



IMPLEMENTASI METODE *K-MEANS* SEBAGAI PENGELOMPOKAN SISWA BERDASARKAN PROSES BELAJAR SISWA

Mohammad Zoqi Sarwani¹⁾, Dian Ahkam Sani²⁾

¹⁾Teknologi Informasi, Universitas Merdeka Pasuruan
Email: zoqi.sarwani@unmerpas.ac.id

²⁾Teknologi Informasi, Universitas Merdeka Pasuruan
Email: dianahkam@unmerpas.ac.id

Abstrak

Siswa adalah bagian dari peserta didik formal yang menempuh jenjang pendidikan dasar, menengah dan menengah atas. Didalam pelaksanaan pembelajaran pada faktor internal (keluarga), para peserta didik memiliki perlakuan yang berbeda-beda, misalnya: mendapat dukungan dan perhatian dari orang tua atau tidak sama sekali. Fasilitas pendukung pembelajaran memadai atau tidak sama sekali. Sama halnya dengan faktor internal, faktor eksternal yang didapat dari peserta didik juga berbeda-beda. Sehingga perlakuan tersebut merupakan bagian utama dari prestasi belajar yang didapat siswa selama menempuh masa pendidikan. Prestasi belajar yang dimaksud adalah nilai yang didapat siswa. Untuk itu prestasi belajar sangatlah penting karena keberhasilan lembaga pendidikan bisa terlihat dari seberapa besar prestasi belajar peserta didik. Pada penelitian ini digunakan metode *K-Means* sebagai pengelompokan siswa berdasarkan proses belajar yang dialami dan silhouette coefficient untuk memvalidasi hasil dari cluster yang diberikan. Dengan menggunakan metode silhouette coefficient pada metode *K-Means* didapatkan hasil 0.69 dengan jumlah cluster sebanyak 3 cluster.

Kata kunci: *Data Mining, Clustering, K-Means, prestasi belajar, siswa*

Abstract

Students are part of the formal learner who pursue elementary, middle and senior secondary education. In the implementation of learning on internal factors (family), the learners have different treatment, for example: get support and attention from parents or not at all. learning support facilities are adequate or not at all. Same thing with internal factors, external factors obtained from learners also vary. So that treatment is a major part of the learning achievement that students get during the education period. Learning achievement in question is the value obtained by students. For that learning achievement is very important because the success of educational institutions can be seen from how much learning achievement learners. In this study used the K-Means method as a grouping of students based on the learning process experienced and silhouette coefficient to validate the results of the given cluster. By using silhouette coefficient method on K-Means method obtained result 0.69 with cluster count as much as 3 cluster.

Keywords: *Data Mining, Clustering, K-Means, learning achievement, student*



I. PENDAHULUAN

Pendidikan adalah salah satu aspek terpenting dalam merubah pola pikir seseorang. Kualitas pendidikan bisa dilihat dari seberapa besar dukungan yang diperoleh secara internal (keluarga) maupun eksternal (lingkungan hidup). Pendidikan adalah usaha, pengaruh, perlindungan dan bantuan yang diberikan kepada anak agar tertuju kepada wawasannya, atau lebih tepatnya membantu anak agar cukup cakap melaksanakan tugas hidupnya sendiri (M. J. Langeveld, 1980). Sedangkan menurut UU No. 20/2003 tentang Sistem Pendidikan Nasional, pendidikan adalah usaha sadar terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukannya, masyarakat bangsa, dan negara.

Data mining merupakan istilah untuk menguraikan informasi yang terdapat dalam sekumpulan data. *Data mining* adalah proses menggunakan teknik statis, matematika, (*artificial intelligent*) kecerdasan buatan, dan machine learning untuk memisah dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban, dkk. 2005). Sedangkan pengertian *data mining* menurut (Last & Maimon, 2000) adalah bagian dari proses *Knowledge Discovery Database* (KDD) yang terdiri dari beberapa tahapan antara lain pemilihan data, *preprocessing*, *transformation*, *data mining*, dan evaluasi hasil. *Data mining* dalam bidang pendidikan umumnya didapat dari sistem informasi akademik.

Clustering adalah pengelompokan sejumlah data kedalam sejumlah grup sehingga masing-masing grup memiliki suatu persamaan yang esensial (Garcia et al, 2002). *Clustering* juga memiliki

peranan penting dalam aplikasi *data mining*.

K-Means adalah salah satu dari sekian banyaknya metode *clustering* pada *data mining*. Pengelompokan pada metode *K-Means* didasari dari titik pusat *cluster* (*centroid*) terdekat dengan data yang ada. Tujuannya adalah memaksimalkan kemiripan data dalam satu klaster dan meminimalkan kemiripan data antar klaster. Ukuran yang digunakan adalah metode jarak (Euclidean). Sehingga kemiripan data secara maksimal didapat dari jarak terpendek antara data terhadap titik pusat klaster (*centroid*).

Penelitian yang berkaitan dengan *clustering* dibidang pendidikan umumnya sudah banyak dilakukan diantaranya oleh (Aranda, Astari, & Natasya, 2016). Penelitian yang dilakukan adalah untuk menentukan pemilihan konsentrasi pada perguruan tinggi. Pada penelitian ini penggunaan metode *K-Means* digunakan untuk membantu dalam menentukan kecenderungan pemilihan konsentrasi studi mahasiswa dari nilai matakuliah yang ada. Adapun input yang digunakan pada system ini adalah Indeks Prestasi Mahasiswa pada semester 3 terkait konsentrasi yang ditawarkan. Hasilnya adalah 4 dari 12 mahasiswa diarahkan untuk mengambil konsentrasi Pemrograman, 4 mahasiswa diarahkan untuk mengambil konsentrasi Multimedia, sementara 3-5 mahasiswa diarahkan untuk mengambil konsentrasi Jaringan Komputer. Selanjutnya penelitian tentang *clustering* juga dilakukan oleh (Munandar & Widyarto, 2013).

Penelitian yang dilakukan untuk mengelompokkan mahasiswa kedalam bidang keahlian berdasar nilai yang diperolehnya. Penelitian ini menggunakan metode *Fuzzy C-Means*. Klaster bidang keahlian dibagi menjadi tiga bidang, diantaranya adalah Multimedia, Pemrograman web dari hasil penelitian penelitian ini terdapat sebanyak 28 mahasiswa masuk kedalam bidang keahlian multimedia, 70 mahasiswa masuk

kedalam bidang keahlian web, dan 28 mahasiswa masuk pada bidang keahlian pemrograman. Siswa adalah bagian dari peserta didik formal yang menempuh jenjang pendidikan dasar, menengah dan menengah atas. Didalam pelaksanaan pembelajaran pada faktor internal (keluarga), para peserta didik memiliki perlakuan yang berbeda-beda, misalnya: mendapat dukungan dan perhatian dari orangtua atau tidak sama sekali. Fasilitas pendukung pembelajaran memadai atau tidak sama sekali. Sama halnya dengan faktor internal, faktor eksternal yang didapat dari peserta didik juga berbeda-beda. Sehingga perlakuan tersebut merupakan bagian utama dari prestasi belajar yang didapat siswa selama menempuh masa pendidikan. Prestasi belajar yang dimaksud adalah nilai yang didapat siswa. Untuk itu prestasi belajar sangatlah penting, karena keberhasilan lembaga pendidikan bisa terlihat dari seberapa besar prestasi belajar peserta didik. Dari pemaparan masalah tersebut maka dilakukanlah penelitian untuk melakukan penggalan data dengan metode *K-Means* terhadap sekumpulan data yang didapat sehingga didapatkan informasi pengelompokan prestasi belajar siswa.

II. KAJIAN LITERATUR

A. Cluster

Cluster merupakan pengelompokan *data record*. Pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam *cluster* lainnya (Nofriansyah, Erwansyah, & Ramadhan, 2016). Ada beberapa kelebihan dan kekurangan dari analisa *cluster* (Raharto, 2008), yaitu:

- a. Dapat mengelompokkan data observasi dalam jumlah besar dan variable yang relative banyak sedemikian, sehingga data yang direduksi dengan kelompok akan mudah dianalisa.

- b. Dapat digunakan dalam skala ordinal data, rasio, dan interval.

B. K-Means

Pada metode *K-Means*, pengelompokan dimulai dengan menentukan jumlah kluster (k) dan centroid awal yang dipilih secara acak. *Centroid* merupakan rata-rata observasi yang berada dalam satu kluster. Observasi yang masuk dalam kluster dihitung berdasar jarak minimum dari observasi ke *centroid* kluster tersebut. Sehingga *centroid* terus melakukan pembaharuan selama terjadi perubahan observasi dalam kluster. Proses akan berhenti setelah kriteria masuk dalam kategori konvergen dengan ciri-ciri setiap observasi tidak ada yang berpindah lagi. Dalam pembentukan kluster Misalkan diberikan matriks data $\{X_{ij}\}$ berukuran $n \times p$ dengan $i = 1, 2, \dots, n$, dan $j = 1, 2, \dots, p$.

1. Asumsikan jumlah kluster awal k .
2. Tentukan *centroid* C_k
3. Hitung jarak setiap objek ke setiap centroid dengan menggunakan jarak *Euclidean* atau dapat ditulis sebagai berikut: $d(x_i, c_i) = \sqrt{(x_i - c_i)^2}$.
4. Setiap objek disusun ke *centroid* terdekat dan kumpulan objek tersebut akan membentuk kluster.
5. Tentukan *centroid* baru dari kluster yang baru terbentuk, dimana *centroid* baru itu diperoleh dari rata-rata setiap objek yang terletak pada kluster yang sama.
6. Ulangi langkah 3, jika *centroid* awal dan baru tidak sama.

Dalam pemilihan jumlah kluster awal, menurut Rencher (2002) dapat ditentukan melalui pendekatan salah satu metode hirarki. Oleh karena itu k (jumlah kluster awal) dapat ditentukan terlebih dahulu dengan metode hirarki, kemudian dilanjutkan pengelompokan objek-objek dengan metode *k-means*.

C. Proses Belajar

Proses belajar adalah tahapan-tahapan yang dilalui dalam pengembangan kemampuan belajar siswa. Kualitas pembelajaran atau pembentukan kompetensi dapat dilihat dari segi proses

dan segi hasil. Segi proses, pembelajaran atau pembentukan kompetensi dikatakan berhasil dan berkualitas apabila seluruhnya atau setidaknya sebagian besar tujuh puluh lima persen peserta didik terlibat secara aktif, di samping menunjukkan kegairahan belajar yang tinggi, semangat belajar yang besar, dan rasa percaya pada diri sendiri. Sedangkan dari segi hasil, proses pembelajaran dikatakan berhasil apabila terjadi perubahan perilaku yang positif pada diri peserta didik seluruhnya atau setidaknya sebagian besar tujuh puluh lima persen (Mulyasa, 2009).

III. METODE PENELITIAN

Pada bagian metode penelitian ini akan diuraikan langkah-langkah sistematis dan terarah yang akan dijadikan acuan sebagai kerangka penelitian. Diantaranya adalah:

1. Studi Literatur

Tahap awal untuk melakukan penelitian adalah dengan melakukan studi literatur berdasarkan topik penelitian yang diambil. Studi literatur yang dikaji berkaitan dengan klaster data dengan menggunakan metode *k-means*. Data penelitian didapat dari hasil survey dan kuisioner terhadap siswa SMK 1 Pasuruan.

2. Analisa dan Perancangan

Analisa merupakan metode setelah studi literatur dan pengumpulan data mengenai kasus yang ada pada penelitian. Analisa yang dilakukan pada penelitian ini dimulai dari masukan data yang digunakan dan dibutuhkan sehingga menghasilkan kelompok pada tiap-tiap data. Data yang diinputkan bersifat nominal dan numerik. Setiap nilai yang ada merupakan kategori dari data tersebut. Tabel 1 adalah gambaran dari data yang disajikan dalam penelitian.

Tabel 1. Dataset Penelitian

No	Macam-macam attribute			
	Keadaan Keluarga	Dukungan Keluarga	Pendidikan Ayah	Pendidikan Ibu
1	1	2	2	4
2	1	1	3	3
3	2	2	4	4

Tabel 2. Macam-macam atribut

No	Macam-macam attribute				
	Akses Internet	Waktu Belajar	Waktu Diam	Sering Keluar	Hubungan Keluarga
1	1	1	4	3	5
2	2	1	2	3	4
3	2	3	1	2	2

No	Macam-macam attribute	
	Kesehatan	Absensi
1	4	70
2	3	60
3	3	50

Keterangan dataset:

- Keadaan keluarga (1: Hidup bersama, 2:Pisah).
- Dukungan dari keluarga (1: iya, 2: tidak).
- Akses Internet saat belajar (1: iya, 2: tidak).
- Pendidikan terakhir Ibu (0: tidak sekolah, 1: SD, 2: SMP, 3: SMA, 4: Perguruan Tinggi).
- Pendidikan terakhir Ayah (0: tidak sekolah, 1: SD, 2: SMP, 3: SMA, 4: Perguruan Tinggi).
- Durasi belajar siswa (1: <2 jam, 2: 2 jam – 5 jam, 3: 5 jam-10 jam, 4:>10 jam).
- Hubungan siswa dengan keluarga (1: sangat buruk, 2: buruk, 3: cukup, 4: baik, 5: sangat baik).
- Durasi berdiam diri saat setelah belajar (1: sangat banyak, 2: banyak, 3: cukup, 4: sedikit, 5: sangat sedikit).
- Kesempatan bermain dengan teman (1: sangat banyak, 2: banyak, 3: cukup, 4: sedikit, 5: sangat sedikit).
- Kesehatan siswa (1: sangat buruk, 2: buruk, 3: cukup, 4: baik, 5: sangat baik).
- Absensi siswa di sekolah (absensi 0-93)

IV. HASIL DAN PEMBAHASAN

Dari dataset yang sudah terkumpul maka dilakukan proses pengelompokan kedalam 3 klaster ($k=3$). Berdasarkan hasil klaster tersebut maka ditentukan titik tengah atau *centroid* sebanyak k secara random atau



acak. Pengukuran jarak pada setiap data terhadap titik tengah dilakukan dengan menggunakan perhitungan jarak *Euclidean*. Dari hasil perhitungan jarak tersebut maka didapatkan nilai jarak terkecil untuk satu *centroid* sehingga data tersebut akan berafiliasi menjadi data klaster dari klaster terdekat. Setelah dipastikan benar bahwa data 1 sampai banyaknya jumlah data masuk kedalam kelompok satu, selanjutnya penentuan *centroid* baru berdasarkan data yang ada pada setiap klaster. Proses ini berulang sampai data masuk pada 3 pengelompokan. Dari hasil ujicoba didapatkan hasil *centroid* sebagaimana berikut:

Tabel 3. Centroid Final

Attribute	Klaster Baik 120 data	Klaster Cukup 104 data	Klaster Kurang 125 data
Keadaan Ortu	1.05	1.1635	1.12
Dukungan Keluarga	1	1	2
Akses Internet	1.1333	1.1346	1.176
Pendidikan Ibu	2.925	2.9135	2.592
Pendidikan Ayah	2.7583	2.6154	2.304
Waktu Belajar	2.125	2.2115	1.88
Hubungan Keluarga	4.075	3.7596	4.016
Waktu Diam	3.3167	3.1058	3.232
Sering Keluar	3.175	2.9423	3.192
Kesehatan	4.7083	2.3462	3.512
Absensi Kelas	5.525	6.5	5.944

Dari hasil clustering menggunakan k-means didapatkan hasil silhouette sebesar 0.69253108828133.

V. KESIMPULAN DAN SARAN

Kesimpulan yang didapat dari hasil ujicoba yang sudah dilakukan diatas adalah bahwa algoritma K-Means mampu melakukan clustering data pada proses belajar siswa dengan nilai silhouette 0.69253108828133. Nilai silhouette sebesar

itu didapat karena jarak antara data sangat dekat.

Saran-sara terhadap penelitian ini adalah perlunya dilakukan proses optimasi jumlah klaster untuk mendapatkan nilai silhouette yang besar.

REFERENSI

- Aranda, Josi., Natasya, Wirda Astari Galvani. 2016, "Penerapan Metode K-Means Cluster Analysis Pada Sistem Pendukung Keputusan Pemilihan Konsentrasi Untuk Mahasiswa International Class TMIK AMIKOM Yogyakarta", Seminar Nasional Teknologi Informasi dan Multimedia.
- Efraim Turban, dkk. 2005. *Decision Support Systems and Intelligent Systems*.
- E. Mulyasa. (2009). *Menjadi Guru Profesional Menciptakan Pembelajaran Kreatif dan Menyenangkan*. Bandung: Remaja Rosdakarya.
- Garcia-Molina, Hector, Ullman,JD., & Wildom, Jennifer. 2002. *Database systems the complete book, International edition*. New Jersey, Prentice Hall.
- Langeveld, M.J. 1980. *Exceptional Children: an Introductory. Survey of Special Education (Sixth Ed)*. Macmillan Publishing. New York.
- Larose, Daniel T. 2005. *Discovering Knowledge in Data : An Introduction to Data Mining*. John Willey & Sons, Inc.
- Munandar, Ali Tb., Widarto, Wahyu Oktri., Harsiti. 2013, "Clustering Data Nilai Mahasiswa Untuk Pengelompokan Konsentrasi Jurusan Menggunakan Fuzzy Cluster Means", Seminar Nasional Aplikasi Teknologi Informasi.
- Nofriansyah, Dicky. 2014. "Konsep Data Mining vs Sistem Pendukung. Keputusan". Edisi I, Yogyakarta.
- O. Maimon and M. Last. 2001. "Knowledge Discovery and Data Mining: The Info-Fuzzy Network (IFN) Methodology", Kluwer Academic Publishers.
- Raharto, I. T. 2008. "Analisis Cluster: Tugas Mata Kuliah Konsep Data Mining dan Teknik Perencanaan." Makasar Program Pasca Sarjana Universitas Hasanuddin.