

PERANCANGAN SISTEM REKOMENDASI PENJURUSAN PADA SEKOLAH MENENGAH KEJURUAN SANTO PAULUS MENGGUNAKAN METODE *K-MEANS CLUSTERING*

Isaias F.X.F.A. Felnditi¹⁾, Desi Arisandi²⁾, Tri Sutrisno³⁾

¹⁾²⁾³⁾Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jl. Letjen S. Parman No. 1, Grogol Petamburan, Jakarta Barat 11440 Indonesia
email: felnditi@gmail.com¹⁾, desia@fti.untar.ac.id²⁾, tris@fti.untar.ac.id³⁾

ABSTRACT

Santo Paulus Vocational School in Jakarta, is one of the Private Schools in Central Jakarta which holds majors for grade X (ten) students who will continue their studies to grade XI (eleven). The obstacle that is often found in the majors in SMK Santo Paulus Jakarta is the difficulty in determining which students meet the criteria to occupy certain majors. This is because the majors process is still done manually, so it requires quite a long time and is considered inappropriate. This study, purpose to the K-Means Clustering algorithm for the decision support system of student majors at SMK Santo Paulus Jakarta. K-Means Clustering algorithm can group data that has attributes and has a large amount of data, so that it can be utilized in the determination of student majors in accordance with students' academic abilities.

Based on the results of the implementation of the K-Means Clustering algorithm for the majors decision support system, the conclusion that can be drawn is the K-Means-Clustering algorithm can be applied in the student majors decision support system.

Key Words

Grades, K-Means Clustering, Majors, Students.

1. Pendahuluan

Seiring dengan perkembangan jaman, komputer semakin banyak berperan di dalam kehidupan masyarakat. Hampir semua bidang kehidupan telah menggunakan komputer sebagai alat bantu. Perkembangan komputer yang begitu pesat dapat langsung dirasakan manfaatnya oleh masyarakat. Namun ada beberapa masalah yang dapat kita lihat masih dilakukan secara manual. Salah satunya adalah masalah penjurusan di SMK Santo Paulus. Masalah penentuan jurusan untuk SMK sering sekali menjadi permasalahan di sekolah, karena sulitnya menentukan siswa mana yang memenuhi kriteria untuk menempati jurusan tertentu. Seorang guru sering mengalami kesulitan dalam menentukan siswa mana yang memenuhi kriteria untuk menempati jurusan tertentu. Hal ini dikarenakan proses

penjurusan masih dilakukan dengan manual. Salah satu algoritma yang digunakan dalam pengelompokan data adalah algoritma *K-Means Clustering*. Algoritma ini mempunyai kelebihan yang dapat diterapkan dan dijalankan untuk aplikasi penentuan jurusan yang ingin dirancang, relatif cepat untuk diadaptasi, dan paling banyak dipraktekkan dalam data mining. Algoritma ini termasuk salah satu algoritma paling penting dalam data mining.

Metode *K-Means Clustering* merupakan metode untuk mengelompokkan objek atau data yang mempunyai atribut dan mempunyai jumlah data yang banyak ke dalam bentuk satu atau lebih kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster/kelompok yang sama. Data yang digunakan untuk proses analisis metode *K-Means Clustering* adalah data nilai. Nilai yang diambil sebagai dasar untuk proses Clustering adalah nilai rata-rata semester I dan semester II di kelas X (sepuluh). Aturan dan metode algoritma *K-Means Clustering* dapat diterapkan pada sebuah Program bantu untuk mengelompokkan data nilai siswa. Langkah-langkah algoritma *K-Means* diterjemahkan ke dalam bahasa pemrograman untuk melakukan tugas klastering data. Sekelompok data tentang nilai siswa peserta didik dimasukkan ke dalam Input Program, kemudian Program melakukan pengolahan data yang digunakan sebagai kriteria yang telah ditetapkan sesuai langkah algoritma *K-Means Clustering*, dan hasilnya berupa klaster data yang menjadi bahan pertimbangan untuk kriteria siswa yang akan menempati jurusan tertentu. Berdasarkan permasalahan yang telah diuraikan, maka dalam hal ini dirancang sebuah sistem dengan judul "Perancangan Sistem Rekomendasi Penjurusan Pada Sekolah Menengah Kejuruan Santo Paulus menggunakan Metode *K-Means Clustering*".

2. Dasar Teori

2.1 Analisis Sistem

Analisis sistem adalah fase pengembangan sistem yang menentukan sistem informasi apa yang harus

dilakukan untuk memecahkan masalah yang sudah ada dengan mempelajari sistem dan proses kerja untuk mengidentifikasi kekuatan, kelemahan dan peluang untuk perbaikan.

Analisis sistem terdiri dari mendefinisikan masalah, mengidentifikasi penyebabnya, menentukan solusi, dan mengidentifikasi kebutuhan informasi yang harus memenuhi dengan solusi sistem. Analisis Sistem dapat dilihat pada Gambar 1.



Gambar 1 Flowchart Analisis Sistem

2.2 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam basis data.

Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam basis data besar.

Menurut Gartner Group Data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika.

Nama Data mining sebenarnya mulai dikenal sejak tahun 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang penting dalam berbagai bidang, mulai dari bidang akademik, bisnis, hingga medis.

Data mining dapat diterapkan pada berbagai bidang yang mempunyai sejumlah data, tetapi karena wilayah penelitian dengan sejarah yang belum lama, maka Data mining masih diperdebatkan posisi bidang pengetahuan yang memilikinya. Maka, Daryl Pregibon menyatakan bahwa “Data mining adalah campuran dari

statistik, kecerdasan buatan, dan riset basis data” yang masih berkembang.

Kita tidak dapat menyamakan pencarian tertentu dari objek individu (jenis apa pun) dengan pencarian Data mining. Pencarian dalam Data mining tidak mencari secara individualis, tetapi set individualis, atau dengan kata lain dikelompokkan dengan kriteria tertentu. Dengan kata lain, perbedaan antara pencarian biasa dengan Data mining dianalogikan dengan pencarian pohon dengan pencarian hutan; “tidak dapat menilai hutan dengan hanya menilai pohon” [2].

2.3 K-Means Clustering

K-Means Clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok.

Data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil.

Langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut:[2]

1. Pilih jumlah cluster k.
2. Inisialisasi k pusat cluster ini dapat dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberi nilai awal dengan angka-angka random.
3. Alokasikan semua data/objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data dengan pusat cluster. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat cluster. Jarak paling antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam cluster mana.

Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak Euclidean Distance yang dirumuskan sebagai berikut:[2]

$$\begin{aligned}
 D(i, j) &= \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \\
 &\dots\dots(1)
 \end{aligned}$$

Keterangan:

$D(i, j)$ = Jarak data ke i ke pusat cluster j.

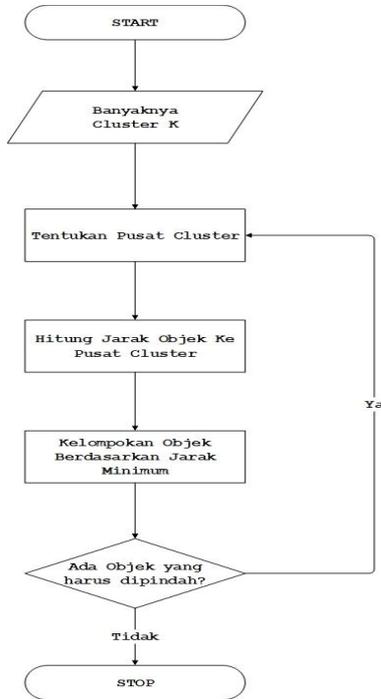
X_{ki} = Data ke i pada atribut data ke k.

X_{kj} = Titik pusat ke j pada atribut ke k.

4. Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata dari semua data/ objek dalam cluster tertentu. Jika dikehendaki dapat juga menggunakan median dari

cluster tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang dapat dipakai.

5. Tugaskan lagi setiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat cluster tidak berubah lagi.



Gambar 2 Flowchart Metode K-Means Clustering

2.4 Clustering

Clustering merupakan salah satu metode analisis Data mining yang dapat digunakan untuk berbagai bidang salah satunya adalah dalam pembuatan perancangan sistem rekomendasi penjurusan pada SMK Santo Paulus. Proses clustering mengelompokkan sejumlah data yang dianalisis sesuai dengan kesamaan karakteristik setiap datanya. Data dari satu cluster dengan cluster lain mempunyai karakteristik yang berbeda.

Karakteristik setiap cluster yang dianalisis tidak ditentukan sebelumnya. Proses clustering harus dilakukan oleh pihak yang memahami karakteristik domain data yang dianalisis [1].

Proses clustering juga dapat digunakan dalam pra-pemrosesan data sebelum data diolah dengan menggunakan teknik Data mining lainnya untuk meningkatkan pemahaman karakteristik domain data. Pemilihan atribut dalam data set yang akan digunakan dalam clustering, penting dilakukan untuk mendapatkan hasil clustering yang lebih baik [4].

3. Hasil Percobaan K-Means Clustering

Berikut adalah contoh perhitungan K-Means Clustering, Prosedur awal K-Means Clustering adalah menentukan jumlah cluster, kemudian sembarang obyek atau element pertama dalam cluster dapat dipilih untuk dijadikan titik tengah (centroid point) cluster, menentukan koordinat titik tengah terhadap cluster, menentukan jarak setiap obyek terhadap koordinat titik tengah dan mengelompokkan obyek - obyek tersebut berdasarkan pada jarak minimumnya. Algoritma K-Means Clustering selanjutnya akan melakukan pengulangan langkah - langkah tersebut hingga terjadi kestabilan. Dikatakan stabil apabila tidak ada obyek yang dapat dipindahkan lagi. Berikut adalah data perhitungan K-Means Clustering :

Tabel 1 Tabel Cluster Acak

C	X1	X2	X3	X4	X5
1	75	76	75	80	71
2	83	85	86	76	84

Dimana:

C = Cluster

X1 = Titik Pusat ke 1 (Nilai Random dari pelajaran Matematika)

X2 = Titik Pusat ke 2 (Nilai Random dari pelajaran Bahasa Indonesia)

X3 = Titik Pusat ke 3 (Nilai Random dari pelajaran Bahasa Inggris)

X4 = Titik Pusat ke 4 (Nilai Random dari pelajaran Pengenalan Akutansi Keuangan)

X5 = Titik Pusat ke 5 (Nilai akhir mata pelajaran Pengenalan Administrasi Perkantoran)

C1 = (75, 76, 75, 80, 71)

C2 = (83, 85, 86, 76, 84)

Tabel 2 Data Awal Siswa

NO	NIS	NAMA	NILAI AKHIR				
			MAT	B.ING	B.INDO	P.AK	P.AP
1	2665	Siswa.Ka-1	83	79	83	83	80
2	2734	Siswa.Ka-2	81	77	83	80	72
3	2735	Siswa.Ka-3	78	82	81	80	77
4	2668	Siswa.Ka-4	84	83	83	83	77
5	2669	Siswa.Ka-5	83	82	80	80	82
6	2670	Siswa.Ka-6	78	87	85	84	86
7	2671	Siswa.Ka-7	78	81	89	80	84
8	2694	Siswa.Ka-8	79	76	87	78	71
9	2695	Siswa.Ka-9	79	84	81	84	72
10	2674	Siswa.Ka-10	77	84	88	79	72

Langkah selanjutnya dilakukan perhitungan antara Data awal Siswa dan Tabel Cluster Acak dan didapatkan hasil iterasi pertama sebagai berikut :

Tabel 3 Iterasi 1

NO	NIS	NAMA	NILAI AKHIR					C1	C2	HASIL
			MAT	BING	BINDO	PAK	PAP			
1	2665	Siswa Ke-1	83	79	83	83	80	13,66637	10,4800840	C2
2	2734	Siswa Ke-2	81	77	83	80	72	10,8995	13,3940482	C1
3	2735	Siswa Ke-3	78	82	81	80	77	10,81665	11,13570073	C1
4	2668	Siswa Ke-4	84	83	83	83	77	13,45962	10,38300324	C2
5	2669	Siswa Ke-5	83	82	80	80	82	13,68439	8,062257748	C2
6	2670	Siswa Ke-6	78	87	85	84	86	21,70253	9,899494937	C2
7	2671	Siswa Ke-7	78	81	89	80	84	18,97489	8,124038405	C2
8	2694	Siswa Ke-8	79	76	87	78	71	12,80625	16,46207785	C1
9	2695	Siswa Ke-9	79	84	81	84	72	11,51256	15,8118883	C1
10	2674	Siswa Ke-10	77	84	88	79	72	13,45962	13,92030828	C2

Suatu data akan menjadi anggota dari suatu cluster yang memiliki jarak terkecil dari pusat clusternya. Misalkan untuk data pertama, jarak terkecil diperoleh pada cluster kedua, sehingga data pertama akan menjadi anggota dari cluster kedua. Demikian juga untuk data kedua, jarak terkecil ada pada cluster pertama, maka data tersebut akan masuk pada cluster pertama. Posisi cluster selengkapnya dapat dilihat Setelah didapatkan data iterasi pertama dilakukan perhitungan untuk mendapatkan Nilai Titik Centroid 1&2 yang baru yaitu titik centroid 1&2 iterasi pertama dan di dapatkan hasil sebagai berikut :

Tabel 4 Titik Centroid 1&2 Iterasi 1

Perbarui Nilai Titik Centroid 1&2 Iterasi 1					
Centroid	MAT	BING	BINDO	PAK	PAP
C1	79,25	79,75	83	80,5	73
C2	80,5	82,66666667	84,66666667	81,5	80,16666667

Langkah selanjutnya adalah melakukan perhitungan untuk mencari data iterasi kedua yaitu dengan cara yang sama seperti mencari data iterasi pertama dan didapatkan hasil sebagai berikut :

Tabel 5 Data Iterasi 2

NO	NIS	NAMA	NILAI AKHIR					C1	C2	HASIL
			MAT	BING	BINDO	PAK	PAP			
1	2665	Siswa Ke-1	83	79	83	83	80	8,559127	4,97487106	C2
2	2734	Siswa Ke-2	81	77	83	80	72	3,446011	10,30120396	C1
3	2735	Siswa Ke-3	78	82	81	80	77	5,18411	5,689260618	C1
4	2668	Siswa Ke-4	84	83	83	83	77	7,441493	5,236026991	C2
5	2669	Siswa Ke-5	83	82	80	80	82	10,49251	5,830893284	C2
6	2670	Siswa Ke-6	78	87	85	84	86	15,47173	8,888877039	C2
7	2671	Siswa Ke-7	78	81	89	80	84	12,66399	6,88954809	C2
8	2694	Siswa Ke-8	79	76	87	78	71	6,554193	12,18249791	C1
9	2695	Siswa Ke-9	79	84	81	84	72	5,847088	8,338767884	C1
10	2674	Siswa Ke-10	77	84	88	79	72	7,167658	6,901702011	C1

Setelah didapatkan data iterasi kedua dilakukan perhitungan untuk mendapatkan Nilai Titik Centroid 1&2 yang baru yaitu titik centroid 1&2 iterasi 2 dan di dapatkan hasil sebagai berikut :

Tabel 6 Titik Centroid 1&2 Iterasi 2

Perbarui Nilai Titik Centroid 1&2 Iterasi 2					
Centroid	MAT	BING	BINDO	PAK	PAP
C1	78,8	80,6	84	80,2	72,8
C2	81,2	82,4	84	82	81,8

Karena titik centroid baru iterasi pertama dan kedua belum sama maka iterasi dilanjutkan dengan mencari data iterasi ketiga dan didapatkan hasil sebagai berikut :

Tabel 7 Data Iterasi 3

NO	NIS	NAMA	NILAI AKHIR					C1	C2	HASIL
			MAT	BING	BINDO	PAK	PAP			
1	2665	Siswa Ke-1	83	79	83	83	80	8,899391	4,47480287	C2
2	2734	Siswa Ke-2	81	77	83	80	72	4,413615	11,41127408	C1
3	2735	Siswa Ke-3	78	82	81	80	77	5,4111	6,814880125	C1
4	2668	Siswa Ke-4	84	83	83	83	77	7,689931	5,785441122	C2
5	2669	Siswa Ke-5	83	82	80	80	82	10,94722	4,841487375	C2
6	2670	Siswa Ke-6	78	87	85	84	86	15,20789	7,35118058	C2
7	2671	Siswa Ke-7	78	81	89	80	84	12,28959	6,782578181	C2
8	2694	Siswa Ke-8	79	76	87	78	71	6,107883	13,68887207	C1
9	2695	Siswa Ke-9	79	84	81	84	72	5,973274	10,79079877	C1
10	2674	Siswa Ke-10	77	84	88	79	72	5,734188	11,88444961	C1

Setelah didapatkan data iterasi ketiga dilakukan perhitungan untuk mendapatkan Nilai Titik Centroid 1&2 yang baru pada iterasi ketiga dan di dapatkan hasil sebagai berikut :

Tabel 8 Titik Centroid 1&2 Iterasi 3

Perbarui Nilai Titik Centroid 1&2 Iterasi 3					
Centroid	MAT	BING	BINDO	PAK	PAP
C1	78,8	80,6	84	80,2	72,8
C2	81,2	82,4	84	82	81,8

Pada tahap ini perhitungan berhenti karena titik centroid baru kedua dan ketiga mendapatkan hasil yang sama dan dapat disimpulkan bahwa perhitungan hasil penjurusan menggunakan 10 Data Siswa berhenti pada iterasi ketiga dan didapatkan hasil sebagai berikut :

Tabel 9 Hasil Penjurusan Siswa

Dapat Disimpulkan Bahwa :						
AK	C1	Siswa Ke-1	Siswa Ke-3	Siswa Ke-8	Siswa Ke-9	Siswa Ke-10
AP	C2	Siswa Ke-1	Siswa Ke-4	Siswa Ke-5	Siswa Ke-6	Siswa Ke-7

4. Kesimpulan

Berdasarkan pembuatan dan pengujian dari aplikasi sistem rekomendasi penjurusan SMK Santo Paulus berbasis web menggunakan metode *K-Means Clustering*, dapat disimpulkan bahwa :

1. Tingkat akurasi dari metode *K-Means Clustering* menggunakan 200 Data Siswa mendapatkan tingkat akurasi sebesar 62,5% dari data minat asli siswa.
2. Aplikasi mempermudah Tata Usaha dalam memasukkan Data Siswa, Data Nilai Siswa, dan Data Penjurusan Siswa.
3. Aplikasi mempermudah Kepala Sekolah, Guru, dan Siswa dalam melihat hasil penjurusan.
4. Aplikasi mempermudah Murid dalam rekomendasi penjurusan pada SMK Santo Paulus.
5. Aplikasi telah berhasil memberikan rekomendasi penjurusan terhadap murid berdasarkan data nilai yang menjadi kriteria dalam penjurusan.
6. Hasil perhitungan *K-Means Clustering* didapatkan berdasarkan jumlah titik centroid akhir atau yang sedang dicari dengan titik centroid sebelumnya bernilai sama.
7. Dalam pengujian data pada metode *K-Means Clustering* hasil sebagian besar sama dengan penjurusan asli pada data siswa tersebut dengan tingkat akurasi 62,5%.

REFERENSI

- [1] Saputra, Angga, Bagus Mulyawan dan Tri Sutrisno. “Rekomendasi Lokasi Wisata Kuliner di Jakarta Menggunakan Metode K-Means Clustering dan Simple Additive Weighting”. Mahasiswa Program Studi Teknik Informatika Universitas Tarumanagara Jakarta. Vol 7, No 1 (2019): Jurnal Ilmu Komputer dan Sistem Informasi.
- [2] Rochita. “Penggunaan Algoritma K-Means dalam Penentuan Jurusan Untuk SMA (Studi Kasus di SMA Negeri 1 Jakarta)”. Hal.3. <https://digilib.esaunggul.ac.id/public/UEU-Undergraduate-8159-JURNAL%20.pdf>.
- [3] Sasmito, Wiro, Ginanjar. “Penerapan Metode Waterfall pada Desain Sistem Informasi Geografis Industri Kabupaten Tegal. Jurnal Informatika: Jurnal Pengembangan IT(JPIT). Vol.2, No.1. Januari 2017.
- [4] Sudharmono, Casaandra, Neil; Ayub, Mewati. “Penerapan Metode K-Means dan Cobweb terhadap Analisis Prestasi Akademik Mahasiswa yang mengikuti Kegiatan Kemahasiswaan”. Jurnal Teknik Informatika dan Sistem Informasi. Vol.1,No.2. Agustus 2015, Hal 103.

Isaias F.X.F.A. Felnditi, mahasiswa Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta.

Desi Arisandi, memperoleh gelar S.Kom. dari Universitas Tarumanagara. Kemudian memperoleh M.T.I. dari Universitas Indonesia. Saat ini aktif sebagai dosen tetap Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta.

Tri Sutrisno, memperoleh gelar S.Si dari Universitas Diponegoro. Kemudian memperoleh M.Sc. dari Universitas Gadjah Mada. Saat ini aktif sebagai dosen tetap Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta.