

## PENERAPAN METODE K – MEANS UNTUK KLASIFIKASI BIDANG PEKERJAAN ALUMNI

Diqy Fakhrun Shiddieq, S.T., M. Kom.<sup>1</sup>, Muhamad Rizal Fadillah<sup>2</sup>

<sup>1</sup>Program Studi Teknik Informatika, STMIK LPKIA BANDUNG

<sup>2</sup>Teknik Informatika, Program Studi Teknik Informatika, STMIK LPKIA BANDUNG

<sup>3</sup>STMIK LPKIA BANDUNG, Jln. Soekarno Hatta No. 456 Bandung 40266

<sup>1</sup>diqy@lpkia.ac.id, <sup>2</sup>muhamadrizalfadillah6@gmail.com

---

### Abstrak

Klasifikasi bidang pekerjaan alumni bertujuan untuk mengetahui informasi sudah sejauh mana keberhasilan pada setiap kompetensi keahlian dengan para lulusan yang sudah bekerja dengan posisi yang mereka dapatkan, apakah posisi yang mereka dapatkan sudah sesuai dengan kompetensi keahlian yang mereka miliki atau tidak sesuai. Untuk memanfaatkan data tersebut diperlukan sebuah metode untuk mengolah data alumni. Metode yang akan digunakan untuk mengolah data tersebut dengan metode *Data Mining*. *Data mining* dapat menggali dan mengumpulkan informasi yang berguna dari kumpulan data. Untuk mendapatkan klasifikasi bidang alumni diperlukan metode Algoritma *K-Means*. Sehingga data historis alumni dapat diklasifikasikan menjadi lulusan yang bekerja sangat sesuai dengan kompetensi keahlian, cukup sesuai dengan kompetensi keahlian dan tidak sesuai dengan kompetensi keahlian. Perancangan basis data untuk klasifikasi pekerjaan bidang alumni ini menggunakan suatu model data berorientasi objek. Dengan metode pemrograman berorientasi objek, produktivitas program dapat meningkat. Pemrograman yang akan dibentuk akan menggunakan bahasa pemrograman PHP dengan database MySQL.

**Kata kunci:** *Data mining, K – Means, klasifikasi, bidang pekerjaan alumni.*

---

### 1. PENDAHULUAN

Pendidikan merupakan aspek penting dalam menunjang karir seseorang untuk masa yang akan datang. Berdasarkan PP No 47 Tahun 2008 tentang wajib belajar menjelaskan bahwa penyelenggara wajib belajar pada jalur formal dilaksanakan minimal pada jenjang pendidikan dasar yang meliputi SD, MI, SMP, MTS, dan bentuk lain yang sederajat. Walaupun wajib belajar di Indonesia untuk skala nasional 9 tahun, di beberapa daerah yang sudah memiliki kesiapan dan kemampuan dapat diatur pada perda masing – masing untuk menjadikan wajib belajar menjadi 12 tahun. Saat peserta didik melanjutkan pendidikan tingkat menengah atas dan memutuskan untuk melanjutkan ke Sekolah Menengah Kejuruan, tujuan yang diharapkan oleh siswa yaitu memiliki kompetensi keahlian yang berkompeten sesuai dengan jurusan yang dipilih. Ketika sudah lulus diharapkan siswa memiliki kompetensi keahlian sehingga akan lebih mudah dalam mendapatkan pekerjaan.

Sekolah Menengah Kejuruan Negeri 6 Bandung merupakan sekolah yang sudah berdiri sejak tahun 1953. Berlokasi di Jalan Soekarno Hatta Komplek Riung Bandung Kecamatan Gede Bage Bandung. Setiap tahunnya berdasarkan data sebaran alumni, SMK Negeri 6 Bandung dapat menciptakan lulusan sebanyak kurang lebih 700 siswa untuk semua jurusan yang ada di sekolah. Namun dari data

lulusan tersebut pihak sekolah kesulitan mendapatkan informasi kesesuaian antara kompetensi keahlian dengan posisi atau bidang pekerjaan yang diisi oleh lulusan. Sehingga sekolah tidak mengetahui keberhasilan penerapan kurikulum yang sudah ada, apakah sasaran kurikulum yang diajarkan sudah berhasil dicapai atau tidak. Menurut Depdiknas tahun 2004 bahwa kurikulum SMK harus memperhatikan serta memberikan sesuatu yang bermakna, baik yang ideal maupun pragmatis sesuai dengan kebutuhan peserta didik.

Maka dari itu untuk mengolah data – data yang ada menjadi sebuah informasi yang dibutuhkan oleh pimpinan sekolah diperlukan suatu metode pengolahan data. Salah satu metode pengolahan data yaitu metode data mining. *Data mining* membahas penggalian atau pengumpulan informasi yang berguna dari kumpulan data. Untuk mengumpulkan informasi tersebut metode yang akan digunakan adalah Algoritma *K – Means*. Metode *K – Means* merupakan salah satu metode clustering berbasis partisi [1].

Berdasarkan latar belakang diatas maka identifikasi permasalahan yang didapat yaitu data alumni belum dimanfaatkan secara optimal untuk mendapatkan informasi kesesuaian kompetensi lulusan dengan bidang pekerjaan, sehingga pihak sekolah kesulitan dalam pengambilan keputusan dalam hal kurikulum.

Sesuai dengan identifikasi permasalahan yang sudah didapatkan agar pembahasan lebih terfokus dengan tujuan utama maka ruang lingkup yang akan dibahas diantaranya:

1. Data alumni yang akan dibahas hanya alumni yang bekerja dengan hasil sesuai dengan kompetensi keahlian, cukup sesuai dengan kompetensi keahlian dan tidak sesuai dengan kompetensi keahlian.
2. Data sebaran alumni yang akan diambil berdasarkan lulusan tahun 2017 – 2018 selama 2 tahun terakhir.
3. Hanya menggunakan beberapa atribut yaitu data siswa, jurusan, dan posisi atau bagian yang didapatkan alumni.

Tujuan perancangan dari penulisan tugas akhir ini yaitu untuk memanfaatkan data alumni secara optimal untuk mendapatkan informasi kesesuaian kompetensi lulusan dengan bidang pekerjaan.

## 2. LANDASAN TEORI

### 2.1 Data Mining

*Data mining* merupakan suatu istilah yang digunakan untuk menguraikan pengetahuan baru di dalam database. *Data mining* merupakan proses dengan menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi serta pengetahuan yang bermanfaat yang terkait dari berbagai database dalam skala besar [2].

### 2.2 K – Means

Metode K-means merupakan salah satu algoritma klustering dengan metode partisi (partitioning method) yang berbasis titik pusat (centroid) selain algoritma K-Medoids berbasis obyek. Algoritma K-Means dalam penerapannya memerlukan tiga parameter yang seluruhnya ditentukan pengguna yaitu jumlah cluster k, inialisasi klaster, dan jarak sistem. Biasanya K-Means dijalankan secara independen dengan inialisasi yang berbeda menghasilkan cluster akhir yang berbeda karena algoritma ini secara prinsip hanya mengekspansi data menuju lokal minimal [3].

Algoritma K-Means mendefinisikan centroid atau pusat cluster dari cluster menjadi rata – rata point dari cluster tersebut. Dalam penerapan algoritma K-Means, jika diberikan sekumpulan data  $X = \{x_1, x_2, \dots, x_n\}$  dimana  $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$  adalah system dalam ruang real  $R^n$ , maka algoritma K-Means akan menyusun partisi X dalam sejumlah k cluster (*a priori*). Setiap cluster memiliki titik tengah (*centroid*) yang merupakan nilai rata – rata (*mean*) dari data – data dalam cluster tersebut. Tahapan dari algoritma *K-Means* antara lain :

- a. Langkah 1 : Pilih jumlah cluster k

- b. Langkah 2 : Inialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberi nilai awal dengan angka-angka random.
- c. Langkah 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* yang dirumuskan sebagai berikut:

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

dimana :

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

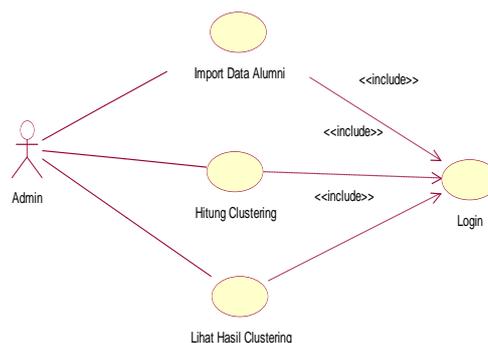
$X_{ki}$  = Data ke i pada atribut data ke k

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

- d. Langkah 4 : Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata dari semua data/ objek dalam cluster tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
- e. Langkah 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 [4].

## 3. GAMBAR SISTEM

### 3.1 Use Case Diagram



Gambar 1. Use Case Diagram

**3.2 Studi Kasus**

Dalam studi kasus ini, data yang digunakan yaitu data alumni SMK Negeri 6 Bandung lulusan tahun 2017 dan 2018. Langkah – langkah dalam penerapan algoritma k – means adalah sebagai berikut:

1. Persiapkan data yang akan dijadikan sebagai data training. Di sini data yang akan digunakan yaitu data alumni angkatan 2017 dan 2018. Berikut adalah data yang akan digunakan:

Tabel 1. Data Training

NO	NAMA	JURUSAN	BIDANG PEKERJAAN
1.	Adrian Sastradihardja	Teknik Gambar Bangunan	Karyawan
2.	Aldi Septian Putra Pratama Hasbi	Teknik Gambar Bangunan	Sipil
3.	Amalia Marinda	Teknik Gambar Bangunan	Kuliah
4.	Azat Syaputra	Teknik Gambar Bangunan	Karyawan
5.	Beben Abdul Rochim	Teknik Gambar Bangunan	Sipil
6.	Dandi Herdiansah	Teknik Gambar Bangunan	Sipil
7.	Diana Nabilah Nur Afifah	Teknik Gambar Bangunan	Karyawan
8.	Ega Kartika Mardiana	Teknik Gambar Bangunan	Sipil
9.	Erik Rizal Syahsani	Teknik Gambar Bangunan	Karyawan
10	Ghoirilman Arief Abdul Azis	Teknik Gambar Bangunan	Sipil
.....	.....	.....	.....
1537.	Warsono	Teknik Kendaraan Ringan	Mekanik

2. Berikutnya data harus di transformasikan ke dalam bentuk numeric agar data bisa diolah dengan algoritma K – Means. Penulis menginisialisasikan nama dengan id, jurusan dengan nomor 1 sampai 6, dan bidang kerja dari no 1 sampai 8. Berikut adalah transformasi dan keterangan dari data yang sudah dirubah ke numeric.

Tabel 2. Keterangan Jurusan

Keterangan Jurusan	
1	Teknik Audio Video
2	Teknik Gambar Bangunan
3	Teknik Instalasi Tenaga Listrik
4	Teknik Kendaraan Ringan
5	Teknik Kontruksi Kayu
6	Teknik Pemesinan

Tabel 3. Keterangan Bidang Pekerjaan

Keterangan Bidang Pekerjaan	
1	Karyawan
2	Sipil
3	Mekanik
4	Teknisi
5	Informatika
6	Kepolisian
7	Kuliah
8	Tidak Bekerja

Tabel 3. Data Yang Sudah Ditransformasikan

No	Nama / ID	Jurusan	Bidang Pekerjaan
1.	K1	2	1
2.	K2	2	2
3.	K3	2	7
4.	K4	2	1
5.	K5	2	2
6.	K6	2	2
7.	K7	2	1
8.	K8	2	2
9.	K9	2	1
10.	K10	2	2
.....	.....	.....	.....
1537.	K1537	4	3

3. Memilih pusat cluster awal. Karena kita akan membuat 3 kategori maka nilai pusat cluster yang akan dibentuk yaitu 3. Berikut nilai cluster awal yang terbentuk:

Tabel 3. Nilai Cluster Awal

Cluster	Jurusan	Bidang Pekerjaan
C0	2	1
C1	1	4
C2	5	7

4. Menghitung jarak antara data centroid dengan data. Centroid yang sudah terbentuk akan dihitung jarak tiap data ke tiap pusat cluster. Sebagai contoh untuk data dengan id k1 maka menghitung jarak ke pusat centroidnya yaitu :

$$C0 = \sqrt{(2 - 2)^2 + (1 - 1)^2} = 0$$

$$C1 = \sqrt{(2 - 1)^2 + (1 - 4)^2} = 3,162278$$

$$C2 = \sqrt{(2 - 5)^2 + (1 - 7)^2} = 6,708204$$

Terus lakukan perhitungan tersebut sampai data ke 1537 sehingga menghasilkan tabel seperti berikut.

Tabel 4. Hasil Jarak Antar Centroid

No	Data			C0	C1	C2
	Id	Kode Jurusan	Kode Bidang			
1	k1	2	1	0	3,162278	6,708204
2	k2	2	2	1	2,236068	5,830952
3	k3	2	7	6	3,162278	3
4	k4	2	1	0	3,162278	6,708204
5	k5	2	2	1	2,236068	5,830952
6	k6	2	2	1	2,236068	5,830952
7	k7	2	1	0	3,162278	6,708204

8	k8	2	2	1	2,236068	5,830952
9	k9	2	1	0	3,162278	6,708204
10	k10	2	2	1	2,236068	5,830952
.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....
1537	K1537	4	3	2,828427	3,162278	4,123106

- Setelah menghitung jarak antara data centroid dengan data, kelompokan data sesuai dengan cluster dengan mencari hasil nilai centroid dengan nilai yang paling kecil untuk dijadikan kelompok cluster 1 sampai dengan cluster 3. Maka di dapatkan hasil sebagai berikut:

Tabel 5. Hasil Pengelompokan Cluster

No	Data			C0	C1	C2	Cluster
	Id	Kode Jurusan	Kode Bidang				
1	k1	2	1	0	3,162278	6,708204	1
2	k2	2	2	1	2,236068	5,830952	1
3	k3	2	7	6	3,162278	3	3
4	k4	2	1	0	3,162278	6,708204	1
5	k5	2	2	1	2,236068	5,830952	1
6	k6	2	2	1	2,236068	5,830952	1
7	k7	2	1	0	3,162278	6,708204	1
8	k8	2	2	1	2,236068	5,830952	1
9	k9	2	1	0	3,162278	6,708204	1
10	k10	2	2	1	2,236068	5,830952	1
.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....
1537	K1537	4	3	2,828427	3,162278	4,123106	1

- Langkah selanjutnya yaitu mencari nilai untuk centroid berikutnya. Cara untuk mengetahui nilai centroid berikutnya dengan mencari rata – rata data sesuai dengan kelompok cluster yang sudah terbentuk. Untuk menghitung rata - rata dapat menggunakan rumus jumlah nilai seluruh data setiap cluster dibagi dengan jumlah data pada setiap kelompok cluster. Hasil yang didapatkan adalah sebagai berikut.

Tabel 5. Hasil Pengelompokan Cluster

	Cluster 0	Cluster 0	Cluster 1	Cluster 1	Cluster 2	Cluster 2
Jumlah Nilai Data	3254	1400	693	1431	1525	1627
Jumlah Data	898	898	335	335	304	304
Cluster berikutnya	3,623608	1,55902	2,068657	4,271642	5,016447	5,3511974

- Setelah mengetahui nilai centroid berikutnya maka ulangi langkah – 3 sampai dengan langkah – 5 sampai nilai cluster tetap dan tidak berubah – rubah lagi. Apabila nilai cluster sudah tetap maka proses clustering sudah selesai. Pada data di atas nilai cluster tidak berubah – ubah pada iterasi keempat dengan nilai cluster sebagai berikut.

Tabel 5. Hasil Akhir Nilai Cluster

	Cluster 0	Cluster 0	Cluster 1	Cluster 1	Cluster 2	Cluster 2
Jumlah Nilai Data	3254	1400	747	1708	1471	1350
Jumlah Data	898	898	367	367	272	2072
Cluster berikutnya	3,623608	1,55902	2,035422	4,653951	5,408088	4,963235

Dari hasil akhir nilai clustering dapat disimpulkan bahwa jumlah anggota cluster yang berada di cluster 0 sebanyak 887, cluster 2 sebanyak 299, cluster 2 sebanyak 203 dan cluster 3 sebanyak 148. Jika keanggotaan dari cluster tersebut di representasikan ke dalam bentuk presentase didapat hasil sebagai berikut:

- Cluster 0 = 58 %
- Cluster 1 = 24 %
- Cluater 2 = 18 %

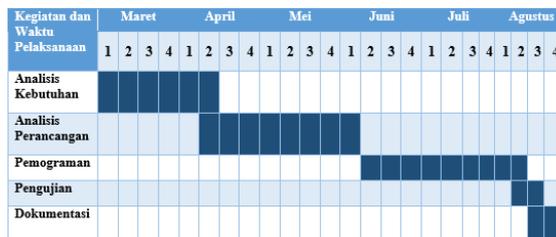
Maka dapat disimpulkan bahwa alumni yang bekerja sangat sesuai dengan bidang kompetensi keahlian sebesar 58 %, alumni yang bekerja cukup sesuai dengan bidang kompetensi keahlian sebesar 24 %, alumni yang bekerja tidak sesuai dengan kompetensi keahlian sebesar 18 %. Sehingga SMK Negeri 6 Bandung cukup berhasil menciptakan lulusan yang bekerja sesuai dengan kompetensi keahlian dengan hasil presentase yang cukup signifikan yaitu sebesar 58%.

#### 4. IMPLEMENTASI

##### 4.1 Implementasi Perangkat Lunak

Pada subbab ini akan dijelaskan langkah-langkah mengenai pembuatan, penggunaan dan jadwal pelaksanaan dalam implementasi rancangan perangkat lunak. Jadwal pelaksanaan digambarkan menggunakan *gantchart*.

Tabel 5. Gantt Chart



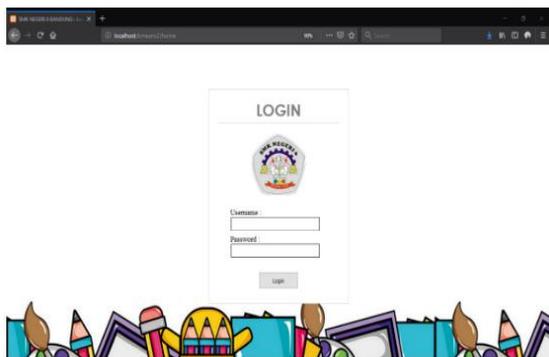
##### 4.2 Lingkup dan Batasan Implementasi

Uraian mengenai lingkup dan batasan implementasi termasuk modul – modul program yang akan diimplementasikan, fungsional system dan lain sebagainya termasuk lingkungan system.

Adapun lingkup dan batasan implementasi agar pada saat digunakan sistem dapat berjalan sesuai dengan modul, fungsional sistem maka lingkup dan batasan implementasi sebagai berikut :

1. Fungsi Login : Sebagai jalan untuk masuk kedalam sistem.
2. Fungsi data alumni : Melakukan import data alumni berbentuk file excel.
3. Fungsi proses clustering: Hasil dari proses pengelompokan data

**4.3 Implementasi Antarmuka**



Gambar 10. Antarmuka Login



Gambar 11. Antarmuka Import Data



Gambar 12. Antarmuka Hasil Clustering

Objek	Data 1	Data 2	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Objek 1	1	4	Null	Null	OK	Null
Objek 2	1	4	Null	Null	OK	Null
Objek 3	1	5	Null	Null	OK	Null
Objek 4	1	4	Null	Null	OK	Null
Objek 5	1	1	Null	OK	Null	Null
Objek 6	2	1	Null	Null	OK	Null
Objek 7	2	2	Null	OK	Null	Null
Objek 8	2	7	Null	Null	Null	OK
Objek 9	2	1	Null	Null	OK	Null
Objek 10	2	2	Null	OK	Null	Null
Objek 11	3	1	OK	Null	Null	Null
Objek 12	3	1	OK	Null	Null	Null
Objek 13	3	1	OK	Null	Null	Null
Objek 14	3	1	OK	Null	Null	Null
Objek 15	3	4	Null	Null	Null	OK
Objek 16	4	3	Null	Null	OK	Null
Objek 17	4	3	Null	Null	OK	Null
Objek 18	4	1	Null	OK	Null	Null
Objek 19	4	7	Null	Null	Null	OK
Objek 20	4	3	Null	Null	OK	Null
Objek 21	5	1	OK	Null	Null	Null
Objek 22	5	2	OK	Null	Null	Null
Objek 23	5	1	OK	Null	Null	Null
Objek 24	5	1	OK	Null	Null	Null
Objek 25	3	1	OK	Null	Null	Null

Gambar 13. Antarmuka Hasil Iterasi Cluster

**5. KESIMPULAN**

Dengan adanya sistem klasifikasi bidang pekerjaan alumni menggunakan data mining K- Means, sekolah dapat memanfaatkan data historis alumni menjadi optimal dan menghasilkan informasi presentase kesesuaian bidang pekerjaan alumni. Presentase yang didapatkan yaitu 58% untuk alumni yang bekerja sangat sesuai, 24% untuk alumni yang bekerja cukup sesuai dan 18% untuk alumni yang bekerja tidak sesuai dengan kompetensi keahlian. Sehingga dengan adanya presentase tersebut pihak sekolah dapat mengetahui kesesuaian bidang pekerjaan alumni dengan jurusan yang diambil dengan akurat serta memudahkan pihak sekolah dalam melakukan evaluasi baik dalam kurikulum pembelajaran maupun kerja sama dengan perusahaan.

**DAFTAR PUSTAKA**

- [1] S. Adinugroho and Y. A. Sari, Implementasi Data Mining Menggunakan Weka, Malang: UB Press, 2018.
- [2] D. Nofriansyah, Konsep Data Mining Vs Sistem Pendukung Keputusan, Yogyakarta: DEEPUBLISH, 2015.
- [3] E. Irwansyah and M. Faisal, Advanced Clustering: Teori dan Aplikasi, DeePublish, 2015.
- [4] J. O. Ong, "Implementasi Algoritma K-Mean Clustering Untuk Menentukan Strategi Marketing President University," *Jurnal Ilmiah Teknik Industri*, p. 13, 2013.