



Pemetaan Kompetensi Siswa Untuk Mengikuti Seleksi Lomba Kejuaraan Multimedia Menggunakan Metode K-Means

Lis Utari^{1*}, Rio Geraldy²

¹Teknik Informatika/Universitas Binaniga Indonesia
Email: lis_utari@yahoo.co.id

²Sistem Informasi/Universitas Binaniga Indonesia
Email: riogeraldy23@gmail.com

ABSTRACT

The process of mapping competent students who will take part in the competition is the selection of students who have the best scores in the vocational field, which is carried out by the academic side in order to include students who are competent in the vocational field to be included in the vocational competition. So far, the school still has difficulties in selecting competent and less competent students due to the limitations of the presentation and processing of the data carried out. To facilitate the mapping of competent students in the vocational field, this activity is computerized into software. Time accuracy and effectiveness are important factors so that the mapping process of competent students can run well and as it should. In this research, an application is made that can cluster competent students so that they can take part in Multimedia vocational competitions appropriately and accurately using the K-Means Algorithm. It applies variables such as Academic Values and Attitude Values.

Keywords: Clusterization, Competent Students, Competition; algorithm k-means.

ABSTRAK

Proses pemetaan siswa berkompeten yang akan mengikuti lomba merupakan penyeleksian siswa yang memiliki nilai terbaik dalam bidang kejuruan, yang dilakukan pihak akademik agar dapat mengikutsertakan siswa-siswa yang berkompeten dalam bidang kejuruan untuk diikutsertakan dalam lomba kejuruan. Selama ini pihak sekolah masih kesulitan dalam menyeleksi siswa yang berkompeten maupun yang kurang kompeten dikarenakan keterbatasan penyajian dan proses data yang dilakukan. Untuk memudahkan jalannya pemetaan siswa berkompeten didalam bidang kejuruan, maka kegiatan ini dikomputerisasikan kedalam bentuk *software*. Keakuratan dan keefektifitasan waktu merupakan faktor penting agar proses pemetaan siswa berkompeten dapat berjalan dengan baik dan sebagaimana mestinya. Pada penelitian ini dibuat sebuah aplikasi yang dapat mengklaster siswa berkompeten agar dapat mengikuti lomba kejuruan Multimedia dengan tepat dan akurat dengan menggunakan Algoritma K-Means. Didalamnya diterapkan variabel-variabel seperti Nilai Akademik dan Nilai Sikap.

Keywords: Klasterisasi; Siswa Berkompeten; Lomba; algoritma k-means.

A. PENDAHULUAN

1. Latar Belakang

Besarnya peluang keberhasilan siswa serta kecilnya peluang kegagalan siswa merupakan cerminan dari kualitas dunia pendidikan. Dunia pendidikan saat ini dituntut untuk memiliki

kemampuan bersaing dengan memanfaatkan semua sumber daya yang dimiliki, selain sumber daya sarana, prasarana dan manusia, sistem informasi merupakan salah satu sumber daya yang dapat digunakan untuk meningkatkan kemampuan bersaing.

Tujuan pendidikan adalah fokus utama dari perubahan yang diinginkan setelah peserta didik mengikuti pendidikan. Berbagai instansi yang berbeda biasanya akan memiliki tujuan pendidikan yang beda pula. Beberapa pendidikan bertujuan untuk menghasilkan peserta didik yang kompeten dalam keahlian tertentu, instansi lain bertujuan secara spesifik untuk melatih aspek afektif pada peserta didik.

Dalam dunia pendidikan, mengikutsertakan siswa dalam perlombaan merupakan salah satu cara pihak sekolah untuk melihat kemampuan siswa yang selama ini dipelajari ketika belajar disekolah agar siswa tersebut dapat melatih kemampuan yang dimilikinya.

Secara umum penilaian siswa yang berkompoten dapat dilihat dari nilai mata pelajaran baik teori dan praktek yang dikategorikan dalam nilai pengetahuan. Evaluasi dan penilaian terhadap siswa berkompoten dilakukan dengan memberi nilai oleh pengajar kepada semua siswa yang mengikuti pelajaran yang diajarkan dan yang diikutinya. Seiring dengan terus bertambahnya jumlah data siswa setiap tahun, maka jumlah data siswa terus meningkat, sehingga menyebabkan penumpukan data yang belum diolah dengan optimal untuk menggali informasi dan pengetahuan baru melalui pola-pola yang terbentuk dari penumpukan data tersebut.

Clustering merupakan salah satu teknik dari salah satu fungsionalitas data mining, algoritma clustering merupakan algoritma pengelompokan sejumlah data menjadi kelompok-kelompok data tertentu (cluster). Setiap cluster memiliki centroid yang merupakan suatu besaran yang dihitung dari rata-rata nilai tiap items dari suatu cluster dan juga memiliki item yang letaknya paling tengah. Sementara jarak antar cluster didefinisikan dengan menggunakan beberapa metode-metode untuk menentukan cluster mana yang berdekatan. (Kusrini,2009)

K-Means Clustering merupakan salah satu metode cluster analysis non hirarki yang berusaha untuk mempartisi objek yang ada kedalam satu atau lebih cluster atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam cluster yang lain. Metode K-Means Clustering berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain.

2. Permasalahan

Adapun masalah yang menjadi dasar pada penelitian ini adalah:

- a. Kurang akuratnya penentuan siswa yang berkompoten maupun yang kurang kompeten dalam bidang kejuruan
- b. Kurang efektifnya pemilihan siswa hanya berdasarkan rata-rata nilai produktif dikelas

B. METODE

Algoritma K-Means merupakan algoritma pengelompokkan iteratif yang melakukan partisi set data ke dalam sejumlah K cluster yang sudah ditetapkan di awal. Algoritma K-Means sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Secara historis, K-Means menjadi salah satu algoritma yang paling penting dalam bidang data mining (Wu dan Kumar, 2009).

Secara historis, bentuk esensial K-Means ditemukan oleh sejumlah peneliti dari lintas disiplin ilmu. yang paling berpengaruh adalah Lloyd (1982), Forgey (1967), Friedman dan Rubin (1967), McQueen (1967). Algoritma K-Means berkembang hingga menjadi konteks yang lebih besar sebagai algoritma hill-climbing, seperti yang disampaikan oleh Gray dan Nuhoff (1998).

K-Means dapat diterapkan pada data yang direpresentasikan dalam r-dimensi ruang tempat. K-Means mengelompokkan set data r-dimensi, $X = \{x_i | i = 1, \dots, N\}$ dimana $x_i \in \mathfrak{R}^d$ yang menyatakan data ke-i sebagai "titik data". Seperti yang dijelaskan sebelumnya bahwa K-Means mempartisi X ke dalam K cluster, Algoritma K-Means mengelompokkan semua titik data dalam X sehingga

setiap titik x_i hanya jatuh dalam satu dari K partisi. Yang perlu diperhatikan adalah titik berada dalam cluster yang mana dilakukan dengan cara memberikan setiap titik sebuah ID cluster. Titik dengan ID cluster yang sama berarti berada dalam satu cluster yang sama, sedangkan titik dengan ID cluster yang berbeda berada dalam cluster yang berbeda. Untuk menyatakan hal ini, biasanya dilakukan dengan vektor keanggotaan cluster m dengan panjang N, di mana m, bernilai ID cluster titik x_i .

Parameter yang harus dimasukkan ketika menggunakan algoritma K-Means adalah nilai K. Nilai K yang digunakan biasanya didasarkan pada informasi yang diketahui sebelumnya tentang sebenarnya berapa banyak cluster data yang muncul dalam X, berapa banyak cluster yang dibutuhkan untuk penerapannya, atau jenis cluster dicari dengan mengeksplorasi/melakukan percobaan dengan beberapa nilai K. Berapa nilai K yang dipilih tidak perlu memahami bagaimana K-Means mempartisi set data X.

Dalam K-Means, setiap cluster dari K cluster diwakili oleh titik tunggal dalam \mathcal{R}^d . Set representatif cluster dinyatakan $C = \{c_j | j = 1, \dots, K\}$. Sejumlah K representatif cluster tersebut disebut juga sebagai *cluster means* atau *cluster centroid* (atau *centroid* saja). Untuk set data dalam X dikelompokkan berdasarkan konsep kedekatan atau kemiripan. Meskipun konsep yang dimaksud untuk data – data yang berkumpul dalam satu cluster adalah data – data yang mirip, tetapi kuantitas yang digunakan untuk mengukurnya adalah ketidakmiripan (*dissimilarity*). Artinya, data-data dengan ketidakmiripan (jarak) yang kecil/dekat maka lebih besar kemungkinannya untuk bergabung dalam satu cluster.

Pada saat data sudah dihitung ketidakmiripan terhadap setiap centroid, maka selanjutnya dipilih ketidakmiripan yang paling kecil sebagai cluster yang akan diikuti sebagai relokasi data pada cluster di sebuah iterasi. Relokasi sebuah data dalam cluster yang diikuti dapat dinyatakan dengan nilai keanggotaan a yang bernilai 0 atau 1. Nilai 0 jika tidak menjadi anggota sebuah cluster dan 1 jika menjadi anggota sebuah cluster. Karena K-Means mengelompokkan secara tegas data hanya pada satu cluster, maka dari nilai a sebuah data pada semua cluster, hanya satu yang bernilai 1, sedangkan lainnya 0 seperti dinyatakan oleh persamaan berikut:

$$a_{ij} = \begin{cases} 1 & \text{arg min } \{(X_i, C_j)\} \\ & j \\ 0 & L = \text{lainnya} \end{cases}$$

$d(X_i, C_j)$ menyatakan ketidakmiripan (jarak) dari data ke- i ke cluster C_j

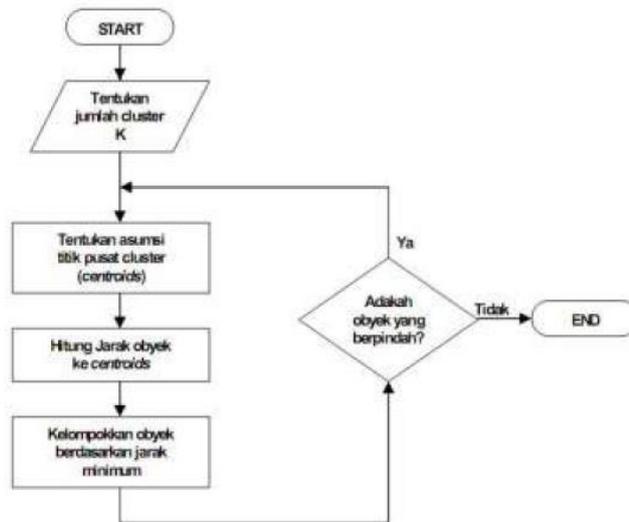
Sementara relokasi centroid untuk mendapatkan titik centroid C didapatkan dengan menghitung rata – rata setiap fitur dari semua data yang tergabung dalam setiap cluster. Rata-rata sebuah fitur dari semua data dalam sebuah cluster dinyatakan oleh persamaan berikut:

$$C_j = \frac{1}{N_k} \sum_{l=1}^{N_k} x_{jl}$$

N_k adalah jumlah data yang tergabung dalam sebuah cluster. Jika diperhatikan dari langkahnya yang selalu memilih cluster terdekat, maka sebenarnya K-Means berusaha untuk meminimalkan fungsi objektif/fungsi biaya non-negatif, seperti dinyatakan oleh persamaan berikut:

$$J = \sum_{i=1}^N \sum_{l=1}^K a_{ic} d(x_i, C_l)^2$$

Dengan kata lain, K-Means berusaha untuk meminimalkan total jarak kuadrat (*squared distance*) di antara setiap titik x_i dan representasi cluster C_j terdekat. Berikut adalah tahapan dalam Algoritma K-Means :



Gambar 1. Tahapan algoritma K-Means

C. HASIL DAN PEMBAHASAN

1. Hasil

Analisa Metode

Metode yang digunakan adalah metode klasterisasi Algoritma K- Means yang akan menghasilkan sebuah klaster yang nantinya akan digunakan untuk mengelompokkan data yang akan diprediksi.

Data yang digunakan pada penelitian ini adalah data nilai produktif siswa dan data nilai sikap siswa dari tahun akademik 2016/2017 yaitu kelas 10 semester ganjil dan genap.

Tabel 1. Data Nilai Produktif

No	Nama Siswa	Nilai Produktif							Rata - Rata Nilai Produktif
		Semester I			Semester II				
1	Afifah Eva Aryanti	83	86	86	79	83	83	83	83
2	Alita Hujahaeni	82	86	85	79	85	86	86	84
3	Alvina Krisma Yanti	83	86	87	79	90	88	89	86
4	Bagus Priyono	82	86	87	78	86	88	88	85
5	David Awaludin	79	85	85	77	79	75	75	79
6	Desti Septiani	82	85	85	79	81	84	84	82
7	Devita Ariyanti	83	86	85	79	88	87	87	83
8	Dina Amelia R. br Sinulingga	81	85	82	78	80	83	83	81
9	Eka Maulana	82	86	85	78	75	75	75	82
10	Felix Jerry mia	86	91	89	86	92	92	94	86
...
50	Zahra Syachrani	78	79	78	78	83	85	78	79

Tabel 2. Data Nilai Sikap

No	Nama Siswa	Penilaian							Nilai	Kriteria
		Semester I			Semester II					
1	Afifah Eva Aryanti	3	4	4	3	4	4	4	A	Baik
2	Alita Hujahaeni	4	4	3	4	4	4	3	A	Baik
3	Alvina Krisma Yanti	4	4	4	4	3	4	4	A	Baik
4	Bagus Priyono	4	4	4	4	4	4	4	A	Baik
5	David Awaludin	3	3	4	3	3	3	3	B	Cukup
6	Desti Septiani	4	4	3	4	3	4	4	A	Baik
7	Devita Ariyanti	4	4	4	3	4	4	4	A	Baik
8	Dina Amelia R. br Sinulingga	4	4	4	4	4	3	4	A	Baik
9	Eka Maulana	3	4	3	3	4	3	3	B	Cukup
10	Felix Jerry mia	4	4	4	4	4	4	3	A	Baik
...
50	Zahra Syachrani	4	4	3	4	4	4	4	A	Baik

a. Perhitungan Algoritma K-Means

1) Siapkan Dataset

Dataset yang digunakan pada perhitungan ini menggunakan variabel data rata – rata pada

Tabel 3. Dataset

No	Nama Siswa	Rata-rata Nilai Produktif	Angka Mutu
1	Nama 1	83	4,00
2	Nama 2	84	4,00
3	Nama 3	86	4,00
4	Nama 4	85	4,00
5	Nama 5	79	3,00

50	Nama 50	79	4,00

2) Tentukan Jumlah Cluster

Dari dataset perhitungan k-means siswa berkompeten akan dibagi menjadi 2 kelompok/cluster sehingga nilai K pada perhitungan ini adalah 2. Dimana cluster 1 untuk siswa berkompeten dan cluster 2 untuk siswa yang masih harus belajar lagi.

3) Pilih titik centroid secara acak

Pada langkah 2 di atas, dataset perhitungan k-means siswa berkompeten akan dibagi menjadi 2 (dua) kelompok/cluster sehingga titik centroid yang dipilih juga sejumlah 2 (dua).

Tabel 4. Titik centroid awal

Data ke-i	Rata-rata Nilai Produktif	Angka Mutu
13	88	4,00
37	82	4,00

4) Kelompokkan data sehingga terbentuk K buah cluster dengan titik centroid dari setiap cluster

1) Data ke - 1

$$C1 = \sqrt{(83 - 88)^2 + (4,00 - 4,00)^2} = 5,00$$

$$C2 = \sqrt{(83 - 82)^2 + (4,00 - 4,00)^2} = 1,00$$

2) Data ke - 2

$$C1 = 4,00$$

$$C2 = 2,00$$

3) Data ke - 3

$$C1 = 2,00$$

$$C2 = 4,00$$

4) Data ke - 4

$$C1 = 3,00$$

$$C2 = 3,00$$

5) Data ke - 5

$$C1 = 9,06$$

$$C2 = 3,16$$

5) Perbarui nilai titik centroid

Perhitungan manual titik centroid lama menjadi titik centroid baru adalah sebagai berikut :

“Belajar” setelah dilakukan prediksi dan keputusannya berubah menjadi “Berkompeten” sebanyak 4 siswa, kemudian data nyata dengan keputusan “Belajar” setelah dilakukan prediksi dan keputusannya tetap “Belajar: sebanyak 34 siswa kemudian dilakukan perhitungan akurasi dengan cara seperti dibawah ini:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Akurasi = \frac{12 + 34}{12 + 0 + 34 + 4} \times 100\% = 92\%$$

Berdasarkan perhitungan akurasi, maka didapatkan hasil akurasi sebesar 92%.

D. KESIMPULAN

Berdasarkan hasil penelitian yang telah diselesaikan, dapat ditarik kesimpulan sebagai berikut:

1. Dapat memberikan hasil yang dibutuhkan untuk pemetaan siswa berkompeten dengan menggunakan metode Algoritma K-Means Clustering, dari hasil pemetaan diperoleh 16 siswa berkompeten mengikuti lomba kejuruan multimedia dan 34 siswa masih harus belajar dengan menggunakan variabel nilai rata – rata produktif dan angka mutu nilai sikap.
2. Dalam pembuatan aplikasi untuk pemetaan siswa berkompeten dengan metode menerapkan metode Algoritma K-Means

E. DAFTAR PUSTAKA

- [1] Agustin Ely Rahayu, Khoiril Hikmah, Nanik Yustia Ningsih, Abd. Charis Fauzan (2019) Penerapan K-Means Clustering Untuk Penentuan Klasterisasi Beasiswa Bidikmisi Mahasiswa ILKOMNIKA: Journal of Computer Science and Applied Informatics Vol. 1, No. 2, Desember 2019, Halaman 82-86
- [2] Ai Ilah Warnilah (2016) ANALISIS ALGORITMA K-MEANS CLUSTERING UNTUK PEMETAAN PRESTASI SISWA STUDI KASUS SMP NEGERI I SUKAHENING Indonesian Journal on Computer and Information Technology Vol 1 No 1 Mei 2016
- [3] Budi Santosa, 2007, *Data Mining (Teori dan Aplikasi)*, Graha Ilmu, Yogyakarta.
- [4] Edhy Sutanta (2011:91). *Basis Data Dalam Tinjauan Konseptual*
- [5] Fitri Larasati Sibuea & Andy Sapta (2017). PEMETAAN SISWA BERPRESTASI MENGGUNAKAN METODE K-MEANS CLUSTERING JURTEKSI (Jurnal Teknologi dan Sistem Informasi) Vol. IV No. 1, Des 2017, hlm. 85 – 92.
- [6] Green F Mandias, Green A Sandag, Susi Susanti, dan Haryanto Reza Musak (2017) Penerapan Algoritma K-Means Untuk Analisis Prestasi Akademik Mahasiswa Fakultas Ilmu Komputer Universitas Klabat Cogito Smart Journal/VOL. 3/NO. 2/DEC 2017
- [7] Hamdan Yuwafi, Fitri Marisa, Indra Darma Wijaya (2019) IMPLEMENTASI DATA MINING UNTUK MENENTUKAN SANTRI BERPRESTASI DI PP. MANAARULHUDA DENGAN METODE CLUSTERING ALGORITMA K-MEANS Jurnal SPIRIT Vol. 11 No. 1 Mei 2019, hal 22 – 29
- [8] Jaroji, Danuri, Fajri Profesio Putra (2016). K-MEANS UNTUK MENENTUKAN CALON PENERIMA BEASISWA BIDIK MISI DI POLBENG JURNAL INOVTEK POLBENG - SERI INFORMATIKA, Vol. 1, No. 1 , Juni 2016.
- [9] Raymond McLeod, Jr, George P. Schell (2007). *Management Information System (Sistem Informasi Manajemen)*, Prentice Hall.
- [10] Sugiyono. (2016). *Metode Penelitian dan Pengembangan (Research and Development/R&D)*. In Bandung: Alfabeta.
- [11] Tan, P.N., Steinbach, M., Kumar, V. (2006). *Introduction to Data Mining*. Boston: Pearson Education.