

# Data Mining Pengelompokan Bidang Keahlian Mahasiswa Menggunakan Algoritma K-Means (Studi Kasus : Universitas Cic Cirebon)

Chairun Nas

Manajemen Informatika, Universitas Catur Insan Cendekia Cirebon  
Jl. Kesambi No. 202, Kota Cirebon  
Email: chairun.nas@cic.ac.id

**Abstrak.** Pada proses perkuliahan, setiap mahasiswa mengikuti semua matakuliah yang telah disediakan oleh pihak kampus. Dari seluruh matakuliah yang ada, beberapa matakuliah membentuk keahlian mahasiswa, seperti keahlian programing, database atau networking. Namun dikarenakan mahasiswa harus mengikuti seluruh proses pembelajaran tersebut, maka mahasiswa sulit mengetahui keahlian apa yang perlu mereka fokuskan. Tujuan dari penelitian ini adalah untuk mengelompokkan mahasiswa, agar mahasiswa tahu keahlian apa yang perlu mereka tingkatkan atau kembangkan. Dalam penelitian ini dilakukan pengolahan data nilai-nilai matakuliah yang membentuk keahlian mahasiswa, diambil dari mahasiswa semester 5. Data akan di uji dengan menggunakan proses Data Mining dengan algoritma K-Means. Dari hasil pengujian yang telah dilakukan, diperoleh pengelompokan mahasiswa berdasarkan keahliannya dengan jumlah iterasi sebanyak 3 kali iterasi. Selanjutnya dari pengujian tersebut, dilakukan pengujian juga dengan menggunakan aplikasi Tanagra, dimana dihasilkan jumlah iterasi dan jumlah anggota kelompok dalam cluster adalah sama, dengan akurasi sebesar 80%. Maka penggunaan Datamining dengan algoritma K-Means mampu mengelompokkan mahasiswa berdasarkan keahliannya.

**Kata kunci:** *Data mining, Clustering, K-Means, Keahlian.*

## 1 Pendahuluan

Dalam proses pembelajaran di perguruan tinggi, mahasiswa dituntut mengikuti seluruh matakuliah yang telah ditawarkan kepada mahasiswa. Dari seluruh matakuliah yang dijalani, terdapat beberapa matakuliah yang membentuk keahlian mahasiswa baik itu secara teori maupun praktek. Namun dikarenakan mahasiswa harus mengikuti seluruh matakuliah tersebut, maka mahasiswa tidak dapat menentukan fokus bidang keahlian yang harus mereka kembangkan. Untuk itu, pihak kampus perlu mengelompokkan mahasiswa berdasarkan keahlian mahasiswa dari nilai-nilai matakuliah yang telah diperoleh mahasiswa, sehingga mahasiswa dapat menentukan fokus bidang keahlian mereka.

Untuk mengelompokkan mahasiswa dalam bidang keahliannya tersebut, maka perlu adanya proses perhitungan yang tepat, sehingga didapatkan hasil yang baik. Dengan tahapan yang jelas maka akan dihasilkan pengelompokan data yang baik. Dengan menggunakan Data mining dalam pengolahan data, maka dapat membantu dalam pengelompokan bidang keahlian mahasiswa. Data mining merupakan suatu proses penggalian data atau penyaringan data dengan memanfaatkan kumpulan data dengan ukuran yang cukup besar melalui serangkaian proses untuk mendapatkan informasi yang berharga dari data

tersebut [1]. *Data mining* adalah bagian dari KDD (*Knowledge Discovery in Database*) yang terdiri dari beberapa tahapan seperti pemilihan data, pra pengolahan data, transformasi, *Data mining*, dan evaluasi hasil [2]. Maka *Data mining* ditujukan dalam melakukan pengolahan data untuk menggali informasi yang berharga. Dalam proses *Data mining* sendiri, terdapat metode-metode yang digunakan dalam mengolah data.

Salah satu metode dalam *data mining* adalah *Clustering*. *Clustering* merupakan sebuah tahapan untuk memisahkan informasi menjadi beberapa bagian dengan tujuan bahwa informasi memiliki kriteria yang sama antara informasi yang satu dengan yang lainnya kedalam suatu tempat yang sama [3]. *Clustering* juga dapat diartikan sebuah metode analisis data yang bertujuan mengelompokkan data dengan karakteristik yang sama ke dalam suatu wilayah yang sama [4]. Salah satu pendekatan atau algoritma yang terdapat dalam *Clustering* yaitu *K-Means*. Algoritma *K-Means* merupakan algoritma pengelompokan iteratif yang melakukan partisi *set* data kedalam sejumlah *Kcluster* yang sudah ditetapkan diawal [1]. *K* dimaksudkan sebagai *konstanta* jumlah *cluster* yang diinginkan dan *means* berarti nilai suatu rata-rata dari suatu *cluster* [5].

Sebelumnya, penelitian *Data mining* dengan menggunakan algoritma *K-Means* telah dilakukan pada SMKN 1 Nguling untuk menentukan keahlian bagi siswa baru. Pada penelitian ini dilakukan pengujian pada 770 siswa baru yang akan dikelompokkan pada 3 program keahlian. Selanjutnya dilakukan proses algoritma *K-Means* dimana diperoleh hasil pada iterasi ke 3. Maka proses *Data mining* dengan algoritma *K-Means* telah dapat membantu dalam pengelompokan keahlian siswa baru [6].

Selain dari itu, algoritma *K-Means* juga dilakukan uji coba dalam penelitian mengelompokkan kunjungan wisatawan ke objek wisata unggulan di Provinsi DKI Jakarta. Dalam penelitian tersebut dilakukan pengujian pada 8 objek wisata unggulan dengan data kunjungan dari tahun 2007 sampai dengan 2013. Selanjutnya dikelompokkan menjadi 3 bagian yaitu Jumlah Kunjungan Tinggi, Jumlah Kunjungan Sedang dan Jumlah Kunjungan Rendah. Dari proses *K-Means* yang dilakukan maka diperoleh pengelompokan wisata unggulan dengan jumlah kunjungan tinggi 1, jumlah kunjungan sedang 1 dan selebihnya jumlah kunjungan rendah. Dengan hasil tersebut maka Provinsi DKI Jakarta dapat meningkatkan kualitas layanan pada objek wisata dengan kunjungan rendah [7].

Maka dari penelitian yang telah dibahas sebelumnya, *Data mining* dengan algoritma *K-Means* cocok digunakan dalam mengelompokkan bidang keahlian mahasiswa.

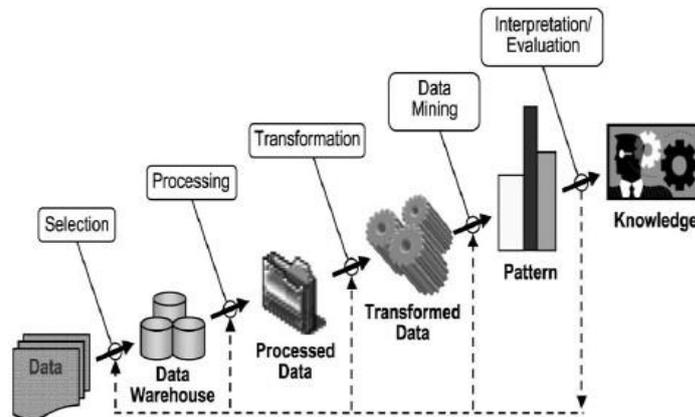
## **2 Dasar Teori**

### **2.1 Data Mining**

*Data mining* adalah suatu ekstraksi implisit dari sekumpulan data untuk dapat digali informasi yang bermanfaat didalamnya, dimana informasi tersebut tidak diketahui sebelumnya [8]. *Data mining* merupakan disiplin ilmu yang mempelajari metode-metode untuk mengekstrakan pengetahuan atau

menemukan pola dari suatu data [1]. Maka dapat disimpulkan bahwasanya *Data mining* bertujuan untuk menggali informasi berharga yang tersembunyi didalam sekumpulan data.

*Data mining* merupakan proses atau bagian dari KDD (*Knowledge Discovery in Databases*). KDD sendiri merupakan tahapan dalam penggalian data untuk mendapatkan informasi yang berharga. Adapun bentuk umum dari Proses KDD dapat dijelaskan pada Gambar 1 berikut [4]:



**Gambar 1.** Tahapan KDD (*Knowledge Discovery in Database*)

Adapun tahapan dari *Knowledge Discovery in Database* berdasarkan Gambar 1 dapat dijelaskan sebagai berikut [9]:

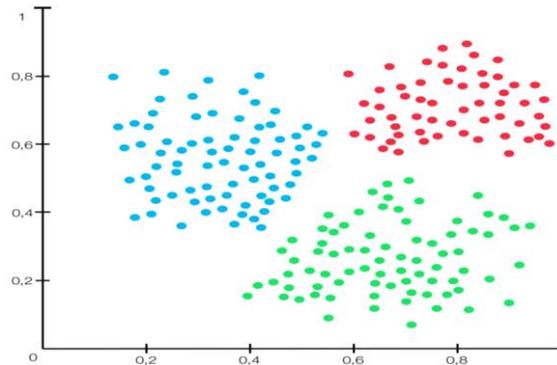
1. Membuat *set* data target yang menjadi tujuan utama dari perolehan informasi yang akan ditemukan.
2. Pembersihan data dari data *noise* dan tidak konsisten
3. Menggabungkan data dari berbagi sumber yang berbeda.
4. Memilih data dari database yang sesuai dengan tujuan analisis..
5. Melakukan transformasi atau mengubah bentuk data agar sesuai untuk proses penggalian data.
6. Proses Data mining dengan memilih algoritma atau metode tertentu untuk memperoleh sebuah pola atau suatu *set* representasi.
7. Mengidentifikasi pola yang diperoleh.
8. Mengkonsolidasikan pengetahuan yang diperoleh agar dapat digunakan oleh pihak yang memiliki kepentingan.

Dalam penelitian ini, proses KDD difokuskan kepada tahapan nomor 6 dan 7, sehingga diperoleh informasi dari data yang diolah.

## 2.2 Clustering

*Clustering* adalah penempatan seperangkat kombinasi objek yang memiliki kesamaan sedemikian rupa dimana objek tersebut memiliki keterkaitan satu sama lainnya dibandingkan kelompok penempatan lainnya [10]. Objek akan dikelompokkan ke dalam satu *cluster* sehingga objek-objek yang berada dalam satu *cluster* akan mempunyai kesamaan yang tinggi antara satu dengan yang lainnya [4].

Dengan melakukan pengelompokan terhadap data atau objek, maka dapat meminimalkan variasi data dalam suatu kelompok. Adapun *cluster* dapat digambarkan seperti pada Gambar 2 berikut [10]:



**Gambar 2.** Pengelompokan / Clustering

Pengelompokan dalam *clustering* dibagi kedalam wilayah-wilayah dimana objek satu dengan objek lainnya yang memiliki kesamaan diletakkan ke dalam wilayah yang sama.

### 2.3 Algoritma *K-Means*

Algoritma *K-Means* adalah sebuah algoritma dalam *Data mining* yang melakukan proses pemodelan tanpa pengawasan dan juga melakukan proses pengelompokan data dalam partisi, dimana data yang terkandung dalam suatu kelompok memiliki karakteristik yang sama dan berbeda dengan data dalam kelompok lain [11]. Algoritma *K-Means* merupakan algoritma yang membutuhkan parameter *input* sebanyak  $k$  dan membagi sekumpulan  $n$  objek kedalam  $k$  *cluster* sehingga tingkat kemiripan antar anggota dalam suatu *cluster* tinggi sedangkan tingkat kemiripan dengan anggota *cluster* lain sangat rendah [5]. Kemiripan anggota terhadap kelompoknya diukur dengan kedekatan objek terhadap nilai *mean* pada *cluster* atau dapat disebut sebagai *centroid cluster* atau pusat kelompok.

Untuk mengelompokan suatu data atau objek kedalam *cluster*, maka dapat mengikuti tahapan dalam algoritma *K-Means*. Adapun tahapan dalam proses algoritma *K-Means* adalah sebagai berikut [12]:

1. Tentukan banyak *cluster* yang ingin dibentuk, dengan nilai  $k$  adalah jumlah *cluster*.
2. Tentukan nilai *centroid* dari setiap *cluster* yang dibentuk. Nilai *centroid* ditentukan secara acak dari data dan jumlah *centroid* sama dengan jumlah kelompok.
3. Hitung jarak setiap data dengan masing-masing *centroid* dengan menggunakan rumus :

$$d(P, Q) = \sum_{j=1}^p \{x_j(P) - x_j(Q)\}^2 \quad \dots\dots\dots (1)$$

Keterangan :

$d$  = Jarak data ke *centroid*

$x$  = Nilai data

**P** = Data pada *record*

**Q** = Nilai *centroid*

4. Dari hasil perhitungan jarak data dengan *centroid*, maka dikelompokkan data berdasarkan pada jarak minimum.
5. Berdasarkan pengelompokan data tersebut, selanjutnya dicari *centroid* baru berdasarkan keanggotaan setiap *cluster* dengan menghitung rata-rata *cluster*. Untuk menentukan *centroid* baru dapat menggunakan rumus:

$$C(i) = \frac{x_1 + x_2 + \dots + x_n}{\sum x} \quad \dots\dots\dots (2)$$

Keterangan :

**x<sub>1</sub> ... x<sub>n</sub>** = Nilai data *record* 1 sampai n

**∑ x** = Jumlah data *record*

6. Selanjutnya lakukan langkah ke 3 kembali sampai diperoleh beberapa iterasi.

Iterasi akan berhenti apabila kelompok iterasi tidak memiliki perubahan pada iterasi sebelumnya.

## 2.4 Keahlian

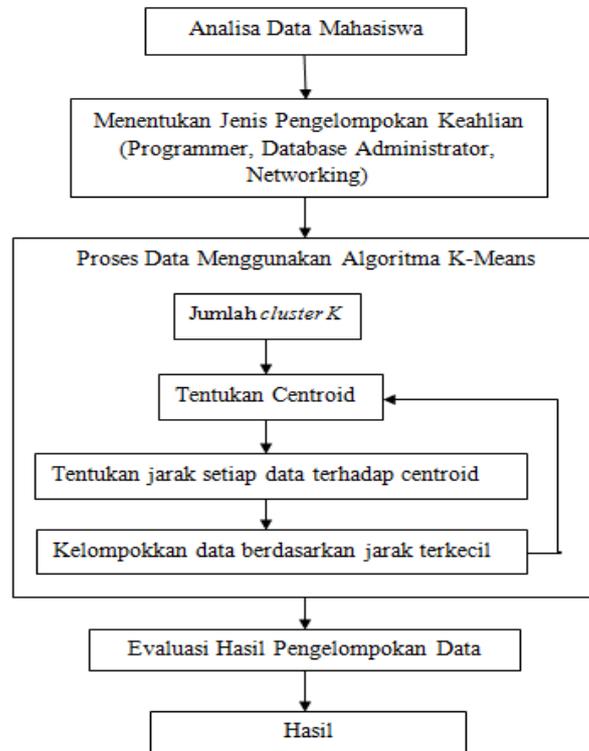
Keahlian merupakan suatu minat atau bakat yang harus dimiliki oleh seseorang. Dengan keahlian yang dimilikinya, memungkinkan untuk dapat menjalankan dan menyelesaikan tugas-tugas secara baik dengan hasil yang maksimal [13]. Keahlian seorang mahasiswa tentu dapat mereka peroleh dari pendidikan formal atau informal. Dengan adanya keahlian, maka akan terbuka peluang yang besar bagi mereka untuk mendapatkan pekerjaan.

Keahlian dapat dikelompokkan menjadi 2 bagian, yaitu [14]:

1. *Hard skills*, merupakan kemampuan teknis, kemampuan teknis diperlukan untuk dapat menyelesaikan masalah atau tugas yang berhubungan dengan kemampuan yang dimiliki untuk mencapai tujuan pekerjaan.
2. *Soft skills*, merupakan kemampuan karakteristik, kemampuan karakteristik berhubungan kepribadian seseorang, seperti sikap interpersonal, kemampuan komunikasi, kemampuan manajemen diri dan waktu, dan lain-lain.

## 3 Metodologi Penelitian

Pada penelitian ini, dapat dijelaskan langkah-langkah dalam proses pengelompokan mahasiswa berdasarkan keahlian bidangnya dengan menggunakan *Data mining*. Adapun prosesnya dapat dilihat pada Gambar 3 berikut:



**Gambar 3.**Tahapan Proses Penelitian

Dari Gambar 3 tersebut, maka dapat dijelaskan tahapan dari penelitian sebagai berikut:

### 3.1 Analisa Data Mahasiswa

Data yang digunakan dalam penelitian ini adalah data berupa nilai-nilai mahasiswa semester 5 berdasarkan matakuliah bidang keahliannya. Matakuliah bidang keahlian yang dijadikan sebagai penilaian dalam penelitian ini dapat dilihat pada Tabel 1 berikut:

**Tabel 1.**Matakuliah Bidang Keahlian

No.	Kode Matakuliah	Nama Matakuliah	Kelompok
1	MK1	Design Web (HTML + CSS)	Programmer
2	MK2	Pemograman Internet	
3	MK3	Framework	
4	MK4	Sistem Basis Data	Database Administrator
5	MK5	Perancangan Basis Data	
6	MK6	Oracle	
7	MK7	Jaringan Komputer	Networking
8	MK8	Management Server	
9	MK9	Networking Administrator	

Adapun dari data matakuliah bidang keahlian yang telah dikumpulkan, maka akan diambil sampel data dari mahasiswa semester 5 sebanyak 10 data sebagai data uji coba. Adapun data tersebut dapat dilihat pada Tabel 2 berikut:

**Tabel 2.**Data Sampel Uji Coba

No.	Mahasiswa	Nilai Matakuliah								
		MK1	MK2	MK3	MK4	MK5	MK6	MK7	MK8	MK9
1	Mhs A	89	85	83	80	85	79	85	83	85
2	Mhs B	87	85	84	78	89	80	89	84	85
3	Mhs C	78	89	82	87	80	83	80	82	89
4	Mhs D	79	80	79	88	75	85	75	79	80
5	Mhs E	80	83	79	82	84	83	84	79	83
6	Mhs F	77	80	85	79	84	76	84	85	80
7	Mhs G	89	85	82	85	79	83	79	82	85
8	Mhs H	79	78	78	78	76	79	80	78	78
9	Mhs I	75	78	82	83	83	83	83	82	78
10	Mhs J	82	80	81	80	85	80	85	81	80

Dari data yang telah diperoleh, selanjutnya data akan dilakukan pengolahan *Data mining*.

### 3.2 Pengelompokan Keahlian

Data yang akan diolah dengan algoritma *K-Means* akan dibagi menjadi 3 kelompok atau *cluster*, yaitu *cluster Programmer*, *cluster Database Administrator* dan *cluster Networking*. Pengelompokan akan diperoleh dari proses algoritma *K-Means* yang dilakukan. Akan dilakukan proses iterasi berkali-kali sampai tidak ada perubahan anggota kelompok dari iterasi terakhir dengan iterasi sebelumnya.

### 3.3 Proses Algoritma K-means

Dari data yang telah dikumpulkan sebelumnya, selanjutnya akan dilakukan proses perhitungan dengan algoritma *K-Means*. Algoritma *K-Means* sendiri memiliki langkah-langkah proses dalam penyelesaiannya sehingga sampai dalam tahap hasil. Hasil dari algoritma *K-Means* sendiri adalah dalam bentuk pengelompokan data. Selanjutnya dari pengelompokan data tersebut akan dilakukan evaluasi untuk mendapatkan keputusan.

### 3.4 Evaluasi Hasil Pengelompokan Data

Setelah diperoleh hasil pengelompokan data dari algoritma *K-Means*, selanjutnya dilakukan evaluasi terhadap hasil pengelompokan data tersebut. Evaluasi bertujuan untuk melihat apakah masih ada terdapat kesalahan dari hasil pengelompokan data dan apakah perlu dilakukan pengujian kembali.

### 3.5 Hasil

Hasil pengelompokan data adalah tujuan akhir dari proses penentuan keahlian mahasiswa. Hasil dari pengelompokan data tersebut akan menjadi sebagai alat bantu bagi pihak kampus dan mahasiswa sendiri untuk menentukan arah yang baik dalam pengembangan keahliannya. Sehingga mahasiswa dapat fokus keahlian apa yang perlu mereka tingkatkan.

#### 4 Hasil Dan Pembahasan

Dari metodologi penelitian yang telah dibahas sebelumnya, maka dapat dijelaskan proses algoritma *K-Means* dalam pengelompokan keahlian mahasiswa. Adapun untuk proses algoritma *K-Means* tersebut dapat dilihat dalam langkah-langkah berikut.

Langkah pertama adalah menentukan banyak *cluster* dalam pengelompokan keahlian mahasiswa. Untuk *cluster* data dapat dilihat pada Tabel 1 dimana *cluster* dalam pengujian ini adalah 3 ( $k = 3$ ) dan untuk data pengujiannya adalah dalam bentuk nilai-nilai mahasiswa pada matakuliah bidang keahlian yang tersedia pada Tabel 2. Data pada Tabel 2 tersebut ditransformasikan dengan dicari nilai rata-ratanya untuk setiap matakuliah bidang keahliannya seperti untuk Bidang Programmer, dicari nilai rata-rata dari MK1 + MK2 + MK3.

Setelah dilakukan transformasi data pada Tabel 2, maka dapat dilihat hasil transformasi data pada Tabel 3 berikut:

**Tabel 3.**Data Hasil Transformasi

No.	Mahasiswa	Programer	Database Administrator	Networking
1	Mhs A	85,67	81,33	84,33
2	Mhs B	85,33	82,33	86
3	Mhs C	83	83,33	83,67
4	Mhs D	79,33	82,67	78
5	Mhs E	80,67	83	82
6	Mhs F	80,67	79,67	83
7	Mhs G	85,33	82,33	82
8	Mhs H	78,33	77,67	78,67
9	Mhs I	78,33	83	81
10	Mhs J	81	81,67	82

Langkah kedua adalah menentukan nilai *centroid* untuk setiap *cluster*. Nilai *centroid* diperoleh secara acak dari nilai yang ada pada setiap data pengujian. Adapun nilai *centroid* yang diambil dapat dilihat pada Tabel 4 berikut:

**Tabel 4.**Nilai *Centroid* Awal Tiap *Cluster*

Nilai Centroid			
Centroid 1 (C1)	85,33	80,67	79,33
Centroid 2 (C2)	81,33	77,67	83
Centroid 3 (C3)	78,67	83,33	84,33

Langkah ketiga adalah menghitung jarak setiap data ke pusat cluster atau nilai *centroid*. Perhitungan jarak (*Eucliden*) dapat dilakukan dengan rumus (1) diatas. Maka berdasarkan rumus tersebut, maka dilakukan proses perhitungan *eucliden* sebagai sampel seperti berikut:

$$d(\text{Data 1, Centroid 1}) = \sqrt{(85,67 - 85,33)^2 + (81,33 - 80,67)^2 + (84,33 - 79,33)^2} \\ = \mathbf{4,185}$$

$$d(\text{Data 1, Centroid 2}) = \sqrt{(85,67 - 81,33)^2 + (81,33 - 77,67)^2 + (84,33 - 83)^2} \\ = \mathbf{8,123}$$

$$d(\text{Data 1, Centroid 3}) = \sqrt{(85,67 - 78,67)^2 + (81,33 - 83,33)^2 + (84,33 - 84,33)^2} = 7,681$$

Maka untuk perhitungan *euclidean* pada iterasi 1 dapat dilihat pada Tabel 5 berikut:

**Tabel 5.** Nilai *Euclidean* Pada Iterasi 1

Data Ke-i	N1	N2	N3	Centroid 1			Centroid 2			Centroid 3		
				85,33	80,67	79,33	82,67	83,33	77,67	81	78,67	86
Mhs A	85,67	81,33	84,33	4,185			8,123			7,681		
Mhs B	85,33	82,33	86	6,116			9,265			7,449		
Mhs C	83	83,33	83,67	5,516			6,237			4,457		
Mhs D	79,33	82,67	78	7,368			2,054			5,621		
Mhs E	80,67	83	82	6,079			4,380			2,746		
Mhs F	80,67	79,67	83	5,221			6,319			5,082		
Mhs G	85,33	82,33	82	3,282			5,933			7,080		
Mhs H	78,33	77,67	78,67	7,467			6,197			8,136		
Mhs I	78,33	83	81	7,911			4,482			2,704		
Mhs J	81	81,67	82	5,098			4,542			3,778		

Langkah keempat adalah mengelompokkan data berdasarkan nilai *euclidean* yang diperoleh. Data dikelompokkan berdasarkan jarak minimum atau jarak terdekat dengan *clusternya*. Dengan membandingkan nilai terkecil antara nilai *euclidean* pada masing-masing *centroid*, sehingga dapat diperoleh data tersebut masuk kedalam *cluster* apa. Maka pengelompokan dari perolehan data diatas dapat dilihat pada Tabel 6 berikut:

**Tabel 6.** Pengelompokan *Cluster* Data Pada Iterasi 1

Data Ke-i	Cluster 1	Cluster 2	Cluster 3
Mhs A	Cluster 1		
Mhs B	Cluster 1		
Mhs C			Cluster 3
Mhs D		Cluster 2	
Mhs E			Cluster 2
Mhs F			Cluster 3
Mhs G	Cluster 1		
Mhs H		Cluster 2	
Mhs I			Cluster 3
Mhs J			Cluster 3

Langkah kelima adalah menentukan nilai *centroid* baru untuk iterasi selanjutnya. Nilai *centroid* baru diperoleh dari nilai data yang telah dikelompokkan berdasarkan *clusternya*. Adapun penentuan nilai *centroid* baru dapat dilihat sebagai berikut:

Nilai Centroid 1 Baru

$$= \frac{85,67 + 85,33 + 85,33}{3}, \frac{81,33 + 82,33 + 82,33}{3}, \frac{84,33 + 86 + 82}{3}$$

$$= (85,4 \quad 82,0 \quad 84,1)$$

Nilai Centroi 2 Baru

$$= \frac{79,33 + 78,33}{2}, \frac{82,67 + 77,67}{2}, \frac{78 + 78,67}{2}$$

$$= (78,8 \quad 80 \quad 78,3)$$

Nilai Centroid 3 baru

$$= \frac{83 + 80,67 + 80,67 + 78,33 + 81}{5}, \frac{83,33 + 83 + 79,67 + 83 + 81,67}{5}, \frac{83,67 + 82 + 83 + 81 + 82}{5}$$

$$= (80,7 \quad 82 \quad 82,3)$$

Langkah selanjutnya adalah melakukan pengulangan pada langkah ketiga dengan menggunakan nilai *centroid* baru. Hitung kembali nilai *euclidean* untuk iterasi ke 2 seperti pada Tabel 7 berikut:

**Tabel 7.**Nilai *Euclidean* Pada Iterasi 2

Data Ke-i	N1	N2	N3	Centroid 1			Centroid 2			Centroid 3		
				85,44	82,0	84,1	78,8	80	78,3	80,7	82	82,3
Mhs A	85,67	81,33	84,33	0,738			9,169			5,385		
Mhs B	85,33	82,33	86	1,923			10,279			5,882		
Mhs C	83	83,33	83,67	2,818			7,472			2,890		
Mhs D	79,33	82,67	78	8,669			2,571			4,587		
Mhs E	80,67	83	82	5,314			4,983			0,930		
Mhs F	80,67	79,67	83	5,425			5,040			2,553		
Mhs G	85,33	82,33	82	2,139			7,768			4,612		
Mhs H	78,33	77,67	78,67	9,946			2,571			6,256		
Mhs I	78,33	83	81	7,828			3,919			2,882		
Mhs J	81	81,67	82	4,930			4,516			0,631		

Dari perhitungan *euclidean* diatas, maka diperoleh nilai *euclidean* untuk masing-masing data. Langkah selanjutnya adalah melakukan pengelompokan data seperti pada langkah keempat. Sama seperti sebelumnya, nilai terdekat akan dikelompokkan kedalam *clusternya*. Adapun pengelompokan *cluster* data untuk iterasi ke 2 dapat dilihat pada Tabel 8 berikut:

**Tabel 8.**Pengelompokan *Cluster* Data Pada Iterasi 2

Data Ke-i	Cluster 1	Cluster 2	Cluster 3
Mhs A	Cluster 1		
Mhs B	Cluster 1		
Mhs C	Cluster 1		
Mhs D		Cluster 2	
Mhs E			Cluster 3
Mhs F			Cluster 3
Mhs G	Cluster 1		
Mhs H		Cluster 3	
Mhs I			Cluster 2
Mhs J			Cluster 3

Dari pengelompokan *cluster* pada iterasi 2, maka ditemukan perbedaan anggota kelompok antara iterasi 1 dan iterasi 2. Karena ada perbedaan *cluster*, maka dilakukan proses iterasi ke 3. Untuk iterasi , tentukan kembali nilai *centroid* baru

seperti pada langkah-langkah sebelumnya. Adapun penentuan nilai *centroid* baru pada iterasi 3 dapat dilihat sebagai berikut:

$$\begin{aligned} & \text{Nilai Centroid 1 Baru} \\ &= \frac{85,67 + 85,33 + 83 + 85,33}{4}, \frac{81,33 + 82,33 + 83,33 + 82,33}{4}, \frac{84,33 + 86 + 83,67 + 82}{4} \\ &= (84,83 \quad 82,3 \quad 84) \end{aligned}$$

$$\begin{aligned} & \text{Nilai Centroid 2 Baru} \\ &= \frac{79,33 + 78,33}{2}, \frac{82,67 + 77,67}{2}, \frac{78 + 78,67}{2} \\ &= (78,8 \quad 80 \quad 78,3) \end{aligned}$$

$$\begin{aligned} & \text{Nilai Centroid 3 baru} \\ &= \frac{80,67 + 80,67 + 78,33 + 81}{4}, \frac{83 + 79,67 + 83 + 81,67}{4}, \frac{82 + 83 + 81 + 82}{4} \\ &= (80,2 \quad 82 \quad 82) \end{aligned}$$

Langkah selanjutnya adalah melakukan pengulangan pada langkah ketiga dengan menggunakan nilai *centroid* baru. Hitung kembali nilai *euclidean* untuk iterasi ke 3 seperti pada Tabel 9 berikut:

**Tabel 9.**Nilai *Euclidean* Pada Iterasi 3

Data Ke-i	N1	N2	N3	Centroid 1			Centroid 2			Centroid 3		
				84,83	82,3	84,0	78,8	80	78,3	80,2	82	82,0
Mhs A	85,67	81,33	84,33	1,345			9,169			5,997		
Mhs B	85,33	82,33	86	2,061			10,279			6,550		
Mhs C	83	83,33	83,67	2,114			7,472			3,612		
Mhs D	79,33	82,67	78	8,148			2,571			4,171		
Mhs E	80,67	83	82	4,666			4,983			1,269		
Mhs F	80,67	79,67	83	5,040			5,040			2,437		
Mhs G	85,33	82,33	82	2,061			7,768			5,186		
Mhs H	78,33	77,67	78,67	9,613			2,571			5,640		
Mhs I	78,33	83	81	7,192			3,919			2,395		
Mhs J	81	81,67	82	4,373			4,516			0,849		

Dari perhitungan *euclidean* diatas, maka diperoleh nilai *euclidean* untuk masing-masing data. Langkah selanjutnya adalah melakukan pengelompokan data seperti pada langkah keempat. Sama seperti sebelumnya, nilai terdekat akan dikelompokkan kedalam *clusternya*. Adapun pengelompokan *cluster* data untuk iterasi ke 3 dapat dilihat pada Tabel 10 berikut:

**Tabel 10.**Pengelompokan *Cluster* Data Pada Iterasi 3

Data Ke-i	Cluster 1	Cluster 2	Cluster 3
Mhs A	Cluster 1		
Mhs B	Cluster 1		
Mhs C	Cluster 1		
Mhs D		Cluster 2	
Mhs E			Cluster 3
Mhs F			Cluster 3

Mhs G	Cluster 1		
Mhs H		Cluster 2	
Mhs I			Cluster 3
Mhs J			Cluster 3

Berdasarkan pengelompokan data diatas, maka antara iterasi 2 dan iterasi 3 tidak ada perubahan anggota kelompok sehingga proses algoritma *K-Means* dihentikan. Dari pengelompokan data tersebut didapatkan hasil untuk *cluster 1* sebanyak 4 data, *cluster 2* sebanyak 2 data dan *cluster 3* sebanyak 4 data. Dari proses clustering diatas, maka dapat ditentukan bidang keahlian mahasiswa tersebut berdasarkan pengelompokan anggota diatas.

Dari perhitungan manual diatas yang dilakukan, perlu dilakukan pengujian dengan menggunakan aplikasi pengujian datamining, yaitu Tanagra. Dari hasil pengujian tanagra, maka diperoleh kelompok anggota seperti pada Gambar 4 Berikut :

	Mahasiswa	Programer	Database Adm	Networking	std_Program	std_Database	std_Networkin	Cluster_KMean
1	Mhs A	85,67	81,33	84,33	1,35507	-0,208759	0,930246	c_kmeans_1
2	Mhs B	85,33	82,33	86	1,23705	0,355462	1,61673	c_kmeans_1
3	Mhs C	83	83,33	83,67	0,428318	0,919682	0,65894	c_kmeans_1
4	Mhs D	79,33	82,67	78	-0,845526	0,547295	-1,67181	c_kmeans_3
5	Mhs E	80,67	83	82	-0,380418	0,733488	-0,0275421	c_kmeans_3
6	Mhs F	80,67	79,67	83	-0,380418	-1,14537	0,383526	c_kmeans_2
7	Mhs G	85,33	82,33	82	1,23705	0,355462	-0,0275421	c_kmeans_1
8	Mhs H	78,33	77,67	78,67	-1,19262	-2,27381	-1,3964	c_kmeans_2
9	Mhs I	78,33	83	81	-1,19262	0,733488	-0,43861	c_kmeans_3
10	Mhs J	81	81,67	82	-0,265876	-0,0169259	-0,0275421	c_kmeans_3

Gambar 4. Pengujian Dengan Menggunakan Tanagra

Cluster size and WSS			
Clusters	3		
Cluster	Description	Size	WSS
cluster n°1	c_kmeans_1	4	2,5740
cluster n°2	c_kmeans_2	2	2,5506
cluster n°3	c_kmeans_3	4	2,7461
R-Square for each attempt			
Number of trials	6		
Trial	R-square		
1	0,633129		
2	0,708493		
3	0,708493		

Gambar 5. Jumlah Anggota Setiap Cluster dan Jumlah Iterasi.

Dari hasil yang diperoleh menggunakan Tanagra, diperoleh jumlah anggota kelompok dalam setiap *cluster* dan jumlah banyak iterasi berjumlah sama dengan perhitungan manual. Namun terdapat perbedaan pengelompokan pada Mhs D dan Mhs F. Maka jika dipersentasikan, akurasi perhitungan manual dengan pengujian menggunakan aplikasi Tanagra adalah sebesar 80%.

## 5 Kesimpulan

Berdasarkan pembahasan yang telah dijelaskan, maka dapat ditarik kesimpulan dan saran dari penelitian ini sebagai berikut:

### 5.1 Simpulan

Algoritma *K-Means* telah dapat melakukan proses pengelompokan atau *clustering* terhadap data-data mahasiswa untuk menentukan bidang keahliannya berdasarkan matakuliah bidang keahlian. Hasil proses algoritma *K-Means* ini dari 10 data pengujian untuk 3 pengelompokan data, diperoleh 4 data mahasiswa cocok untuk dikelompokkan dibidang *Programmer*. Selanjutnya 2 data mahasiswa cocok untuk dikelompokkan dalam bidang *Database Administrator* dan 4 data mahasiswa dikelompokkan dalam bidang *Networking*. Berdasarkan pengujian dengan menggunakan aplikasi Tanagra, diperoleh jumlah iterasi dan jumlah anggota setiap *cluster* adalah sama, dengan tingkat akurasi sebesar 80%. Perolehan hasil sangat dipengaruhi oleh nilai *centroid* awal, sehingga apabila menggunakan data yang banyak, perlu menentukan nilai *centroid* yang pas. Hasil dari pengelompokan data tersebut selanjutnya dapat menjadi keputusan bagi mahasiswa atau pihak kampus dalam mengembangkan bidang keahlian mahasiswa.

### 5.2 Saran

Untuk penelitian selanjutnya, dibutuhkan kriteria bidang matakuliah yang lebih luas lagi dengan pengelompokan data yang lebih banyak lagi, sehingga adanya proses yang mendalam untuk mendapatkan hasil keputusan yang baik. Selanjutnya algoritma *K-Means* dapat dikembangkan dengan menggunakan algoritma-algoritma lainnya sehingga hasil yang diperoleh dapat lebih kuat lagi.

## 6 Referensi

- [1] H. Sulastri and A. I. Gufroni, "Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia," *Jurnal Nasional Teknologi Dan Sistem Informasi*, vol. 03 No. 02, pp. 299–305, 2017, doi: <https://doi.org/10.25077/TEKNOSI.v3i2.2017.299-305>.
- [2] A. E. Wicaksono, "Implementasi Data Mining Dalam Pengelompokan Data Peserta Didik Di Sekolah Untuk Memprediksi Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means (Studi Kasus SMAN 16 Bekasi)," *Jurnal Teknologi Rekayasa*, vol. 21 No. 03, pp. 206–216, 2016.
- [3] S. Bhardwaj, "Data Mining Clustering Techniques -A Review," *IJCSMC*, vol. 6 No. 05, pp. 183–186, 2017, ISSN: 2320-088X.
- [4] Y. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *Jurnal Media Infotama*, vol. 12 No. 02, pp. 148–157, 2016, ISSN: 1858-2680.
- [5] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Islam Indragiri)," *Jurnal SISTEMASI*, vol. 7 No. 03, pp. 238–249, 2018, ISSN: 2302-8149.
- [6] E. W. Astuti, "Clustering Program Keahlian Pada Pendaftaran Siswa Baru (PSB) Dengan Menggunakan Algoritma K-Means: Studi Kasus Di SMKN

- 1 Nguling,” *Jurnal SPIRIT.*, vol. 7 No. 02, pp. 58–65, 2015, ISSN: 2085-3092.
- [7] L. Maulida, “Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. DKI Jakarta Dengan K-Means,” *Jurnal Informatika Sunan Kalijaga.*, vol. 2 No. 03, pp. 167–174, 2018, ISSN: 2527-5836.
- [8] K. Ahmed and T. Jesmin, “Comparative Analysis of Data Mining Classification Algorithms in Type-2 Diabetes Prediction Data Using WEKA Approach,” *IJSE.*, vol. 7 No. 02, pp. 155–160, 2014, ISSN: 2086-5023.
- [9] N. Rikhi, “Data Mining and Knowledge Discovery in Database,” *IJETT.*, vol. 23 No. 02, pp. 64–70, 2015, ISSN: 2231-5381.
- [10] S. Manikandan, A. L. Caroline and D. Kanniamma, “The Study on Clustering Analysis in Data Mining,” *International Journal of Data Mining Techniques and Applications.*, vol. 07 No. 01, pp. 46–49, 2018, ISSN: 2278-2419.
- [11] T. Widiyaningtyas, M. I. W. Prabowo and M. A. M. Pratama, “Implementation of K-Means Clustering to Distribution of High School Teachers,” *Procciding EECSI.*, Yogyakarta, 19-21 September 2017, pp. 49-54.
- [12] W. Purba, S. Tamba and J. Saragih, “The Effect of Mining Data K-Means Clustering Toward Students Profile Model Drop Out Potential,” *IOP Conf. Series: Journal of Physics.*, pp. 1–6, 2018, doi :10.1088/1742-6596/1007/1/012049.
- [13] M. Yona, “Pengaruh Penempatan Kerja, Keahlian dan Kepuasan Kerja Terhadap Peningkatan Kinerja Karyawan Pada PT. OSI Electronics Batam,” *Jurnal BENING.*, vol. 05 No. 01, pp. 27–38, 2018, ISSN: 2252-52672.
- [14] M. U. Manara, “Hard Skills dan Soft Skills Pada Bagian Sumber Daya Manusia di Organisasi Industri,” *Jurnal Psikologi Tabularasa.*, vol. 09 No. 01, pp. 37–47, 2014.