 2020.
- H. Sy, Rismayani, and A. Syam. *Data Mining Menggunakan Algoritma K-Means Pengelompokan Penyebaran Diare Di Kota Makassar*. SISITI Semin. Ilm. Sist. Inf. dan Teknol. Inf., vol. 8, no. 1, pp. 73–82, 2019.