

Implementasi Metode K-Means Clustering Dalam Sistem Pemilihan Jurusan Di SMK Swasta Harapan Baru

Yerianus Lase¹, Erwin Panggabean²

Address: STMIK Pelita Nusantara, Teknik Informatika, Indonesia^{1,2}

Email : yerianuslase1111@gmail.com¹, erwinpanggabean⁸@gmail.com²

Abstrak

SMK Swasta Harapan Baru memiliki dua jurusan yang dapat dipilih oleh siswa/siswinya, yaitu jurusan Teknik Kendaraan Ringan (TKR) dan jurusan Rekayasa Perangkat Lunak (RPL). Pemilihan jurusan merupakan hak dari siswa/siswi, namun kesalahan pemilihan jurusan dapat menyebabkan penurunan motivasi belajar dan rendahnya prestasi siswa/siswi. Permasalahan ini dapat diselesaikan dengan menggunakan metode *K-Means Clustering*. Proses pengelompokan data ini dilakukan berdasarkan kemiripan data antar kelompok pada titik tengah (*cluster*) dari 2 jurusan. Means merupakan metode yang membutuhkan parameter input sebanyak K dan membagi sekumpulan n objek ke dalam K cluster sehingga tingkat kemiripan antar anggota dalam suatu cluster tinggi, sedangkan tingkat kemiripan dengan anggota pada cluster lain sangat rendah. Setiap data siswa dihitung jarak kedekatannya dengan *centroid* dari masing-masing jurusan, setelah itu dilakukan update terhadap nilai *centroid* berdasarkan nilai rata-rata dari masing-masing kelompok. Bila nilai *centroid* masih berubah, maka dilakukan perhitungan jarak ulang hingga nilai *centroid* tidak berubah, dan data clustering *stabil*. Hasil penelitian adalah aplikasi dapat digunakan untuk membantu proses pemilihan jurusan pada siswa SMK Swasta Harapan Baru secara otomatis, sehingga dapat membantu siswa/siswi yang masih bingung dengan pemilihan jurusan dan dapat menghindari kesalahan dalam pemilihan jurusan.

Keywords – *Selection of Majors, Clusters, Centroids, K-Means Clustering*

1. Latar Belakang

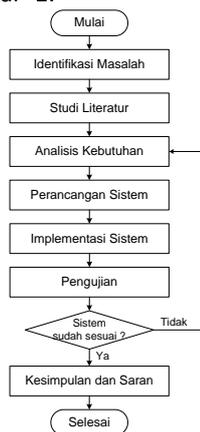
Pemilihan jurusan merupakan hak dari siswa/siswi, namun tidak jarang siswa/siswi masih bingung dengan pemilihan jurusan ini. Akibat dari kesalahan pemilihan jurusan adalah penurunan motivasi belajar dan menyebabkan rendahnya prestasi siswa/siswi karena jurusan yang dipilih ternyata tidak sesuai dengan bakat dan kemampuan siswa/siswi. Salah satu teknik yang dibahas dalam ilmu data mining, yaitu clustering atau pengelompokan data. Clustering dapat digunakan untuk mengidentifikasi kelompok alami dari sebuah kasus yang didasarkan pada kelompok atribut, atau dapat dikatakan clustering mengelompokkan data yang memiliki kemiripan atribut. Salah satu metode clustering paling populer adalah metode K-Means Clustering [1]. K-Means (KM) merupakan salah satu metode data clustering yang

berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda, dikelompokkan ke dalam kelompok yang lainnya. Dengan menggunakan K-Means Clustering, siswa/siswi SMK akan diidentifikasi melalui kemiripan atribut, yaitu nilai akademik dan dibagi menjadi 2 (dua) kelompok, yaitu kelompok yang sesuai dan cocok untuk memilih jurusan Teknik Kendaraan Ringan (TKR) dan kelompok lainnya yang cocok dengan jurusan Rekayasa Perangkat Lunak (RPL). Pengertian clustering dalam keilmuan data mining adalah pengelompokan sejumlah data atau objek ke dalam cluster (group) sehingga setiap data di dalam cluster tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam cluster yang lainnya

Metode K-Means memiliki ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk pengolahan objek dalam jumlah besar. Metode K-Means tidak terpengaruh terhadap urutan objek. Metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya [2].

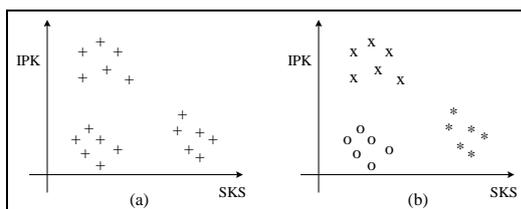
2. Metode

Perancangan aplikasi implementasi metode K-Means Clustering dalam sistem pemilihan jurusan di SMK Swasta Harapan Baru menggunakan metodologi Evolutionary Prototype. Metodologi penelitian dapat dilihat pada gambar 1.



Gambar 1. Metodologi Penelitian *Prototype*

Segmentation Clustering digunakan untuk mengidentifikasi kelompok alami dari sebuah kasus yang didasarkan pada kelompok atribut, atau dapat dikatakan clustering mengelompokkan data yang memiliki kemiripan atribut [3]. Strategi untuk mengelompokkan himpunan data tak berlabel tersebut ke dalam dua kelompok dapat dilakukan dengan melihat sebaran data dan melakukan clustering [4]. Tujuan dari analisa cluster adalah meminimalkan jarak di dalam cluster dan memaksimalkan jarak antar cluster [5].



Gambar 2. Pengelompokan Prestasi Mahasiswa (a) Data Sebelum Pengelompokan (b) Data Setelah Pengelompokan

Clustering merupakan pekerjaan yang memisahkan data atau vektor ke dalam sejumlah kelompok (cluster)

menurut karakteristiknya masing-masing. Data-data yang mempunyai kemiripan karakteristik akan berkumpul dalam cluster yang sama, dan data-data dengan karakteristik berbeda akan terpisah dalam cluster yang berbeda [6]. Tidak diperlukan label kelas untuk setiap data yang diproses dalam suatu clustering, karena nantinya label baru bisa diberikan ketika cluster sudah terbentuk. Karena tidak adanya target label kelas untuk setiap data, maka clustering sering disebut juga pembelajaran tidak berbimbing (*unsupervised learning*) [7]. Ada dua metode clustering yang dikenal, yaitu *hierarchical clustering* dan *partitioning*. Metode *hierarchical clustering* terdiri dari metode complete linkage clustering, single linkage clustering, average linkage clustering dan centroid linkage clustering, sedangkan metode *partitioning* terdiri dari k-means, k-harmonic means dan fuzzy k-means [8]. K-Means (KM) merupakan salah satu metode data clustering non hierarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda, dikelompokkan ke dalam kelompok yang lainnya [9]. K-Kemiripan anggota terhadap cluster diukur dengan kedekatan objek terhadap nilai mean pada cluster atau dapat disebut sebagai centroid cluster atau pusat massa [10]. Cara kerja metode K-Means Clustering:

1. Tentukan k sebagai jumlah cluster yang akan dibentuk.
2. Tentukan k centroid awal secara random atau acak,
3. Hitung jarak setiap objek ke masing-masing centroid dari masing-masing cluster dengan menggunakan metode Euclidian Distance.

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \dots\dots\dots 1$$

4. Alokasikan masing-masing objek ke dalam centroid yang paling dekat.
5. Lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan 2.

$$v = \sum_{i=1}^n x_i/n \quad ; i=1,2,3\dots n \dots\dots\dots 2$$

v adalah centroid pada cluster; xi adalah objek ke-i; n adalah banyaknya objek atau jumlah objek yang menjadi anggota cluster. Ulangi langkah 3 jika nilai centroid baru berbeda dengan centroid lama.

3. Hasil

Analisis kebutuhan sistem merupakan proses identifikasi dan evaluasi permasalahan-permasalahan yang ada,

sehingga sistem yang dibangun sesuai dengan kriteria yang diharapkan. Aplikasi implementasi metode K-Means Clustering dalam sistem pemilihan jurusan di SMK Swasta Harapan Baru harus memenuhi syarat dan kebutuhan sebagai berikut:

- Aplikasi mampu menerima input, berupa data siswa yang disertai dengan nilai yang diperoleh oleh nilai akademik siswa dalam 5 (lima) bidang mata pelajaran, yaitu Matematika, Bahasa Indonesia, Bahasa Inggris, Ilmu Pengetahuan Alam (IPA), dan Ilmu Pengetahuan Sosial (IPS).
- Aplikasi dapat melakukan proses pemilihan jurusan dengan metode K-Means Clustering dan membagi siswa ke dalam dua kelompok siswa/siswi yang sesuai dengan kedua jurusan pada SMK.
- Aplikasi dapat membagi siswa/i ke dalam dua jurusan, yaitu jurusan Teknik Kendaraan Ringan (TKR) dan Rekayasa Perangkat Lunak (RPL).
- Admin dapat mengubah nilai akademik yang dianggap sebagai syarat atau rekomendasi untuk memasuki suatu jurusan.
- Aplikasi menyediakan output berupa laporan pemilihan jurusan bagi siswa/i.

Data penelitian diperoleh dari SMK Swasta Harapan Baru yang berlokasi di jalan Pintu Air, IV Gg. Lingga Raya, Medan Johor, Siti Rejo I, Medan Kota, Kota Medan, Sumatera Utara 20143. Data penelitian berupa data dan nilai siswa untuk dimasukkan sebagai *input* dalam sistem pemilihan jurusan. Contoh data siswa dapat dilihat pada tabel 1.

Tabel 1. Data Siswa

| No. | Nama Siswa | Matematika | IPA | Bahasa Inggris | Bahasa Indonesia | IPS |
|-----|------------|------------|-----|----------------|------------------|-----|
| 1 | Engelina | 85 | 90 | 75 | 80 | 90 |
| 2 | Prencky | 90 | 85 | 70 | 75 | 70 |
| 3 | Risnawati | 75 | 80 | 75 | 70 | 65 |
| 4 | Afriadi | 55 | 45 | 50 | 60 | 65 |
| 5 | Eliel | 50 | 60 | 40 | 65 | 50 |
| 6 | Ferdi | 35 | 45 | 40 | 60 | 60 |

Penelitian menggunakan metode *K-Means Clustering* untuk melakukan pengelompokan siswa, berdasarkan nilai akademik ke dalam 2 jurusan, yaitu jurusan TKR dan jurusan RPL. Data yang mirip akan dikelompokkan dalam satu *cluster* yang sama. Berdasarkan data siswa pada tabel 1, data yang akan Data(1) = [85, 90, 75, 80, 90] Jarak dengan *Centroid-1*

$$= \sqrt{((85 - 85)^2) + (85 - 85)^2 + (85 - 85)^2 + (85 - 85)^2 + (85 - 85)^2} = 0$$

Jarak dengan *Centroid-2*

$$= \sqrt{((85 - 75)^2) + (85 - 75)^2 + (85 - 75)^2 + (85 - 75)^2 + (85 - 75)^2} = 182.21$$

Jarak terdekat dengan *Centroid-1*, maka data-1 termasuk pada *Cluster* ke-1

$$\text{Data}(2) = [90, 85, 70, 75, 70]$$

Jarak dengan *Centroid-1*

$$= \sqrt{((90 - 85)^2) + (90 - 85)^2 + (90 - 85)^2 + (90 - 85)^2 + (90 - 85)^2} = 170.5$$

Jarak dengan *Centroid-2*

$$= \sqrt{((90 - 75)^2) + (90 - 75)^2 + (90 - 75)^2 + (90 - 75)^2 + (90 - 75)^2} = 158.31$$

Jarak terdekat dengan *Centroid-2*, maka data-2 termasuk pada *Cluster* ke-2

$$\text{Data}(3) = [75, 80, 75, 70, 65]$$

Jarak dengan *Centroid-1*

$$= \sqrt{((75 - 85)^2) + (75 - 85)^2 + (75 - 85)^2 + (75 - 85)^2 + (75 - 85)^2} = 182.21$$

Jarak dengan *Centroid-2*

$$= \sqrt{((75 - 75)^2) + (75 - 75)^2 + (75 - 75)^2 + (75 - 75)^2 + (75 - 75)^2} = 0$$

Jarak terdekat dengan *Centroid-2*, maka data-3 termasuk pada *Cluster* ke-2

$$\text{Data}(4) = [55, 45, 50, 60, 65]$$

Jarak dengan *Centroid-1*

$$= \sqrt{((55 - 85)^2) + (55 - 85)^2 + (55 - 85)^2 + (55 - 85)^2 + (55 - 85)^2} = 313.82$$

Jarak dengan *Centroid-2*

$$= \sqrt{((55 - 75)^2) + (55 - 75)^2 + (55 - 75)^2 + (55 - 75)^2 + (55 - 75)^2} = 209.09$$

Jarak terdekat dengan *Centroid-2*, maka data-4 termasuk pada *Cluster* ke-2

$$\text{Data}(5) = [50, 60, 40, 65, 50]$$

Jarak dengan *Centroid-1*

$$= \sqrt{((50 - 85)^2) + (50 - 85)^2 + (50 - 85)^2 + (50 - 85)^2 + (50 - 85)^2} = 320.73$$

Jarak dengan *Centroid-2*

$$= \sqrt{((50 - 75)^2) + (50 - 75)^2 + (50 - 75)^2 + (50 - 75)^2 + (50 - 75)^2} = 239.77$$

Jarak terdekat dengan *Centroid-2*, maka data-5 termasuk pada *Cluster* ke-2

$$\text{Data}(6) = [35, 45, 40, 60, 60]$$

Jarak dengan *Centroid-1*

$$= \sqrt{((35 - 85)^2) + (35 - 85)^2 + (35 - 85)^2 + (35 - 85)^2 + (35 - 85)^2} = 338.84$$

Jarak dengan *Centroid-2*

$$= \sqrt{((35 - 75)^2) + (35 - 75)^2 + (35 - 75)^2 + (35 - 75)^2 + (35 - 75)^2} = 256.97$$

Jarak terdekat dengan *Centroid-2*, maka data-6 termasuk pada *Cluster* ke-2

Hasil perhitungan jarak dapat dilihat pada tabel 2.

Tabel 2. Jarak Data dengan *Cluster*

| i | Data ke-i | Jarak dengan <i>Centroid</i> ke-j | | Cluster Terdekat |
|---|----------------------|-----------------------------------|--------|------------------|
| | | j = 1 | j = 2 | |
| 1 | [85, 90, 75, 80, 90] | 0 | 182.21 | 1 |
| 2 | [90, 85, 70, 75, 70] | 170.5 | 158.31 | 2 |
| 3 | [75, 80, 75, 70, 65] | 182.21 | 0 | 2 |
| 4 | [55, 45, 50, 60, 65] | 313.82 | 209.09 | 2 |
| 5 | [50, 60, 40, 65, 50] | 320.73 | 239.77 | 2 |
| 6 | [35, 45, 40, 60, 60] | 338.84 | 256.97 | 2 |

1. Lakukan perhitungan ulang dan ganti nilai *centroid* dengan jumlah rata-rata (*mean*) dari semua anggota *cluster* sebagai berikut:

- a. $Centroid-1 = Data(1) = [85, 90, 75, 80, 90]$
- b. $Centroid-2 = (Data(2) + Data(3) + Data(4) + Data(5) + Data(6)) / 5$. $Centroid-2 = ([90, 85, 70, 75, 70] + [75, 80, 75, 70, 65] + [55, 45, 50, 60, 65] + [50, 60, 40, 65, 50] + [35, 45, 40, 60, 60]) / 5$. Nilai-1 = $(90 + 75 + 55 + 50 + 35) / 5 = 61$

- Nilai-2 = $(85 + 80 + 45 + 60 + 45) / 5 = 63$
 Nilai-3 = $(70 + 75 + 50 + 40 + 40) / 5 =$ Nilai-3 = 55
 Nilai-4 = $(75 + 70 + 60 + 65 + 60) / 5 =$ Nilai-4 = 66
 Nilai-5 = $(70 + 65 + 65 + 50 + 60) / 5 =$ Nilai-5 = 62

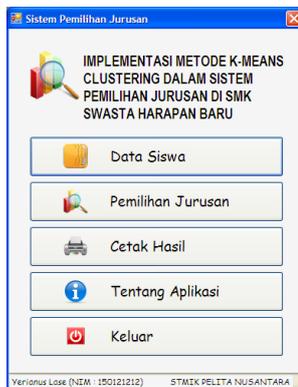
1. $Centroid-2 = [61, 63, 55, 66, 62]$
2. Periksa apakah nilai *centroid* berubah atau tidak.
 - a. *Centroid-1* tidak berubah.
 - b. *Centroid-2* berubah.
3. Oleh karena nilai *centroid* masih berubah, maka ulangi perhitungan jarak setiap data dengan *centroid*. Update nilai *centroid* hingga nilai *centroid* tidak berubah.

Perulangan proses *clustering* dilakukan hingga nilai *centroid* tidak berubah. Pada perulangan ke-3, nilai *centroid* tidak mengalami perubahan lagi, sehingga hasil *clustering* pada setiap data dan pemilihan jurusan dapat dilihat pada tabel 3.

Tabel 3. Hasil Pemilihan Jurusan

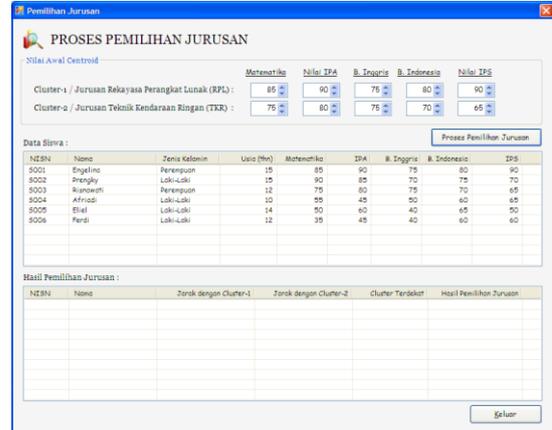
| No. | Nama Siswa | Nilai Akademik Bidang Studi [Matematika, IPA, B. Inggris, B. Indonesia, IPS] | Termasuk Cluster | Pemilihan Jurusan |
|-----|------------|--|------------------|-------------------|
| 1 | Engelina | [85, 90, 75, 80, 90] | 1 | RPL |
| 2 | Prengky | [90, 85, 70, 75, 70] | 1 | RPL |
| 3 | Risnawati | [75, 80, 75, 70, 65] | 1 | RPL |
| 4 | Afriadi | [55, 45, 50, 60, 65] | 2 | TKR |
| 5 | Eliel | [50, 60, 40, 65, 50] | 2 | TKR |
| 6 | Ferdi | [35, 45, 40, 60, 60] | 2 | TKR |

Hasil implementasi metode *K-Means Clustering* dalam sistem. Form Utama akan muncul saat aplikasi dijalankan, seperti terlihat pada, Gambar 3.

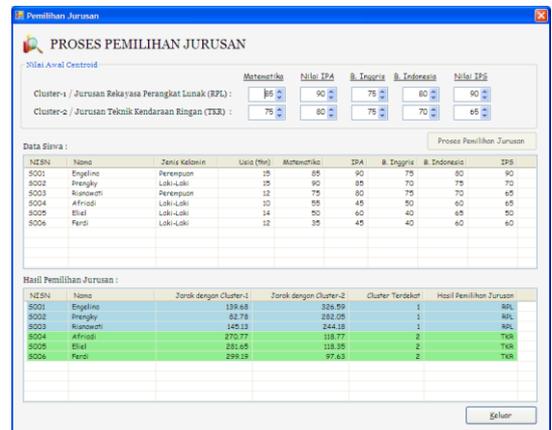


Gambar 3. Form Utama

Form Pemilihan Jurusan berfungsi untuk melakukan *clustering* terhadap data siswa berdasarkan nilai akademik, dan melakukan pemilihan jurusan terhadap data siswa. Nilai awal *centroid* untuk setiap cluster dapat dimasukkan pada form ini, Gambar 4. Pengguna dapat menekan tombol "Proses Pemilihan Jurusan" pada form Pemilihan Jurusan untuk memulai proses pemilihan jurusan, Gambar 5.



Gambar 4. Form Pemilihan Jurusan



Gambar 5. Form Proses Pemilihan Jurusan

4. Kesimpulan

1. Berdasarkan data yang ada, siswa akan dibagi ke dalam 2 jurusan, yaitu jurusan Teknik Kendaraan Ringan (TKR) dan jurusan Rekayasa Perangkat Lunak (RPL).
2. Analisa cluster berfungsi untuk menemukan kumpulan objek hingga objek-objek ditempatkan di dalam kelompok yang sama (yang mempunyai hubungan) dengan yang lain, dan berbeda (tidak mempunyai hubungan) dengan objek-objek di dalam kelompok lain.

Acknowledgement

Terima Kasih yang sebesar-besarnya pada STMIK Pelita Nusantara.

References

- [1] Apriadi Bahar, Bambang Pramono, dan Laode Hasnuddin S Sagala, "Penentuan Strategi Penjualan Alat-Alat Tattoo di Studio Sonyxtattoo Menggunakan Metode K-Means Clustering", *Seman TIK*, Vol.2, No.2 Jul-Des 2016.
- [2] Sheih Al Syahdan, Anita Sindar, Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota, *Jurnal Nasional Komputasi dan Teknologi Informasi*, Vol. 1 No. 2, Oktober 2018, hal: 56-63.
- [3] Herman Mulyana, Pemakaian Metode Asosiasi Dalam Data Mining Untuk Penjualan Lebih Dari Satu Jenis Produk Pada Perusahaan, *Jurnal Pilar Nusa Mandiri*, Vol.X No.1, Maret 2014, hal 47-56.
- [4] Anita Sindar Sinaga, Implementasi OLAP Menggunakan Dashboard Holistics Software Pada LPPM STMIK Pelita Nusantara, *Jutikomp*, Volume 2 Nomor 1 Tahun 2019, hal : 55-59.
- [5] Saut Parsaoran Tamba, Felix Toknady Kesuma, Feryanto, Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering, *Jurnal Sistem Informasi Ilmu Komputer Prima (JUSIKOM PRIMA)*, Vol. 2 No. 2, Maret 2019, hal: 67-72.
- [6] Anita Sindar RM Sinaga, Data Mining Informasi Tingkat Kelulusan Mahasiswa STMIK Pelita Nusantara, *Jurnal Manajemen dan Informatika Komputer Pelita Nusantara*, Tahun 2017 hal 27-35.
- [7] T. Taslim, Penerapan Algorithma K-Mean Untuk Clustering Data Obat Pada Puskesmas Rumbai, *Jurnal Teknologi Informasi & Komunikasi Digital Zone*, Volume 7, Nomor 2, November 2016:108-114.
- [8] Linda Maulida, Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisataunggulan Di Prov. Dki Jakarta Dengan K-Means, *JISKa (Jurnal Informatika Sunan Kalijaga)*, Vol. 2, No. 3, Januari, 2018, Pp. 167–174.
- [9] Mardalius, Implementasi Algoritma K-Means Clustering Untuk Menentukan Kelas Kelompok Bimbingan Belajar Tambahan (Studi Kasus : Siswa Sma Negeri 1 Ranah Pesisir), *Proceding SEMILOKA ROYAL 2017 "Teknologi Mobile" 02 Desember 2017 Hal-105.*
- [10] B. Poerwanto, R.Y. Fa'rifah, Analisis Cluster K-Means Dalam Pengelompokan Kemampuan Mahasiswa, *Jurnal Scientific Pinisi*, Volume 2, Nomor 2, Oktober 2016, hlm. 92-96.