



PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN KELAS KELOMPOK BIMBINGAN BELAJAR TAMBAHAN

Silvia Ningsih^{1(*)}

¹ITB Ahmad Dahlan, Jakarta

Abstract

Additional learning is a learning program that is carried out outside the school's intracurricular program. This means that additional learning activities are carried out after the regular learning program at school has ended. The learning mentor comes from the subject teacher at the school concerned. It is commonly called the afternoon additional learning program. The afternoon additional learning program has its own planning schedule. Additional learning is carried out by students to improve their understanding and deepening of subject material. This goal is related to the preparation of a student to face exams at school, both midterm exams, semester final exams and national final exams. In determining the class of this additional study group using the K-Means Clustering algorithm, the number of data samples will be used is 26 students majoring in science.

Kata Kunci: Data Meaning, K-Means, Clustering, Kelompok Belajar

Juli - Desember 2022, Vol 3 (2) : hlm 73-82
©2022 Institut Teknologi dan Bisnis Ahmad Dahlan.
All rights reserved.

(*) Korespondensi: silvianingsih417@gmail.com (Silvia Ningsih)

PENDAHULUAN

SMA Negeri 1 Pangkalan Kerinci dalam menghadapi Ujian Nasional (UN) akan mengadakan suatu kelompok bimbingan belajar. Hal ini bertujuan untuk memantapkan materi – materi mata pelajaran yang akan di ujikan. Dalam perkembangannya sistem pembagian kelompok bimbingan belajar yang ada saat ini dirasa kurang efektif dikarenakan penekanan pembelajaran yang diberikan untuk semua kelompok (kelas). Sedangkan tingkat kemampuan siswa dalam setiap mata pelajaran tidak sama. Sebagian siswa hanya tanggap dalam beberapa mata pelajaran, dan ada sebagian yang tanggap dalam semua mata pelajaran. Masalah tersebut mengakibatkan kurangnya minat siswa dalam mengikuti proses bimbingan belajar. Belum adanya system pembagian kelompok bimbingan belajar, menyebabkan pembagian kelompok bimbingan belajar tidak sesuai harapan. Permasalahan yang timbul adalah pihak sekolah mengalami kesulitan dalam menentukan kelompok bimbingan dalam menentukan tingkat kemampuan siswa. Ini dikarenakan tiap-tiap siswa mempunyai kemampuan memahami mata pelajaran yang berbeda-beda. Pengelompokan siswa pada kelompok yang tepat akan mampu mengimprovisasi hasil pembelajaran menjadi lebih baik (Henry, 2013).

Knowledge discovery in Database (KDD) didefinisikan sebagai ekstraksi informasi potensial, implisit dan tidak dikenal dari sekumpulan data. Proses *knowledge discovery* melibatkan hasil dari proses *data mining* (proses mengekstrak kecenderungan pola suatu data), kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami (Ronald, 2015). Proses *knowledge discovery in database (KDD)* secara garis besar terdiri dari *Data Selection*, *Pre-processing/Cleaning*, *Transformation*, *Data mining*, dan *Interpretation/Evaluation*.

Algoritma *K-Means* merupakan algoritma non hirarki yang berasal dari metode data *clustering*. Menurut Eko Prasetyo (2012) mengatakan bahwa metode *K-Means* ini mempartisi data kedalamkelompok sehingga data berkarakteristik sama dimasukan kedalam sat kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain. Adapun tujuan dari pengelompokan data ini adalah untuk meminimalkan fungsi objektif yang diset dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi didalam suatu kelompok dan memaksimalkan variasi antar kelompok.

METODE PENELITIAN

Menurut Eko Prasetyo (2012) mengatakan bahwa metode *K-Means* ini mempartisi data kedalam kelompok sehingga data berkarakteristik sama dimasukan kedalam sat kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain.

Langkah-langkah melakukan clustering dengan metode *K-Means* adalah sebagai berikut:

1. Tentukan nilai k sebagai jumlah klaster yang ingin dibentuk.
2. Inisialisasi k pusat *cluster* ini bisa dilakukan dengan berbagai cara, namun yang paling seringdilakukan adalah dengan cara random yang di ambil dari data yang ada.
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* :

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

Dimana :

d : titik dokumen

x_i : data kriteria

μ_j : *centroid* pada cluster ke- j

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

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in s_j} x_j$$

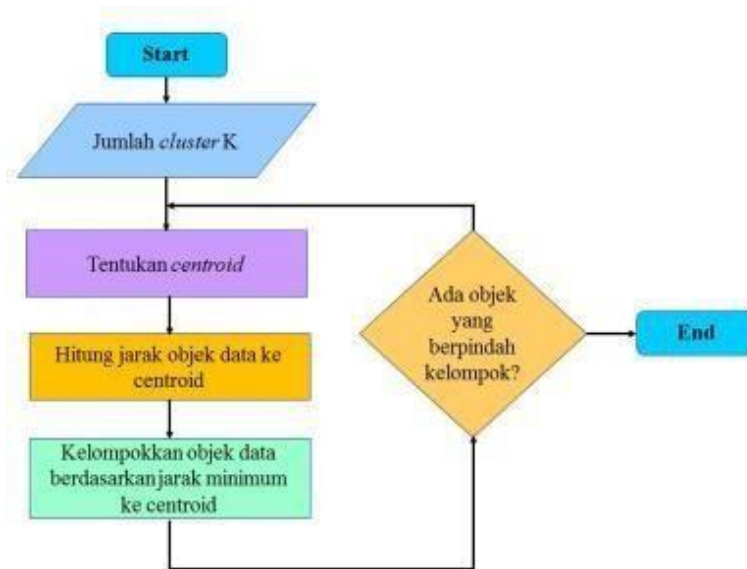
Dimana:

$\mu_j(t+1)$: *centroid* baru pada iterasi ke $(t+1)$ N_{sj} : banyak data pada cluster s_j ,

1. Melakukan perulangan dari langkah 2 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah.

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

Gambar 1. Flowchart Algoritma K-Means Clustering



HASIL DAN PEMBAHASAN

1. Analisa Sistem

Tahap analisa sistem merupakan tahap yang paling penting karena pada tahap ini akan dilakukan evaluasi kinerja, identifikasi terhadap masalah dan cara kerja sistem. Analisa system bertujuan untuk mengetahui semua pengetahuan yang berkaitan dengan data nilai mata pelajaran Ujian Nasional.

SMA Negeri 1 Pangkalan Kerinci dalam menghadapi Ujian Nasional (UN) akan mengadakan suatu belajar tambahan, hal ini bertujuan untuk memantapkan materi-materi mata pelajaran yang akan di ujikan, maka pihak sekolah harus memberikan belajar tambahan kepada siswa yang kemampuannya rendah karena syarat untuk lulus dalam ujian nasional ini semua mata pelajaran yang diujikan harus memenuhi nilai standar lulus yang telah ditentukan jika salah satu mata pelajaran ujian nasional tidak lulus maka dinyatakan siswa tersebut tidak lulus dalam ujian nasional.

2. Pengumpulan Data

Pada proses pengumpulan data ada enam parameter yang akan digunakan dalam pengolahan data yaitu nilai mata pelajaran Ujian Nasional SMA Jurusan IPA. Ujian Nasional adalah sistem standar pendidikan dasar dan menengah yang dilakukan secara nasional sesuai keputusan menteri pendidikan dan berdasarkan UU nomor 20 tahun 2003. Pada penelitian ini data yang dikumpulkan yaitu data nilai Ujian Nasional SMA jurusan IPA yang meliputi data nilai bahasa indonesia, bahasa inggris, matematika, fisika, biologi, dan kimia Data tersebut diambil dari SMA Negeri 1 Pangkalan Kerinci, dengan jumlah keseluruhan data 166 record.

Tabel 1. Jumlah Keseluruhan Data

NO	NAMA	BI	BING	MTK	FIS	BIO	KIM
1	ANGGI MALDO JEFRA	89	88	80	81	83	80
2	DIO HARVANDY	86	80	78	87	81	82
3	ELDERA RIZKI NURAINI	85	83	80	77	70	80
4	ERGA SHINTA ANGELA	80	80	82	80	75	81
5	ERIK KRISMON EFENDI	84	79	80	84	82	85
6	EXSIS YUPITA SARI	86	78	80	78	83	80
7	FEBI KEMBARA PUTRA	88	87	85	87	81	85
8	IBNU MUHAMMAD IKHSAN	80	80	80	80	81	78
9	ICE MAIDIKA SARI	84	87	80	81	79	80
10	IFNI SUKANTI	84	80	80	81	82	80
11	ILHAM RINALDO PUTRA	86	84	80	83	75	80
12	ILHAM YULIANDRA	86	84	85	80	83	81
13	LAILATUL WARAHMAH	84	80	84	81	79	81
14	MEIDI KRISTIAN	83	89	80	80	83	79
15	MELLA ELVIANA	86	89	87	82	85	78
16	MITA TRISNAWATI	89	85	80	78	82	81

17	NURANI AULIA PUTRI	7 5	85	79	80	82	79
18	PANZHER PUTRA	9 0	84	98	80	95	90
19	RANDI ZONNANDA PUTRA	9 1	83	78	87	85	81
20	REZA SURYA PUTRI	7 5	89	80	80	85	90
21	RICE PRIANI	9 0	82	80	81	81	79
22	SATRIA BUANA	8 0	79	79	83	81	80
23	SHERLY WULANDARI	7 5	87	80	79	83	79
24	SILVIA	8 0	90	70	80	78	91
25	SISKA	9 0	86	80	86	84	80
26	SUCI NUR INDAH SARI	9 5	90	92	90	96	90

Data diatas merupakan data sampel nilai UAS Semester IV SMAN 1 Pangkalan Keinci jurusan IPA berjumlah 26 record, data tersebut akan di kluster menjadi tiga kelompok yaitu kemampuan siswa pintar, siswa sedang dan siswa kurang pintar.

Proses Clustering Menggunakan Algoritma *K-Means*

Data yang sudah dijadikan sampel akan dilakukan pengolahan data dengan proses clustering dengan menggunakan algoritma K-Means, sehingga didapatkan hasil pengelompokan data yang diinginkan. Adapun langkah dalam cluster dengan menggunakan algoritma *K-Means*, yaitu :

1. Menentukan Jumlah Cluster

Menentukan jumlah cluster yang digunakan pada data nilai ujian SMAN 1 Pangkalan Kerinci, sebanyak tiga cluster, diantaranya pintar, sedang, dan kurang pintar, berdasarkan mata pelajaran ujian nasional jurusan IPA.

2. Menentukan Centroid

Menentukan pusat awal cluster (centroid) ditentukan secara random atau acak yang diambil dari data yang ada. Nilai cluster 0 diambil dari baris ke-26, nilai cluster1 pada baris ke-10, nilai cluster2 pada baris ke-20.

3. Menghitung jarak

Menghitung jarak antara titik *centroid* dengan titik tiap objek dengan menggunakan *Euclidian Distance*. Adapun penghitungan *centroid* awal secara manual.

Tabel 2. Perhitungan Jarak dan Pengelompokan Data Iterasi Ke-1

NO	NIS	DC0	DC1	DC2	C0	C1	C2
1	14137	23.11	9.49	17.38	0	1	0
2	14138	25.98	7.00	18.30	0	1	0
3	14139	35.19	13.04	21.68	0	1	0
4	14140	32.36	8.43	17.06	0	1	0
5	14141	25.36	5.92	15.20	0	1	0
6	14142	27.96	4.24	18.71	0	1	0

7	14143	19.13	12.33	16.97	0	1	0
8	14144	30.63	4.69	16.31	0	1	0
9	14145	27.28	7.62	14.90	0	1	0
10	14146	27.24	0.00	16.49	0	1	0
11	14147	29.17	8.54	18.84	0	1	0
12	14148	22.72	6.93	16.00	0	1	0
13	14149	27.13	5.10	17.20	0	1	0
14	14150	26.06	9.22	13.75	0	1	0
15	14151	20.88	12.17	17.83	0	1	0
16	14152	25.02	7.75	17.49	0	1	0
17	14153	31.80	10.44	12.12	0	1	0
18	14154	14.07	25.42	25.96	1	0	0
19	14155	21.73	10.39	20.64	0	1	0
20	14156	27.68	16.49	0.00	0	0	1
21	14157	25.69	6.48	20.30	0	1	0

Dari hasil perhitungan di atas didapatkan pusat *cluster* baru seperti Tabel sebagai berikut :

Tabel 3. Centroid Baru Iterasi Ke-2

CLUSTER	BI	BING	MTK	FIS	BIO	KIM
C0	92.50	87.00	95.00	85.00	95.50	90.00
C1	85.05	83.24	80.81	81.76	80.81	80.48
C2	76.67	88.67	76.67	79.67	82.00	86.67

- Iterasi selanjutnya melakukan perhitungan lagi dengan menggunakan titik centroid yang baru Iterasi Ke-1.

Tabel 4. Perhitungan Jarak dan Pengelompokan Data Iterasi Ke-2

NIS	DC0	DC1	DC2	C0	C1	C2
14137	22.59	6.67	14.52	0	1	0
14138	25.66	7.01	15.51	0	1	0
14139	33.34	11.85	17.56	0	1	0
14140	30.01	8.63	14.00	0	1	0
14141	23.86	6.83	13.41	0	1	0
14142	25.56	6.94	16.13	0	1	0
14143	18.96	9.40	16.06	0	1	0
14144	28.45	6.78	13.18	0	1	0
14145	26.18	4.47	11.08	0	1	0
14146	25.39	3.81	13.65	0	1	0

14147	28.29	6.14	14.98	0	1	0
14148	20.33	5.22	14.54	0	1	0
14149	24.73	5.09	15.02	0	1	0
14150	24.93	6.94	10.55	0	1	0
14151	19.33	9.81	16.84	0	1	0
14152	23.53	5.93	14.54	0	1	0
14153	29.91	8.98	10.68	0	1	0
14154	7.04	24.81	28.89	1	0	0
14155	22.42	9.41	18.28	0	1	0
14156	25.89	15.69	5.85	0	0	1
14157	24.57	5.43	17.17	0	1	0
14158	28.12	6.97	12.91	0	1	0

Dari *centroid* baru iterasi ke-2, dilakukan perhitungan kembali, sehingga di dapatkan hasil Tabel Perhitungan Jarak dan Pengelompokan Data Iterasi Ke-1 dimana C0 memiliki 2 anggota, C1 memiliki 22 anggota, C2 memiliki 2 anggota.

Berdasarkan dari perhitungan di atas bahwa pusat *cluster* baru iterasi ke-1 seperti Tabel sebagai berikut :

Tabel 5. Centroid Baru Iterasi Ke-2

CLUSTER	BI	BING	MTK	FIS	BIO	KIM
C0	92.50	87.00	95.00	85.00	95.50	90.00
C1	85.55	83.15	80.90	81.85	80.75	80.55
C2	76.25	87.75	77.25	79.75	82.00	84.75

Dari hasil *centroid* yang didapatkan pada iterasi ke-2, kemudian lakukan lagi perhitungan yang sama sampai anggota tiap *cluster* tidak ada lagi yang berubah.

Tabel 6. Perhitungan Jarak dan Pengelompokan Data Iterasi Ke-3

NO	NIS	DC0	DC1	DC2	C0	C1	C2
1	14137	22.59	6.51	13.98	0	1	0
2	14138	25.66	6.87	14.72	0	1	0
3	14139	33.34	11.85	16.76	0	1	0
4	14140	30.01	8.87	12.64	0	1	0
5	14141	23.86	6.81	12.74	0	1	0
6	14142	25.56	6.91	14.98	0	1	0

7	14143	18.96	9.17	15.88	0	1	0
8	14144	28.45	7.18	11.33	0	1	0
9	14145	26.18	4.70	10.07	0	1	0
10	14146	25.39	3.97	12.32	0	1	0
11	14147	28.29	6.04	14.10	0	1	0
12	14148	20.33	5.14	13.58	0	1	0
13	14149	24.73	5.09	13.80	0	1	0
14	14150	24.93	7.24	9.42	0	1	0
15	14151	19.33	9.81	15.85	0	1	0
16	14152	23.53	5.72	13.96	0	1	0

17	14153	29.91	6.73	11.21	0	1	0
18	14154	7.04	24.67	28.82	1	0	0
19	14155	22.42	9.11	17.79	0	1	0
20	14156	25.89	16.04	6.88	0	0	1
21	14157	24.57	5.01	16.29	0	1	0
22	14158	28.12	7.30	11.31	0	1	0
23	14159	29.06	6.66	11.94	0	1	0
24	14160	33.50	17.80	11.26	0	0	1
25	14161	21.58	7.54	16.29	0	1	0
26	14162	7.04	25.45	30.05	1	0	0

Tabel 7. Tabel Centroid baru Iterasi Ke-3

CLUSTER	BI	BING	MTK	FIS	BIO	KIM
C0	92.50	87.00	95.00	85.00	95.50	90.00
C1	85.55	83.15	80.90	81.85	80.75	80.55
C2	76.25	87.75	77.25	79.75	82.00	84.75

Karena pada Iterasi Ke-2 dan Ke-3 posisi cluster tidak berubah maka iterasi dihentikan dan hasil akhir yang diperoleh yaitu ;

1. C0 memiliki 2 anggota yang diartikan bahwa kelompok pertama adalah kategori kemampuan siswa pintar.
2. C1 memiliki 22 anggota yang diartikan bahwa kelompok kedua adalah kategori kemampuan siswa sedang.
3. C2 memiliki 2 anggota yang diartikan bahwa kelompok ketiga adalah kategori kemampuan siswa kurang pintar dan siswa inilah yang akan diberi belajar tambahan

KESIMPULAN

Berdasarkan uraian yang sudah dikemukakan pada bab-bab sebelumnya, maka dapat ditarik beberapa kesimpulan sebagai berikut :

1. Metode *Clustering* dengan algoritma *K-Means* dapat digunakan untuk mengelompokkan data siswa berdasarkan nilai ujian semester mata pelajaran Ujian Nasional, yaitu kemampuan siswa pintar, siswa sedang dan siswa kurang pintar. sehingga dapat mengetahui siswa yang mana saja yang akan diberi belajar tambahan agar dapat mencapai nilai standar kelulusan Ujian Nasional.
2. Metode *Clustering* dapat digunakan untuk membantu pihak sekolah untuk melakukan belajar tambahan terhadap siswa yang akan mengikuti Ujian Nasional.

DAFTAR PUSTAKA

- Sri Tria Siska (2016). *Analisa Penerapan Data Mining Untuk Menenyukan Kubikasi Air Terjual Berdasarkan Pengelompokan Pelanggan Menggunakan Algoritma K-Means Clustering*. Vol. 9. 1 April 2016, ISSN : 2086-4981
- Bendri Melpa Metisen dan Herlina Latipa Sari (2015). *Analisis Clustering Menggunakan Metode K- Means Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila*. Vol. 11 No. 2, September 2015, ISSN : 1858 – 2680.
- Asroni dan Ronald Adrian (2015). *Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interfase Studi Kasus Pada Jurusan Teknik Informasi UMM Magelang*. Vol. 18, No. 1, 76-82, Mei 2015.
- Nurul Romawati W, *et al* (2015). *Implementasi Algoritma K-means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa*. Vol. 01, No. 2, 30 April 2015, ISSN : 2407-3911
- Fina Nasari dan Surya Darma (2015). *Penerapan K- Means Clustering Pada Data Penerimaan Mahasiswa Baru*. ISSN : 2302-3