

## Data Mining Pengelompokan Akta Nikah Berdasarkan Usia Nikah atau Domisili menggunakan Metode Clustering: Studi Kasus Kemenag Langkat

Nurlaila<sup>1</sup>, Relita Buatun<sup>2</sup>, Rusmin Saragih<sup>3</sup>

<sup>1,2,3</sup>Prodi Sistem Informasi STMIK Kaputama Binjai, Sumatera Utara

[nurlaila13122000@gmail.com](mailto:nurlaila13122000@gmail.com), [bbcbuatun@gmail.com](mailto:bbcbuatun@gmail.com), [evitha12014@gmail.com](mailto:evitha12014@gmail.com)

### ABSTRACT

*One way to find out the marriage certificate data is to group the prospective bride and groom data that has been recorded at the Langkat Ministry of Religion. The analysis was carried out using the Clustering method using the K-Means method which was translated into a software. This software is used for grouping data. The output is to find out the grouping of marriage certificate data with the closest relationship between the age group of the groom, occupation and address. The results showed that the age of the groom-to-be between 19-28 years old had a civil servant job at the address/district of Kutambaru.*

**Keywords:** *age, occupation, address/location and clustering*

### ABSTRAK

Salah satu cara untuk mengetahui data akta nikah adalah dengan mengelompokkan data-data calon pengantin yang telah terdaftar di Kemenag Langkat. Analisa dilakukan dengan *metode Clustering* yang menggunakan *metode K-Means* yang kemudian diterjemahkan dalam sebuah perangkat lunak. Perangkat lunak ini yang digunakan untuk mengelompokkan data. Output untuk mengetahui pengelompokan data akta nikah dengan hubungan terdekat antara kelompok usia calon pengantin laki-laki, pekerjaan dan alamat. Hasil penelitian menunjukkan terdapat usia calon pengantin laki-laki antara 19-28 tahun memiliki pekerjaan PNS berada di alamat/kecamatan Kutambaru.

**Kata kunci :** *usia, pekerjaan, alamat/lokasi dan clustering*

### PENDAHULUAN

Perkembangan teknologi informasi semakin hari semakin bertambah pesat. perkembangan itu berpengaruh pada sistem kerja perusahaan atau instansi pemerintah yang semua sudah berbasis teknologi informasi. Tidak hanya internet tetapi semua teknologi digunakan demi kemajuan instansi tersebut. Baik itu untuk melakukan pengelolaan data, promosi, mau pun hanya menyimpan data. Pada kenyataannya, memang teknologi saat ini sangat berdampak pada kinerja perusahaan atau instansi pemerintah agar tidak ketinggalan dalam persaingan saat ini.

Salah satu cara untuk mengetahui data akta nikah adalah dengan mengelompokkan data-data calon pengantin yang telah terdaftar di Kemenag Langkat. Analisa dilakukan dengan *metode Clustering* yang menggunakan *metode K-Means* yang kemudian diterjemahkan dalam sebuah perangkat lunak. Perangkat lunak ini yang digunakan untuk mengelompokkan data. Output untuk mengetahui

pengelompokan data akta nikah dengan hubungan terdekat antara kelompok usia calon pengantin laki-laki, pekerjaan dan alamat.

Berdasarkan permasalahan diatas maka penulis akan mempengaruhi kriteria-kriteria (variabel) apa saja untuk mengelompokkan berdasarkan variabel usia calon pengantin laki-laki, pekerjaan dan alamat.

## TINJAUAN LITERATUR

### Data Mining

Data Mining sering juga disebut *Knowledge Discovery in Database*, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Menurut Alfa Saleh, (2015 : 208), menjelaskan bahwa: “Data Mining merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi-informasi yang nantinya dapat digunakan”.

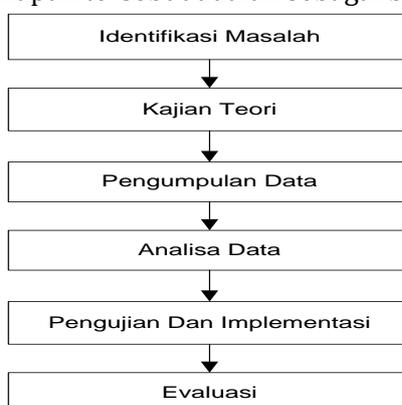
### Clustering

*Clustering* merupakan metode penganalisaan data, yang sering dimasukan sebagai salah satu metode data mining, yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama. Menurut Hermawati (2012, h:123) “Analisa kluster (clustering) yaitu menemukan objek-objek dalam satu kelompok sama (punya hubungan) dengan yang lain dan berada (titik punya hubungan) dengan objek-objek dalam kelompok lain”.

Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data /objek ke dalam *cluster* (kelompok) sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin. Metode *clustering* berusaha untuk menempatkan obyek yang mirip (jaraknya dekat) dalam datu kelompok dan membuat jarak antar kelompok sejauh mungkin. Ini berarti obyek dalam satu kelompok sangat mirip satu sama lain dan berbeda dengan obyek dalam kelompok-kelompok yang lain.

## METODE PENELITIAN

Ada beberapa tahapan metodologi penelitian yang di lakukan dalam penyelesaian masalah. Tahapan tersebut adalah sebagai berikut :



Gambar I Alur Kerja Penelitian

## HASIL DAN PEMBAHASAN

### Analisis Sistem

Dengan menganalisa data-data akta nikah yang menumpuk dan data hanya diproses di Arsip, maka dapat dilihat permasalahan selama ini yaitu menumpuknya data akta nikah tanpa mengetahui data yang sebenarnya yang terdapat di Arsip tersebut. Penyeleksian data – data yang ada hanya secara manual sehingga untuk mengetahui data akta nikah berdasarkan usia nikah dan domisili.

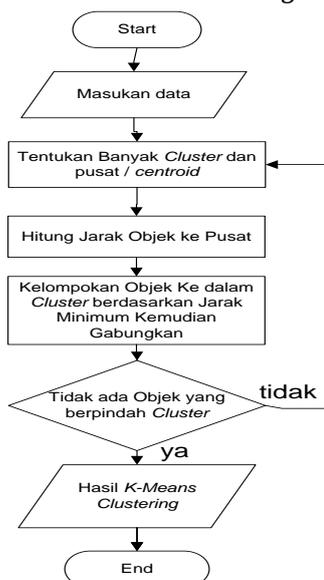
Dari banyaknya data yang telah terkumpul belum diterapkan pengelompokan usia, pekerjaan dan alamat. Sehingga data yang sudah ada menjadi kurang informative. Penerapan data mining pada data akta nikah diolah dengan Aplikasi Matlab, serta menggunakan metode *clustering* yang nantinya menjadi tolak ukur keberhasilan pengembangan sistem. *Clustering* adalah suatu metode pengelompokan berdasarkan ukuran kedekatan (kemiripan). Sebuah *cluster* tidak harus sama akan tetapi pengelompokannya berdasarkan pada kedekatan dari suatu karakteristik sampel yang ada, salah satunya dengan menggunakan rumus jarak *Euclidean*. Desain konseptual untuk menggambarkan aktivitas-aktivitas yang dilakukan dengan menggunakan *flowchart*.

### Perancangan Sistem

Berdasarkan analisis yang dilakukan, penulis akan menggali data untuk dikelompokkan dengan menggunakan metode *clustering* dengan algoritma *K-Means*, dimana variabel-variabel dari data akta nikah yang akan diambil untuk penelitian tersebut adalah usia, pekerjaan dan alamat. Kemudian dari data tersebut di masukan kedalam *Matlab* sehingga terdapat hasil 3 kelompok.

### Perancangan Flowchart

Adapun *flowchart* algoritma hirarki *clustering* dan *K-means* adalah :



Gambar 2 Flowchart Algoritma Hirarki *Clustering* Dan *K-means*

Pengelompokan data dengan metode *K-Means* ini secara umum dilakukan dengan cara sebagai berikut :

1. Masukan data
2. Tentukan jumlah kelompok
3. Alokasi data ke dalam kelompok secara acak.
4. Hitung pusat kelompok (centroid/ rata-rata) dari data yang ada di masing-masing kelompok.
5. Alokasikan masing-masing data ke centroid/rata-rata terdekat.
6. Kembali ke langkah 3 apabila masih ada data yang berpindah kelompok, atau apabila ada perubahan nilai centroid di atas nilai ambang yang ditentukan, atau apabila perubahan nilai fungsi objektif yang digunakan masih di atas nilai ambang yang ditentukan.

### Analisis Transformasi Data

Transformasi data merupakan proses pengubahan atau penggabungan data kedalam format yang sesuai untuk diproses dalam data mining. Sering kali data yang akan digunakan dalam proses data mining mempunyai format yang belum langsung bisa digunakan, oleh karena itu perlu dirubah formatnya.

### Usia

Dari data akat nikah yang ada di Kemenag Langkat terdapat data usia. Berikut ini adalah tabel usia:

**Tabel 1 Usia**

Kode	Usia
1	19-28
2	29-38
3	39-48
4	49-58
5	59-68

### Pekerjaan

Dari data akta nikah yang ada di Kemenag Langkat terdapat data pekerjaan. Berikut ini adalah tabel pekerjaan:

**Tabel 2 Pekerjaan**

Kode	Pekerjaan
1	Wiraswasta
2	PNS
3	TNI
4	Petani
5	Peternak
6	Karyawan Honorer
7	Guru
8	Perdagangan

9	Karyawan BUMN
10	Polisi
11	Karyawan Swasta

## Alamat

Dari data akta nikah yang ada di Kemenag Langkat terdapat data alamat. Berikut ini adalah tabel alamat:

**Tabel 3 Alamat**

Kode	Alamat
1	Bahorok
2	Kutambaru
3	Salapian
4	Kuala
5	Selesai
6	Binjai
7	Sei Bingai
8	Stabat
9	Secanggang
10	Wampu
11	Hinai
12	Padang Tualang
13	Batang Serangan
14	Sawit Seberang
15	Tanjung Pura
16	Gebang
17	Babalan
18	Sei Lapan
19	Tanjung Pura
20	Pangkalan Susu
21	Brandan Barat
22	Besitang
23	Pematang Jaya

## Data Yang Akan di-Clustering

Data yang sudah ada dimasukkan ke tabel untuk dilakukan Transformasi, data yang dibutuhkan sebanyak 20 data terlebih dahulu. Berikut ini adalah data yang akan di transformasikan.

**Tabel 4 Data yang akan di-Clustering**

No	Nama Calon Pengantin Laki-laki	Usia Calon Pengantin Laki-laki	Pekerjaan	Alamat/ Kecamatan
1	M.YUSUF SIDIK	21	WIRASWASTA	Padang Tualang

2	FAHRUL AGRIYAN	26	TNI	Batang Serangan
3	AGUS SULEIMAN	23	WIRASWASTA	Bahorok
4	DEVI KURNIA	23	TNI	Kutambaru
5	ARIMBI	23	WIRASWASTA	Salapian
6	SUGIAT	30	PETANI	Padang Tualang
7	HANDRIA	23	PETERNAK	Batang Serangan
8	HERDI HARIYONON	23	KARYAWAN HONORER	Bahorok
9	ANDRIAN	22	WIRASWASTA	Kutambaru
10	FAUZI HIDAYAT	24	GURU	Salapian
11	SUHARDIMAN	27	PERDAGANGAN	Kuala
12	ABDUL KARIM	25	WIRASWASTA	Selesai
13	FERI IRAWAN	26	KARYAWAN BUMN	Binjai
14	YAYOK	40	PETANI	Sei Bingai
15	BISTON TARIGAN	35	WIRASWASTA	Stabat
16	SANDY THADASY	28	WIRASWASTA	Secanggang
17	RAMLISAH	33	PETANI	Wampu
18	SUKARNO PUTRA	24	PERDAGANGAN	Hinai
19	T.NAZLA PARHAN	22	WIRASWASTA	Padang Tualang
20	RUDINI MARBUN	24	WIRASWASTA	Padang Tualang

#### 4.0 Data Yang di Transformasi

Data yang sudah di transformasikan di dapat dari tiga variabel yang ada, kemudian ditransformasikan sesuai kode yang terdapat di tabel masing-masing variabel. Berikut ini adalah tabel data yang sudah di transformasikan :

**Tabel 5 Data yang di Trasformasi**

No	Nama Calon Pengantin Laki-laki	Usia Calon Pengantin Laki-laki	Pekerjaan	Alamat/ Kecamatan
1	A	1	2	1
2	B	1	3	14
3	C	1	3	1
4	D	1	1	2

5	E	1	4	3
6	F	2	5	13
7	G	1	6	14
8	H	1	1	1
9	I	1	7	2
10	J	1	8	3
11	K	1	1	4
12	L	1	9	6
13	M	1	4	7
14	N	3	1	8
15	O	2	1	9
16	P	1	1	10
17	Q	2	8	11
18	R	1	1	12
19	S	1	1	13
20	T	1	2	20

Lakukan *cluster* menjadi 3 kelompok ( $K=3$ ) dan tentukan titik pusat centroid. Adapun proses perhitungan *clustering* seperti dibawah ini.  $K=3$  Centroid

$C_1 = (1,2,1)$  diambil dari data A

$C_2 = (1,3,14)$  diambil dari data B

$C_3 = (1,1,10)$  diambil dari data P

Iterasi 1 :

1. A (1,2,1)

$$C_1(1,2,1) = \sqrt{(1-1)^2 + (2-2)^2 + (1-1)^2} = 0,00$$

$$C_2(1,3,14) = \sqrt{(1-1)^2 + (2-3)^2 + (1-14)^2} = 13,04$$

$$C_3(1,1,10) = \sqrt{(1-1)^2 + (2-1)^2 + (1-10)^2} = 9,06$$

2. B (1,3,14)

$$C_1(1,2,1) = \sqrt{(1-1)^2 + (3-2)^2 + (14-1)^2} = 13,04$$

$$C_2(1,3,14) = \sqrt{(1-1)^2 + (3-3)^2 + (14-14)^2} = 0,00$$

$$C_3(1,1,10) = \sqrt{(1-1)^2 + (3-1)^2 + (14-10)^2} = 4,47$$

3. C (1,1,1)

$$C_1(1,2,1) = \sqrt{(1-1)^2 + (1-2)^2 + (1-1)^2} = 1,00$$

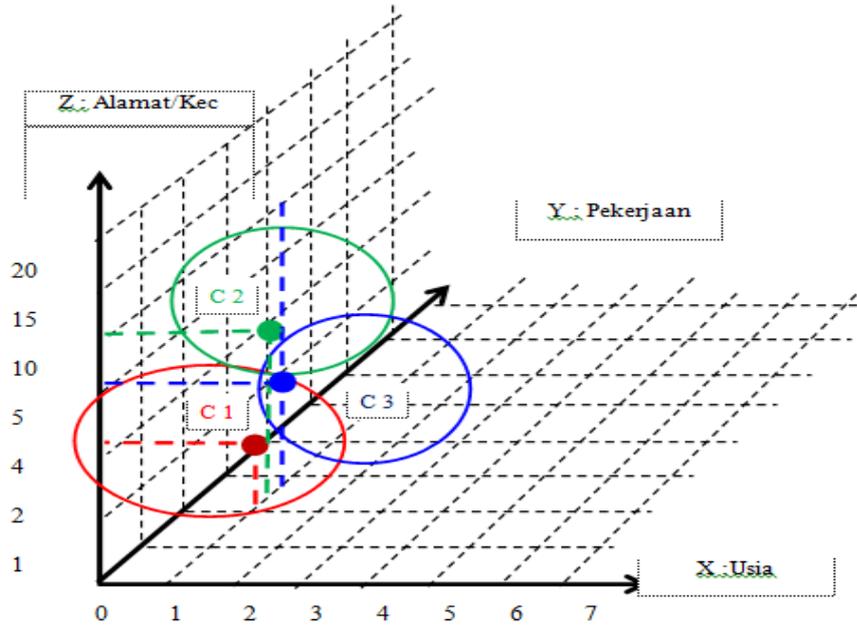
$$C_2(1,3,14) = \sqrt{(1-1)^2 + (1-3)^2 + (1-14)^2} = 13,15$$

$$C_3(1,1,10) = \sqrt{(1-1)^2 + (1-1)^2 + (1-10)^2} = 9,00$$

Dan seterusnya sampai data ke 20

**Grafik Clustering**

Membuat grafik *cluster* berdasarkan perhitungan yang telah dilakukan. Adapun grafik yang diperoleh adalah sebagai berikut :



**Gambar 3 Grafik Cluster berdasarkan perhitungan yang telah dilakukan**

- Cluster 1 :1,00; 2,13;2,13
- Cluster2 :1,17; 2,83;14,33
- Cluster3 :1,67; 2,67;8,50

**Penjelasan Grafik :**

Dari 20 data diperoleh 3 *group*, Cluster 1 terdapat 8 data, Cluster 2 terdapat 6 data, dan Cluster 3 terdapat 6 data. Dan diperoleh Group terbanyak adalah cluster 1.

**1. Cluster 1 Terdapat 8 Data**

1,00; 2,13; 2,13

Dapat diketahui pada cluster 1 terdapat usia calon pengantin laki-laki antara 19-28 tahun memiliki pekerjaan PNS berada di alamat/kecamatan Kutambaru.

**2. Cluster 2 Terdapat 6 Data**

1,17; 2,83; 14,33

Dapat diketahui pada cluster 2 terdapat usia calon pengantin laki-laki antara 19-28 tahun memiliki pekerjaan PNS dan TNI berada di alamat/kecamatan Sawit Seberang.

**3. Cluster 3 Terdapat 6 Data**

1,67; 2,67; 8,50

Dapat diketahui pada *cluster* 3 terdapat usia calon pengantin laki-laki antara 19-28 dan 29-38 tahun memiliki pekerjaan PNS dan TNI berada di alamat/kecamatan Stabat.

## KESIMPULAN DAN SARAN

1. Dapat diketahui pada *cluster* 1 terdapat usia calon pengantin laki-laki antara 19-28 tahun memiliki pekerjaan PNS berada di alamat/kecamatan Kutambaru.
2. Dapat diketahui pada *cluster* 3 terdapat usia calon pengantin laki-laki antara 19-28 dan 29-38 tahun memiliki pekerjaan PNS dan TNI berada di alamat/kecamatan Stabat.
3. Dapat diketahui pada *cluster* 2 terdapat usia calon pengantin laki-laki antara 19-28 tahun memiliki pekerjaan PNS dan TNI berada di alamat/kecamatan Sawit Seberang.

## DAFTAR PUSTAKA

- Izzaty, R.E., Astuti, B., & Cholimah, N. (1967). *Angewandte Chemie International Edition*, 6(11), 951-952., 2006,5-24.
- Sembiring, F., Octaviana, O., & Saepudin, S. (2020). Implementasi Metode K-Means Dalam Pengklasteran Daerah Pungutan Liar Di Kabupaten Sukabumi (Studi Kasus : Dinas Kependudukan Dan Pencatatan Sipil). *Jurnal Tekno Insentif*, 14(1), 40-47. <https://doi.org/10.36787/jti.v14i1.165>.
- Pranata, Dana, Hamdani Hamdani, and Dyna Marisa Khairina. 2015. "Rancang Bangun Website Jurnal Ilmiah Bidang Komputer (Studi Kasus : Program Studi Ilmu Komputer Universitas Mulawarman)." *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer* 10(2):25. doi: 10.30872/jim.v10i2.187.
- Eko Prasetyo, *Data Mining Konsep Dan Aplikasi Menggunakan Matlab*, PT.Andi, Yogyakarta, 2012.
- Yatini B, *Definisi Flowchart*, Jakarta PT. Gramedia Widia Sarana Indonesia, 2006
- Kusrini, Emha Taufiq Luthfi, *Algoritma Data Mining*, Penerbit C.V Andi, Yogyakarta 2009.