



# Sistem Informasi Pemetaan Pendidikan Menggunakan Algoritma Data Mining

Olha Musa<sup>a\*</sup>, Suhartono<sup>b</sup>

<sup>a</sup> STMIK Ichsan Gorontalo

<sup>b</sup> Jurusan Informatika, Fakultas Sains dan Matematika, Universitas Diponegoro

*Naskah Diterima : 3 Januari 2015; Diterima Publikasi : 16 Maret 2015*

---

## Abstract

in this study to identify the increase in educational services based on the quality of non-formal education is an indicator, having tiered in terms of education, non-formal education (training) to be one of the prerequisites in multiplying the potential for self-development. Data mining algorithms is a basic k-means clustering to put the object based on the average (Means) nearest cluster. Aims to design mapping information system education with the k-means cluster. Application k-means cluster is part of a non-hierarchical method, the mapping system of education in 171 samples of data Islam Students Association (HMI) were tested in this study showed that the non-hierarchical method (k-means cluster) has a good degree of accuracy because they specify the number of clusters in advance. Education information system mapping is used to cluster the data level, corresponding formal education and training has been followed. Education information system mapping is used to cluster the data level, corresponding formal education and training has been followed. The test results have in me some real, the spread of the data in each cluster are similar. At the time of the iteration process is not visible difference in the results of the mapping study using the k-means cluster. Results of a cluster centroid information models with variable 4 educated members include S1, S2, has entered basic training cluster 0, educated S1, S2, S3 has entered basic training cluster 1, S1 has educated basic training and training of incoming intermediate cluster 2, educated S1 has entered basic training cluster 3. formal education, education tiered seen in cluster 1 for non-formal education (training) tiered education seen in cluster 2. Based the test results k-means cluster.

**Keywords:** Information systems; Educational mapping; Cluster; K-means

## Abstrak

in this study to identify the increase in educational services based on the quality of non-formal education is an indicator, having tiered in terms of education, non-formal education (training) to be one of the prerequisites in multiplying the potential for self-development. Data mining algorithms is a basic k-means clustering to put the object based on the average (Means) nearest cluster. Aims to design mapping information system education with the k-means cluster. Application k-means cluster is part of a non-hierarchical method, the mapping system of education in 171 samples of data Islam Students Association (HMI) were tested in this study showed that the non-hierarchical method (k-means cluster) has a good degree of accuracy because they specify the number of clusters in advance. Education information system mapping is used to cluster the data level, corresponding formal education and training has been followed. Education information system mapping is used to cluster the data level, corresponding formal education and training has been followed. The test results have in me some real, the spread of the data in each cluster are similar. At the time of the iteration process is not visible difference in the results of the mapping study using the k-means cluster. Results of a cluster centroid information models with variable 4 educated members include S1, S2, has entered basic training cluster 0, educated S1, S2, S3 has entered basic training cluster 1, S1 has educated basic training and training of incoming intermediate cluster 2, educated S1 has entered basic training cluster 3. formal education, education tiered seen in cluster 1 for non-formal education (training) tiered education seen in cluster 2. Based the test results k-means cluster.

**Keywords :** Information Systems , Educational Mapping , Cluster , K - Means

---

## 1. Pendahuluan

Sistem informasi pemetaan pendidikan sebagai dasar meningkatkan layanan pendidikan. Pemetaan pendidikan salah satu penyuplai informasi yang berguna bagi pembuatan keputusan. Gambaran nyata

dari suatu kondisi di wilayah tertentu, menjadi tolak ukur pembangunan pendidikan. Secara jangka panjang, kebijakan yang dihasilkan ditujukan untuk meningkatkan ketersediaan, keterjangkauan, mutu, relevansi, kesetaraan, dan kepastian dalam

---

\*) Penulis korespondensi: [0lh4mu54@gmail.com](mailto:0lh4mu54@gmail.com)

memperoleh layanan pendidikan di Indonesia (Priadi, 2012).

Data mining suatu proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa sekumpulan data besar yang tersimpan dalam penyimpanan dengan menggunakan teknik statistik dan matematika (Larose, 2005)

Data mining dibagi menjadi 6 kelompok salah satunya yaitu *cluster* yang mengelompokkan record, pengamatan dan membentuk kelas obyek-obyek yang memiliki kemiripan. *cluster* adalah kumpulan record yang mempunyai kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan record dalam *cluster* yang lain. Berbeda dengan klasifikasi, pada *cluster* tidak ada variabel target. Pengklasteran tidak melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target, akan tetapi, algoritma pengklasteran melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan, yang mana kemiripan record dalam suatu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record lain akan bernilai minimal (Larose, 2005).

Metode *K-Means* adalah data *clustering* yang digolongkan sebagai metode pengklasifikasian yang bersifat *unsupervised* (tanpa arahan). Pengkategorian metode-metode pengklasifikasian data antara *supervised* dan *unsupervised classification* di dasarkan pada adanya *dataset* yang data itemnya sudah sejak awal mempunyai label kelas dan *dataset* yang data itemnya tidak mempunyai label kelas. Untuk data yang sudah mempunyai label kelas, metode pengklasifikasian yang digunakan merupakan metode *supervised classification* (Agusta, 2007).

*K-Means* algoritma untuk pembentukan *cluster* yang menggunakan metode *partitional clustering*. *K-Means* algoritma *clustering* yang banyak digunakan karena kemudahannya untuk di implementasikan. Ide utama dari algoritma *k-means* yaitu bahwa suatu titik tengah dapat merepresentasikan suatu *cluster*. Titik tengah yang dimaksud adalah titik tengah atau rata-rata dari suatu kumpulan titik, yang biasa disebut *centroid*. Algoritma *K-Means* mengelompokkan  $n$  objek ke dalam  $k$  *cluster* dimana objek-objek dalam satu *cluster* memiliki tingkat kemiripan yang tinggi, sedangkan tingkat kemiripan antara tiap *cluster* rendah (Han dan Kamber, 2006).

Metode pengelompokan dalam analisis *cluster* meliputi metode hirarkis, memulai pengelompokan dengan dua atau lebih obyek yang mempunyai kesamaan paling dekat. Kemudian diteruskan pada obyek yang lain dan seterusnya hingga *cluster* akan membentuk semacam 'pohon' dimana terdapat tingkatan (hirarki) yang jelas antar obyek, dari yang paling mirip hingga yang paling tidak mirip. Alat yang membantu untuk memperjelas proses hirarki ini disebut "dendrogram". Metode non-hirarkis, dimulai dengan menentukan terlebih dahulu jumlah *cluster*

yang diinginkan (dua, tiga, atau yang lain). Setelah jumlah *cluster* ditentukan, maka proses *cluster* dilakukan dengan tanpa mengikuti proses hirarki. Metode ini biasa disebut "*K Means Cluster*" (Dunham, 2003). Tujuan penelitian ini adalah merancang sistem informasi pemetaan pendidikan menggunakan analisis *K-Means Cluster*, studi kasus Himpunan Mahasiswa Islam (HMI).

## 2. Kerangka Teori

### 2.1. Cluster

Teknik *data mining* merupakan sebuah proses ekstraksi informasi untuk menggali pengetahuan (*knowledge discovery*) dan menemukan pola (*pattern recognition*) pada tumpukan data dalam *database* yang biasanya berskala besar. Fungsi-fungsi dalam *data mining* antara lain: fungsi deskripsi, fungsi estimasi, fungsi prediksi, fungsi klasifikasi, fungsi *clustering* dan fungsi asosiasi (Larose, 2005).

Proses pengelompokan sekumpulan obyek kedalam kelas-kelas obyek yang sama disebut *clustering*/pengelompokan. Pengklasteran merupakan satu dari sekian banyak fungsi proses data mining untuk menemukan kelompok atau identifikasi kelompok obyek yang hampir sama. Analisis *cluster* (*clustering*) merupakan usaha untuk mengidentifikasi kelompok obyek yang mirip-mirip dan membantu menemukan pola penyebaran dan pola hubungan dalam sekumpulan data yang besar. Hal penting dalam proses pengklasteran adalah menyatakan sekumpulan pola ke kelompok yang sesuai yang berguna untuk menemukan kesamaan dan perbedaan sehingga dapat menghasilkan kesimpulan yang berharga (Giyanto, 2008).

*Clustering* digunakan untuk menganalisis pengelompokan berbeda terhadap data, mirip dengan klasifikasi, namun pengelompokan belum didefinisikan sebelum dijalankannya *tool* data mining. *Clustering* membagi item menjadi kelompok-kelompok berdasarkan yang ditemukan *tool* data mining. Prinsip dari *clustering* adalah memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar *cluster*. *Clustering* dapat dilakukan pada data yang memiliki beberapa atribut yang dipetakan sebagai ruang multidimensi.

### 2.2. Pengukuran Kualitas Cluster

Untuk mengetahui kualitas hasil dari suatu *clustering*, ada dua cara yang dapat digunakan yaitu dengan pengukuran kualitas internal (*internal quality measure*) dan pengukuran kualitas eksternal (*external quality measure*). *Internal quality measure* dilakukan dengan membandingkan *cluster-cluster* yang berbeda tanpa referensi dari pengetahuan luar (informasi di luar sistem). Sedangkan *external quality measure* dilakukan dengan cara membandingkan kelompok-kelompok yang dihasilkan dari *clustering* dengan

suatu label kelas yang telah diketahui (manual oleh manusia).

### 2.3. Algoritma K-Means

Algoritma k-means menggunakan pendekatan “top-down” dari awal sudah ditetapkan jumlah cluster kemudian dilakukan pengamatan (Myatt, 2007). Pendekatan ini menghitung dengan cepat dan dapat menangani jumlah data yang besar. Dalam algoritma pengelompokan, sebuah data dikatakan serupa dengan mengukur jarak nilai dari satu data dengan data lain (Myatt, 2007). Proses pengukuran jarak antara dua objek ini dinamakan *Euclidean distance* atau jarak Euclidean dengan persamaan sebagai berikut.

Konsep dasar dari algoritma *k-means* adalah pencarian pusat *cluster* (*centroid points*) secara iteratif. Pusat *cluster* ditetapkan berdasarkan jarak setiap data ke pusat *cluster*. Proses *clustering* dimulai dengan mengidentifikasi data yang akan di-*cluster*,  $x_{ij}$  ( $i=1, \dots, n$ ;  $j=1, \dots, m$ ) dengan  $n$  adalah jumlah data yang akan di-*cluster* dan  $m$  adalah jumlah variabel. Pada awal iterasi, pusat setiap *cluster* ditetapkan secara bebas (sembarang),  $c_{kj}$  ( $k=1, \dots, k$ ;  $j=1, \dots, m$ ). Kemudian dihitung jarak antara setiap data dengan setiap pusat *cluster*. Untuk melakukan penghitungan jarak data ke- $i$  ( $x_i$ ) pada pusat *cluster* ke- $k$  ( $c_k$ ), diberi nama ( $d_{ik}$ ), dapat digunakan formula Euclidean, yaitu:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (2.1)$$

Suatu data akan menjadi anggota dari *cluster* ke- $j$  apabila jarak data tersebut ke pusat *cluster* ke- $j$  bernilai paling kecil jika dibandingkan dengan jarak ke pusat *cluster* lainnya.

$$c_{kj} = \frac{\sum_{h=1}^p y_{hj}}{p}; y_{hj} = x_{ij} \in \text{cluster ke } k \quad (2.2)$$

Alternatif penerapan metode *k-means* yaitu *distance space* telah diimplementasikan dalam menghitung jarak (*distance*) antara data dan centroid termasuk di antara dua titiknya L1 (Manhattan/City Block) *distance space*, L2 (Euclidean) *distance space*, dan Lp (Minkowski) *distance space*.

$$D_{L1}(X_2, X_1) = \|X_2 - X_1\|_1 = \sum_{j=1}^p |x_{2j} - x_{1j}| \quad (2.3)$$

Dimana:

$P$  : Dimensi data  
 $|\cdot|$  : Nilai absolute

Sedangkan untuk L2 (*Euclidean distance space*), Jarak antara dua titik dihitung dengan menggunakan rumus sebagai berikut:

$$D_{L2}(X_2, X_1) = \|X_2 - X_1\|_2 = \sqrt{\sum_{j=1}^p (x_{2j} - x_{1j})^2} \quad (2.4)$$

Algoritma *K-Means* adalah metode *clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster* dan algoritma ini hanya bekerja pada atribut numerik.

Algoritma ini pada awalnya mengambil sebagian dari banyaknya komponen dari populasi yang akan dijadikan sebagai pusat *cluster* awal. Pada step ini pusat *cluster* dipilih secara acak dari sekumpulan populasi data (Huang, 2005).

Penentuan berapa *cluster* yang akan dibentuk. Obyek atau elemen pertama dalam *cluster* dapat dipilih untuk dijadikan sebagai titik tengah (*centroid point*) *cluster*. Algoritma K-Means selanjutnya akan melakukan pengulangan langkah-langkah berikut:

1. menentukan koordinat titik tengah setiap *cluster*,
2. menentukan jarak setiap obyek terhadap koordinat titik tengah,
3. Hitung pusat kelompok (sentroid/rata-rata) dari data yang ada di masing masing kelompok
4. mengelompokkan obyek-obyek tersebut berdasarkan pada jarak minimumnya.
5. Kembali kelangkah 3, apabila masih ada data yang berpindah kelompok, atau ada perubahan nilai sentroid diatas nilai ambang yang ditentukan.

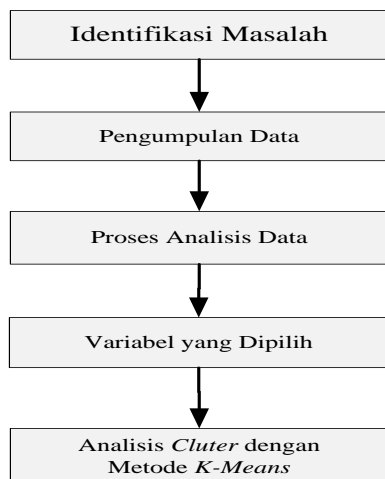
### 2.4. Rapid Miner

RapidMiner (YALE) adalah perangkat lunak open source untuk knowledge discovery dan data mining. RapidMiner memiliki kurang lebih 400 prosedur (operator) data mining, termasuk operator untuk masukan, output, data preprocessing (Dominikus dan Feri, 2010).

## 3. Metodologi

Penelitian sistem informasi pemetaan pendidikan terdapat tahapan-tahapan penelitian yang mengacu pada metodologi *k-means* dan analisis *cluster*. Tahapan pemetaan pendidikan dapat dilihat pada prosedur penelitian terdapat pada Gambar 1.

Data penelitian yang digunakan adalah kasus data organisasi pada Himpunan Mahasiswa Islam. Analisis data menggunakan *clustering* dengan metode *k-means*.



Gambar 1. Prosedur penelitian sistem pemetaan pendidikan dengan metode K-Means

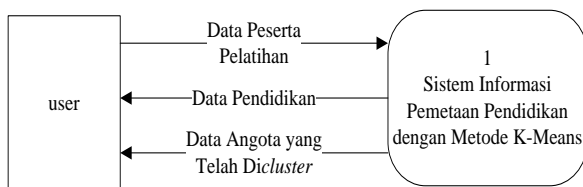
### 3.1. Pendidikan Formal HMI

Training formal berjenjang yang diikuti oleh anggota, dan setiap jenjang merupakan prasyarat untuk mengikuti jenjang selanjutnya. Latihan ini berfungsi memberikan kemampuan tertentu kepada para pesertanya sesuai dengan tujuan dan target pada masing masing jenjang latihan. Latihan kader terdiri dan 3 (tiga) jenjang, yaitu:

- a. Basic Training/ Pelatihan dasar (latihan Kader 1)
- b. Intermediate Training/Pelatihan Menengah (latihan Kader II)
- c. Advance Training/Pelatihan Lanjutan (latihan Kader III)

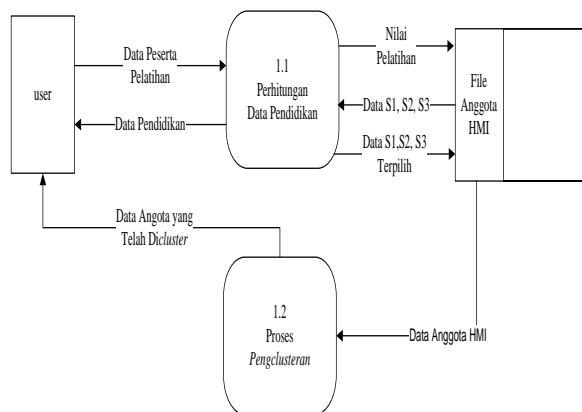
### 3.2. Analisis Perancangan Sistem Pemetaan Pendidikan

Analisis perancangan sistem pada penelitian ini digambarkan dengan menggunakan Data Flow Diagram (DFD). Data Flow Diagram merupakan sebuah teknik grafis yang menggambarkan bagaimana data diproses melalui sistem dalam bentuk input dan output. DFD dimulai dengan menggambarkan keseluruhan sistem dan dilanjutkan dengan merinci masing-masing area fungsi (Gambar 2).



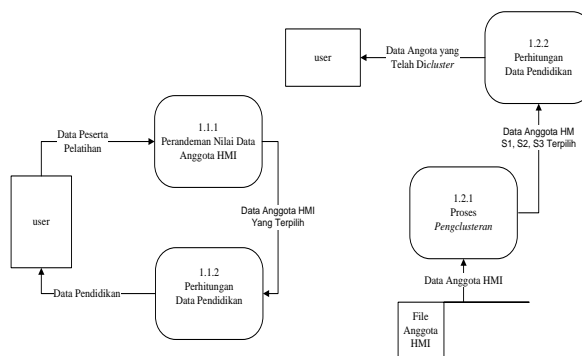
Gambar 2. Diagram Konteks Clustering (DFD level 0)

DFD level 0 dikembangkan lebih dalam lagi untuk memperjelas dan menambah detail aliran data dan proses dalam sistem informasi. Dikembangkan dan menghasilkan DFD level 1, seperti yang tertera pada Gambar 3.



Gambar 3. Teknik Clustering (DFD level 1)

Pada proses 1.1 DFD level 1 diperinci lagi menjadi level 2. Pada DFD level 2 terdapat 4 proses, mencari secara random nilai data anggota HMI, proses perhitungan data pendidikan anggota HMI, proses perandeman data S1, S2, S3, pengclustering data LK1, LK2, LK3, LKK. Output yang dihasilkan ada 4 cluster. Adapun DFD level 2 untuk proses clustering terlihat pada Gambar 4.



Gambar 4. Teknik Clustering (DFD level 2)

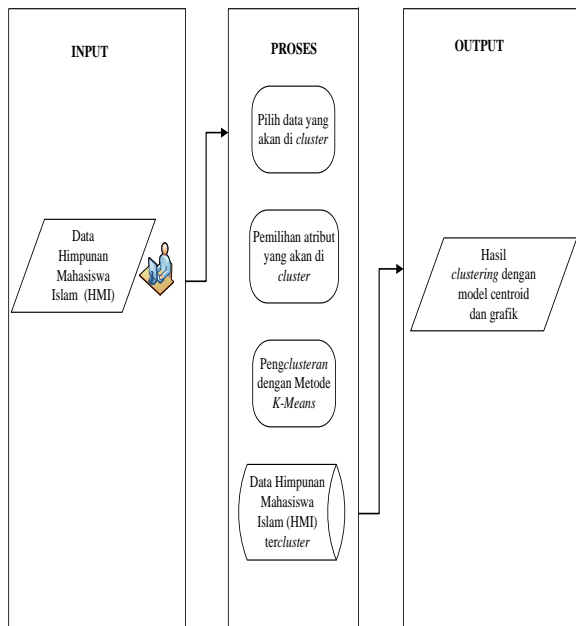
## 4. Hasil dan Pembahasan

### 4.1. Hasil Penelitian

Hasil penelitian ini adalah sebuah sistem informasi pemetaan pendidikan yang dirancang, merupakan sebuah sistem yang dapat mengelompokkan data-data yang mirip menjadi satu kelompok atau cluster menurut tingkat kemiripan datanya. Sistem dapat mengelompokkan data dengan menggunakan K-Means clustering. Tujuan dari sistem ini yaitu mendapatkan kumpulan-kumpulan data yang memiliki kemiripan sehingga masuk dalam suatu cluster yang memetakan pendidikan formal yang di seimbangkan dengan pendidikan non-formal (training) sebagai pendukung dalam peningkatan kualitas diri dalam mengali potensi yang di miliki.

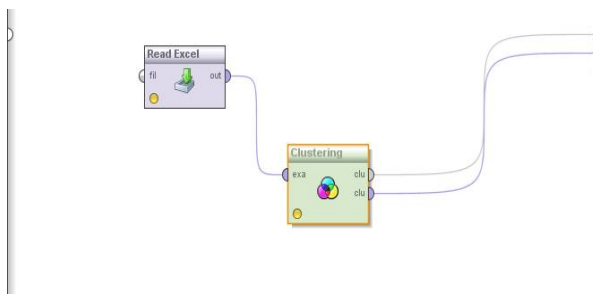
Penelitian ini menerapkan metode K-Means cluster pada perangkat lunak untuk mendapatkan

pengekelompokan data pendidikan formal (s1, s2, s3) dengan pendidikan non formal (training lk1, lk2, lk3 dan lkk), proses clustering dengan metode *k-means* dalam penelitian ini menggunakan alat bantu aplikasi Rapid Miner, ada pun penjabaran aplikasi tertera pada Gambar 5 berikut ini.



Gambar 5. Kerangka sistem informasi pemetaan pendidikan dengan metode *K-Means*

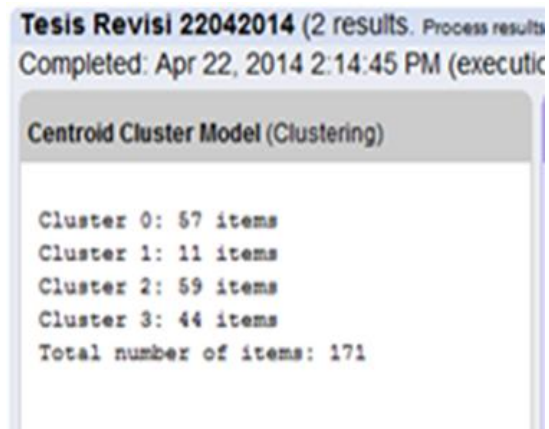
Pada Gambar 5 ini menjelaskan proses pembuatan aplikasi sistem informasi pemetaan pendidikan terbagi menjadi tiga tahapan subsistem yaitu subsistem masukan, subsistem proses dan subsistem output. Sedangkan pemetaan tertera pada Gambar 6 berikut ini.



Gambar 6. Relasi data pemetaan pendidikan dengan metode *K-Means*

#### 4.2. Output Clustering Pemetaan Pendidikan dengan Metode *K-Means*

Dari hasil *cluster* dengan metode *k-means* di dapatkan *centroid cluster* model yang telah di tentukan terlebih dahulu di bagi menjadi empat *cluster*, seperti terlihat pada Gambar 7.



Gambar 7. Hasil pemetaan pendidikan *Centroid Cluster* Model 4

Keseluruhan data terdiri dari 171 data Himpunan Mahasiswa Islam dengan dengan pendidikan formal S1, S2, S3 dan pendidikan non-formal Jenis pelatihan LK I, LK II, LK II, LKK, akan dihitung menggunakan Rapid Miner 5. Hasil statistik dan jarak dalam setiap atribut dapat dilihat pada Tabel 4.1. Nilai titik tengah dapat dilihat pada Tabel 1.

Hasil nilai statistik data set asli dari 10 atribut didapatkan bahwa Cluster nominal mode = cluster\_0 (57), cluster\_1 (11), cluster\_2 (59) cluster\_3 (44).

Hasil nilai statistik data set asli dari 10 atribut didapatkan bahwa :

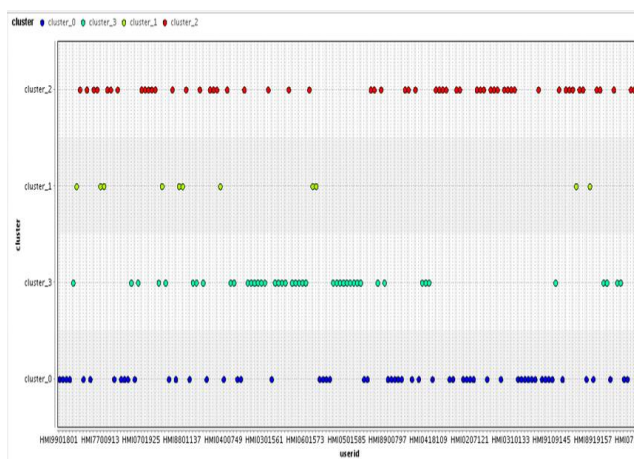
- Atribut lk1(X1) dengan type data integeravg = 1 +/- 0 range = [1.000 ; 1.000
- Atribut lk2 (X2) dengan type data integer avg = 0.684 +/- 0.466 range = [0.000; 1.000]
- Atribut lk3 (X3) dengan type data integer avg = 0.298 +/- 0.459 range = [0.000; 1.000]
- Atribut lkk (X4) dengan type data integer avg = 0.158 +/- 0.366 range = [0.000; 1.000]
- Atribut s1 (X5) dengan type data integer avg = 1 +/- 0 [1.000 ; 1.000] 0.0
- Atribut s2 (X6) dengan type datainteger avg = 0.795 +/- 0.982 [0.000 ; 2.000] 0.0
- Atribut s3 (X7) dengan type datainteger avg = 0.193 +/- 0.738 [0.000 ; 3.000] 0.0

Hasil *centroid* rata-rata nilai *cluster* pemetaan pendidikan dengan total data yang mengikuti pendidikan formal 171 yang bergelar S1, 58 bergelar S2 dan 11 yang bergelar S3 sedang untuk total data yang mengikuti pelatihan (pendidikan non-formal)171 yang ikut LK I, 117 yang ikut LK II, 51 yang ikut LK III dan 27 yang ikut LKK, *centroid* rata-rata nilai *cluster* seperti terlihat pada Tabel 1.

Tabel 1. Centroid rata-rata nilai Cluster pemetaan pendidikan

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
lk1 (X1)	1	1	1	1
lk2 (X2)	0.860	0.818	1	0
lk3 (X3)	0.386	0.364	0.424	0
lk4 (X4)	0.316	0.091	0.119	0.023
s1 (X5)	1	1	1	1
s2 (X6)	2	2	0	0
s3 (X7)	0	3	0	0

Output pemetaan pendidikan dengan grafik *cluster scatter* menunjukan hasil *cluster* masing-masing data yang mirip di kelompokkan menjadi satu *cluster* dengan simbol warna, *cluster\_0* dengan warna biru, *cluster\_1* dengan warna hijau, *cluster\_2* dengan warna merah, *cluster\_3* dengan warna hijau toska, titik-titik dengan masing-masing warna *cluster* ini menunjukan anggota dari masing-masing *cluster*, seperti yang terlihat pada Gambar 8.



Gambar 8. Output pemetaan pendidikan grafik Cluster Scatter

### 4.3. Pengujian Hasil Clustering Pemetaan Pendidikan

Setelah data melalui pemrosesan *clustering* dengan sepuluh variabel, terpilih delapan atribut dengan userid di pilih sebagai label di dapatkan hasil uji yang memperlihatkan bahwa tidak data yang kosong (missing values) sehingga semua data pemetaan pendidikan dapat tercluster dengan baik. Hasil uji dapat dilihat pada Tabel 2.

Penelitian ini melakukan pengelompokan pada data Himpunan Mahasiswa Islam berdasarkan kemiripan data menggunakan teknik *cluster* dengan metode *k-means*. Dari hasil pengujian dengan 171 data, dari pengelompokan ini didapatkan hasil *clustering* nilai akademisi dan nilai\_pelatihan (training berjenjang) data Himpunan Mahasiswa Islam yang menjadi studi kasus, tidak terdapat perubahan pada data yang dilakukan pengujian,

mendapatkan hasil 171 data (valid) seperti pertama kali melakukan pengujian dengan metode *k-means*, dari hasil pengujian dengan metode *k-means* didapatkan empat model *cluster*.

Tabel 2. Hasil Uji Cluster pemetaan pendidikan dengan metode K-means

Role	Name	Type	Range	Missings	Comment
-	lk1 (X1)	integer	= [1.000...1.000]; mean =1.000	no missing values	-
-	lk2 (X2)	integer	= [0.000...1.000]; mean =0.684	no missing values	-
-	lk3 (X3)	integer	= [0.000...1.000]; mean =0.298	no missing values	-
-	lk4 (X4)	integer	= [0.000...1.000]; mean =0.158	no missing values	-
-	s1 (X5)	integer	= [1.000...1.000]; mean =1.000	no missing values	-
-	s2 (X6)	integer	= [0.000...2.000]; mean =0.795	no missing values	-
-	s3 (X7)	integer	= [0.000...3.000]; mean =0.193	no missing values	-
label	userid	polynomial	= {HMI0000496, HMI0001154, HMI0001812, HMI0101163, HMI0101556, HMI0101558, HMI0101559, HMI0101567, HMI0101583, ...}	no missing values	-
id	id	integer	= [1.000...171.000]; mean =86.000	no missing values	-
cluster	cluster	nominal	= {cluster_0, cluster_1, cluster_2, cluster_3}	no missing values	-

Dari hasil *cluster* menggunakan metode *k-means* data dikelompokkan ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada pada kelompok lain. Metode *k-means* meminimalkan variasi antar data yang ada dalam satu *cluster* dan memaksimalkan variasi dengan data yang ada pada *cluster* lain. Seperti yang terlihat pada hasil pemetaan pendidikan dengan metode *k-means*.

### 4.4. Clustering dengan Metode K-Means

Dari hasil komputasi yang dilakukan dengan alat bantu aplikasi Rapid Miner yang masuk menjadi anggota masing-masing *cluster* di ambil dari data maksimal yang berbobot 1, 2 dan 3, bobot ini menunjukan pada masing-masing anggota *cluster* telah mengikuti pelatihan dan berpendidikan s1,s2 dan s3. Contoh *cluster* 0 seperti terlihat pada Tabel 3.

Tabel 3. Contoh penjelasan dari output Clustering

Cluster	User Id (Label)	X1	X2	X3	X4	X5	X6	X7
C <sub>0</sub>	HMI9901801	1	0	0	1	1	2	0
C <sub>1</sub>	HMI8701506	1	0	0	0	1	2	3
C <sub>2</sub>	HMI1201807	1	1	0	0	1	0	0
C <sub>3</sub>	HMI0801505	1	0	0	0	1	0	0

Hasil uji yang telah dilakukan, penyebaran data pada masing-masing *cluster* memiliki kemiripan. Setiap pelatihan mempunyai syarat yang berjenjang yang harus di ikuti oleh setiap anggota HMI,

sehingga pada saat proses iterasi tidak terlihat perbedaan yang signifikan pada hasil data pemetaan pendidikan dengan metode *k-means*, jika di lihat dari pendidikan formal yang berjenjang terlihat pada *cluster\_1* dengan gelar S1 (X5), S2 (X6), S3 (X7). Dalam pendidikan formal pada studi kasus dengan pendidikan non-formal (pelatihan), terlihat pada *cluster\_2*, yang mengikuti pelatihan/training LK I (X1), LK II (X2), dengan pendidikan S1 (X5). Seperti terlihat pada Tabel. 4.

Tabel 4. Rekapitulasi jumlah anggota untuk setiap *cluster*

<i>Cluster</i>	0	1	2	3
LK I (X1)	57	11	59	44
LK II (X2)	49	9	59	0
LK III (X3)	22	4	25	0
LKK (X4)	17	1	7	1
S1 (X5)	57	11	59	44
S2 (X6)	57	11	0	0
S3 (X7)	0	11	0	0
Valid 171 Data	57	11	59	44

Dari hasil *cluster* hanya ada beberapa variabel yang terpilih pada masing-masing *cluster*, yang di tentukan oleh ikut atau tidaknya pendidikan S2, S3 dan pada pelatihan berjenjang. Pada hasil komputasi pemilihan *cluster* di lihat pada pendidikan formal dan pendidikan non-formal yang terakhir di ikuti, yang berpendidikan lebih tinggi tidak akan menjadi anggota pada *cluster* di bawahnya, seperti terlihat pada Tabel 10.

Tabel 10. Output *Cluster* yang terpilih

<i>Cluster</i>	<i>Variabel</i>			
0	LK I (x1)	S1 (x5)	S2 (x6)	0
1	LK I (x1)	S1 (x5)	S2 (x6)	S3 (x7)
2	LK I (x1)	LK II (x2)	S1 (x5)	0
3	LK I (x1)	S1 (x5)	0	0

## 5. Kesimpulan

Sistem informasi pemetaan pendidikan dengan metode *k-means* digunakan untuk mengcluster data pendidikan yang memiliki kemiripan. Dari beberapa hasil uji yang telah dilakukan, penyebaran data pada masing-masing *cluster* memiliki tingkat kemiripan data. Pada saat proses iterasi tidak terlihat perbedaan yang signifikan pada hasil pemetaan pendidikan dengan metode *k-means*. Pada pendidikan formal,

pendidikan berjenjang terlihat pada *cluster 1* untuk pendidikan non-formal (pelatihan) pada studi kasus, pendidikan berjenjang terlihat pada *cluster 2*.

*Cluster* data yang dibuat berhasil menentukan klasifikasi atau pengelompokan data pemetaan pendidikan berdasarkan nilai akademisi dan nilai pelatihan, dalam jenjang pendidikan non-formal (training) dan pendidikan formal dengan rincian sebagai berikut, dari hasil pengelompokan ini didapatkan hasil *clustering* nilai pelatihan dan nilai akademisi data Himpunan Mahasiswa Islam yang menjadi studi kasus sejumlah 171. Informasi hasil centroid model cluster 4 dengan variabel meliputi anggota yang berpendidikan S1, S2, telah mengikuti pelatihan dasar masuk cluster 0, berpendidikan S1, S2, S3 telah mengikuti pelatihan dasar masuk cluster 1, berpendidikan S1 telah mengikuti pelatihan dasar dan pelatihan menengah masuk cluster 2, berpendidikan S1 telah mengikuti pelatihan dasar masuk cluster 3. Berdasarkan pengujian hasil *k-means cluster*.

## Daftar Pustaka

- Agusta, Y., 2007. K-Means- Application, Problems and Related Methods.
- Dunham, M.H., 2003, Data Mining Introductory and Advanced Topics, Southern Methodist University, Pearson Education Inc : New Jersey.
- Dominikus, J. and Feri, S., 2010, Rapidminer (YALE), PT. Alex Media Komputindo.
- Giyanto, H., 2008. Penerapan algoritma Clustering K-Means, K-Medoid, Gath Geva. Tesis, Yogyakarta: Universitas Gajah Mada.
- Han, J. and Kamber, M., 2007. Data Mining: Concepts and Techniques. San Fransisco: Elsevier Inc.
- Huang, Z., 2005. A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining. CSIRO Mathematical and Information Sciences – Australia.
- Larose, 2005. Discovering Knowledge in Data: An Introduction to Data Mining.
- Myatt, G.J., 2007. Making Sense of Data : A Practival Guide to Exploratory Data Analysis and Data Mining. Hoboken: John Wiley and Sons.
- Priadi, S., 2012. Mapping of education as a basis for improving education services, *Papers ICEMAL (International Conference Educational Management, Administration and Leadership)*, 4-5.