



# Algoritma K-Means untuk Pengelompokan Dokumen Akta Kelahiran pada Tiap Kecamatan di Kabupaten Simalungun

Flora Sabarina Napitupulu\*, Irfan Sudahri Damanik, Ilham Syahputra Saragih, Anjar Wanto

Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: <sup>1,\*</sup>florasabarinanapitupulu@gmail.com, <sup>2</sup>irfansudahri@amiktunasbangsa.ac.id,

<sup>3</sup>ilhamsaragih@amiktunasbangsa.ac.id, <sup>4</sup>anjarwanto@amiktunasbangsa.ac.id

## Abstrak

Akta kelahiran merupakan salah satu dokumen yang wajib untuk dimiliki oleh seorang warga negara. Dokumen ini berisikan informasi seputar kelahiran seseorang dan merupakan pencatatan resmi bukti pengakuan negara akan keberadaan orang tersebut. Tujuan dari penelitian ini adalah untuk mengelompokkan dokumen akta kelahiran pada tiap kecamatan di Kabupaten Simalungun. Data penelitian diperoleh dari Dinas Kependudukan dan Pencatatan Sipil Kabupaten Simalungun. Algoritma pengelompokan yang digunakan adalah Algoritma K-means yang merupakan salah satu algoritma Data Mining yang baik digunakan untuk kasus pengelompokan. Dengan menggunakan algoritma ini data yang telah diperoleh dapat dikelompokkan ke dalam beberapa cluster, di mana penerapan proses Clustering K-Means menggunakan alat RapidMiner. Data dibagi menjadi 3 kelompok: tinggi (C1), sedang (C2) dan rendah (C3). Hasil yang diperoleh dari penelitian ini adalah bulan Desember masuk kedalam klaster level tinggi (C1), bulan Agustus, September dan Oktober masuk kedalam klaster sedang (C2), dan bulan Januari, Februari, Maret, April, Mei, Juni, Juli dan November masuk kedalam klaster rendah (C3).

**Kata Kunci:** K-Means, Pengelompokan, Kecamatan, Akta Kelahiran, Simalungun

## Abstract

A birth certificate is a document that must be owned by a citizen. This document contains information about the birth of a person and is an official record of proof of state recognition of the person's existence. The purpose of this study is to group birth certificate documents in each district in Simalungun Regency. The research data was obtained from the Population and Civil Registry Office of Simalungun Regency. The grouping algorithm used is the K-means algorithm which is one of the Data Mining algorithms that is good for the case of grouping. By using this algorithm the data that has been obtained can be grouped into several clusters, where the application of the K-Means Clustering process uses the RapidMiner tool. Data is divided into 3 groups: high (C1), medium (C2) and low (C3). The results obtained from this study are in December entered into a high level cluster (C1), in August, September and October entered into a medium cluster (C2), and in January, February, March, April, May, June, July and November entered into the low cluster (C3).

**Keywords:** K-Means, Clustering, District, Birth Certificate, Simalungun

## 1. PENDAHULUAN

Akta Kelahiran adalah bukti sah mengenai status dan peristiwa kelahiran seseorang yang di keluarkan oleh Dinas Kependudukan dan Pencatatan Sipil yang merupakan salah satu dokumen yang wajib untuk dimiliki oleh seorang warga negara [1]. Bayi yang dilaporkan kelahirannya akan terdaftar dalam Kartu Keluarga dan diberi Nomor Induk Kependudukan (NIK) sebagai dasar untuk memperoleh Pelayanan Masyarakat lainnya [2]. Dalam pelaksanaannya, Akte Lahir menjadi salah satu dokumen kependudukan yang banyak manfaatnya. namun demikian, kepemilikan Akte Kelahiran masih belum sebanding dengan jumlah penduduknya sehingga tidak menutup kemungkinan Dinas Kependudukan dan Pencatatan Sipil Kabupaten Simalungun membuka pelayanan terpadu jemput bola ke daerah-daerah yang masih minim dalam kepemilikan Akte Kelahiran yang sejalan dengan melaksanakan program pemerintah dalam percepatan kepemilikan Akte Kelahiran yang di pimpin langsung oleh Dirjen Dukcapil. Target kepemilikan Akte Kelahiran yang di berikan kepada Dinas Kependudukan dan Pencatatan Sipil (Disdukcapil) daerah-daerah harus mencapai 92% untuk usia 1-17 tahun. Disdukcapil Kab. Simalungun telah melakukan beberapa program layanan masyarakat seperti, membuka layanan masyarakat ke tiap-tiap kecamatan (jemput bola) yang ada di Kab. Simalungun, namun belum mendapatkan hasil yang memuaskan karena hasil dari jemput bola ke tiap-tiap kecamatan masih jauh dari target yang di berikan Dirjen Dukcapil. Hal ini disebabkan oleh karena ketiadaan pemetaan dan clustering data kepemilikan Akte Kelahiran per-Kecamatan setelah pelaksanaan program jemput bola. Sehingga pelaksanaan program jemput bola selama hampir 1 (satu) tahun belum tepat sasaran. Untuk itu dibutuhkan suatu metode pengelompokan data kepemilikan Akta Kelahiran pada semua kecamatan yang ada di Kab. Simalungun.

Banyak cabang ilmu komputer yang mampu menyelesaikan masalah-masalah yang beraneka ragam dan kompleks. Hal ini terbukti dari penelitian terdahulu oleh para peneliti dalam menyelesaikan permasalahan dalam Sistem Pendukung Keputusan [3]–[10], dalam bidang Jaringan Saraf Tiruan [11]–[23], maupun dalam bidang Data Mining [24]–[31]. Berdasarkan permasalahan yang telah dijabarkan tersebut maka penulis melakukan penelitian dengan memanfaatkan data kependudukan Kab. Simalungun dari tahun 2016 sampai dengan tahun 2017 seperti nomor induk kependudukan (NIK), nama penduduk, nomor kartu keluarga, alamat penduduk, jenis pekerjaan, tanggal kelahiran penduduk, jumlah penduduk per-kecamatan dimana data-data tersebut akan di proses dengan menggunakan data mining clustering yakni algoritma K-Means yang nantinya menampilkan informasi daerah atau lokasi kecamatan mana yang diprioritaskan untuk dikunjungi. Hal ini dikarenakan K-



Means merupakan suatu algoritma yang mampu melakukan pengelompokkan secara pertisi yang memisahkan data ke dalam kelompok yang berbeda – berbeda.

## 2. METODE PENELITIAN

### 2.1 Metode Pengumpulan Data

Penelitian ini menggunakan data kuantitatif dari Disdukcapil Kabupaten Simalungun. Pengumpulan data yang digunakan untuk penelitian terdapat beberapa metode yang terdiri dari :

- a. Penelitian Kepustakaan (*Libarry Research*) yaitu memanfaatkan perpustakaan, buku, prosiding atau journal sebagai media untuk bahan referensi dalam menentukan faktor, parameter dan label yang digunakan untuk penelitian.
- b. Penelitian Lapangan (*Field Work Research*) yaitu penelitian yang dilakukan secara langsung ke kantor Disdukcapil. Penulis memperoleh informasi dengan mengumpulkan data, mempelajari data, validasi data dan mencari referensi terkait dengan kasus pada penelitian. Luaran dari studi literature ini adalah tersusun dan terkoleksinya referensi yang baik dan benar dengan penelitian.

### 2.2 Data Penelitian

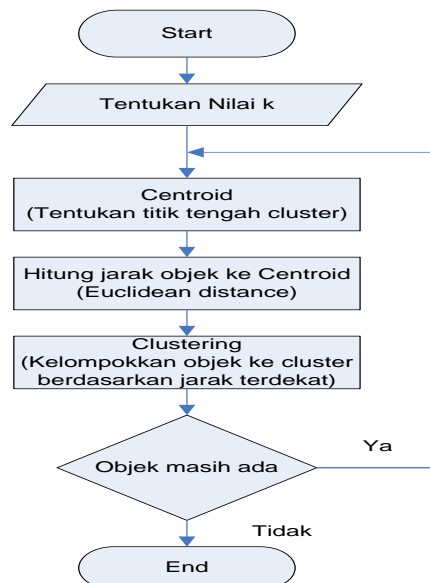
Data penelitian yang digunakan adalah data akta kelahiran di Kabupaten Simalungun tahun 2016 – 2017 berdasarkan bulan. Data akta kelahiran diperoleh dari Disdukcapil Kabupaten Simalungun. Data tersebut dapat dilihat pada tabel berikut:

**Tabel 1.** Data Akta Kelahiran Kab. Simalungun Tahun 2016-2017

No	Bulan	2016	2017
1	Januari	1.162	991
2	Februari	1.512	1.277
3	Maret	1.613	974
4	April	1.253	730
5	Mei	1.850	736
6	Juni	1.965	707
7	Juli	1.668	1.375
8	Agustus	4.430	956
9	September	5.182	1.360
10	Oktober	5.133	2.576
11	November	2.006	241
12	Desember	1.211	2.616
<b>Jumlah</b>		<b>31.001</b>	<b>14.539</b>

### 2.3. Diagram Alir K-Means

Berikut ini akan disajikan Diagram alir dari algoritma *Clustering K-Means* [32].



**Gambar 1.** Diagram alir algoritma *K-Means*



Langkah-langkah algoritma *K-Means* dapat dijelaskan sebagai berikut [33]–[38]:

- a. Tentukan jumlah *cluster* (*k*) pada data set.
- b. Tentukan nilai pusat (*Centroid*).

Penentuan nilai *Centroid* pada tahap awal dilakukan secara *random*, sedangkan pada tahap iterasi digunakan rumus seperti pada persamaan (1) berikut ini.

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \tag{1}$$

Keterangan :

$V_{ij}$  = *Centroid* rata-rata *cluster* ke-*i* untuk variabel ke-*j*

$N_i$  = Jumlah anggota *cluster* ke-*i*

*i, k* = Indeks dari *cluster*

*j* = Indeks dari variabel

$X_{kj}$  = nilai data ke-*k* variabel ke-*j* untuk *cluster* tersebut

- c. Pada masing-masing *record*, hitung jarak terdekat dengan *Centroid*.

Ada beberapa cara yang dapat digunakan untuk mengukur jarak data ke pusat kelompok, diantaranya *Euclidean*, *Manhattan/City Block*, dan *Minkowsky*. Setiap cara memiliki kelebihan dan kekurangan masing-masing. Untuk penulisan pada bab ini, jarak *Centroid* yang digunakan adalah *Euclidean Distance*, dengan rumus seperti dibawah ini:

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2} \tag{2}$$

Keterangan :

$De$  = *Euclidean Distance*

*i* = Banyaknya objek <sup>2</sup>

(*x, y*) = Koordinat objek

(*s, t*) = Koordinat *Centroid*

- d. Kelompokkan objek berdasarkan jarak ke *Centroid* terdekat

- e. Ulangi langkah ke-3 hingga langkah ke-4, lakukan *iterasi* hingga *Centroid* bernilai optimal.

### 3. ANALISA DAN PEMBAHASAN

#### 3.1 Langkah-langkah Perhitungan K-Means

- a. Menentukan data yang akan diolah

Data yang digunakan adalah data akta kelahiran di Kabupaten Simalungun tahun 2016 – 2017 berdasarkan bulan. Data akta kelahiran diperoleh dari Disdukcapil Kabupaten Simalungun. Data tersebut dapat dilihat pada tabel 1 pada pembahasan sebelumnya

- b. Menentukan Jumlah Cluster

Ada 3 Cluster yang ditetapkan pada penelitian ini, diantaranya adalah Rendah (C1), Sedang (C2) dan Tinggi (C3) berdasarkan data akta kelahiran pada tahun 2016 dan 2017.

- c. Menentukan Centroid

Penentuan pusat awal Cluster (*Centroid*) ditentukan secara manual atau *random* yang diambil dari data yang ada dalam range. Nilai Cluster 0 diambil dari data paling rendah, Nilai Cluster 1 diambil dari data rata-rata atau nilai tengah pada data dan Nilai Cluster 2 diambil dari data yang paling tertinggi.

**Tabel 2.** Nilai Centroid

Data Cluster	2016	2017
(C1) Cluster Tinggi	1162	991
(C2) Cluster Normal	1668	1375
(C3) Cluster Rendah	5182	1360

- d. Menghitung Jarak dari Centroid

Untuk menghitung jarak antara titik Centroid dengan titik tiap objek menggunakan *Euclidian Distance*. Rumus untuk menghitung jarak dari Centroid adalah :

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

Maka perhitungan untuk jarak dari *Centroid* ke-1 adalah sebagai berikut :

$$D_{x1,cl} = \sqrt{(1162 - 1162)^2 + (991 - 991)^2} = 0$$

$$D_{x2,cl} = \sqrt{(1512 - 1162)^2 + (1277 - 991)^2} = 451,99$$

$$D_{x3,cl} = \sqrt{(1613 - 1162)^2 + (974 - 991)^2} = 451,32$$

$$D_{x4,cl} = \sqrt{(1253 - 1162)^2 + (730 - 991)^2} = 276,41$$

$$D_{x5,cl} = \sqrt{(1850 - 1162)^2 + (736 - 991)^2} = 733,74$$



Dan seterusnya sampai dengan  $D_{x12,c1}$ . Selanjutnya perhitungan untuk jarak dari *Centroid* ke-2 adalah sebagai berikut :

$$D_{x1,c2} = \sqrt{(1162 - 1668)^2 + (991 - 1357)^2} = 635,21$$

$$D_{x2,c2} = \sqrt{(1512 - 1668)^2 + (1277 - 1357)^2} = 184,23$$

$$D_{x3,c2} = \sqrt{(1613 - 1668)^2 + (974 - 1357)^2} = 404,75$$

$$D_{x4,c2} = \sqrt{(1253 - 1668)^2 + (730 - 1357)^2} = 766,97$$

$$D_{x5,c2} = \sqrt{(1850 - 1668)^2 + (736 - 1357)^2} = 664,41$$

Dan seterusnya sampai dengan  $D_{x12,c2}$ . Selanjutnya perhitungan untuk jarak dari *Centroid* ke-3 adalah sebagaiberikut :

$$D_{x1,c3} = \sqrt{(1162 - 5182 + (991 - 1360))^2} = 4036,90$$

$$D_{x2,c3} = \sqrt{(1512 - 5182 + (1277 - 1360))^2} = 3670,94$$

$$D_{x3,c3} = \sqrt{(1613 - 5182 + (974 - 1360))^2} = 3589,81$$

$$D_{x4,c3} = \sqrt{(1253 - 5182 + (730 - 1360))^2} = 3979,19$$

$$D_{x5,c3} = \sqrt{(1850 - 5182 + (736 - 1360))^2} = 3389,93$$

Dan seterusnya sampai dengan  $D_{x12,c3}$ . Sehingga didapat table jarak dari *Centroid* dan mencari nilai minimal dari ketiga *Centroid*. Tabel Jarak dari *Centroid* adalah sebagai berikut:

**Tabel 3.** Jarak *Centroid* Iterasi ke-1

C1	C2	C3	Jarak <i>Centroid</i>
0,00	635,21	4036,90	0,00
451,99	184,23	3670,94	184,23
451,32	404,75	3589,81	404,75
276,41	766,97	3979,19	276,41
733,74	664,41	3389,93	664,41
851,74	731,05	3282,61	731,05
635,21	0,00	3514,03	0,00
3268,19	2793,60	853,65	853,65
4036,90	3514,03	0,00	0,00
4275,64	3667,24	1216,99	1216,99
1129,09	1183,30	3367,36	1129,09
1625,74	1322,47	4164,90	1322,47

e. Menentukan Cluster

Dalam menentukan *Cluster* dengan mencari nilai *Cluster* berdasarkan nilai minimal dari nilai *Cluster* dan diletakkan pada *Cluster* yang sesuai dengan nilai minimal pada Iterasi 1. Berikut table *Cluster* pada Iterasi 1 sebagai berikut:

**Tabel 4.** *Cluster* Iterasi ke-1

C1	C2	C3
1		
	1	
	1	
1		
	1	
	1	
	1	
		1
		1
		1
1		
	1	

Selanjutnya dalam metode *K-Means*, perhitungan berhenti apabila *Cluster* pada iterasi yang dihasilkan sama pada iterasi sebelumnya. Maka selanjutnya mencari *Cluster* pada iterasi selanjutnya sampai nilai Iterasinya sama. Untuk mencari nilai *Centroid* selanjutnya dengan menggunakan *Centroid* baru pada Iterasi ke-1 dengan menjumlahkan nilai sesuai yang tertera pada *Cluster* di table diatas. Adapun *Centroid* baru untuk



mencari *Cluster* selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada *Cluster* tersebut kemudian membagikannya sebanyak jumlah nilai sebagai berikut :

$$C_{1,2016} = (1162 + 1253 + 2006)/3 = 1474$$

$$C_{2,2016} = (1512 + 1613 + 1850 + 1965 + 1668 + 1211)/6 = 1637$$

$$C_{3,2016} = (4430 + 5182 + 5133)/3 = 4915$$

$$C_{1,2017} = (991 + 730 + 241)/3 = 654$$

$$C_{2,2017} = (1277 + 974 + 736 + 707 + 1375 + 2616)/6 = 1281$$

$$C_{3,2017} = (956 + 1360 + 2576)/3 = 1631$$

Maka, data *Centroid* baru Iterasi ke-1 adalah sebagai berikut :

**Tabel 5.** *Centroid* Baru Iterasi Ke-1

Data Cluster	2016	2017
(C1) Cluster Tinggi	1474	654
(C2) Cluster Normal	1637	1281
(C3) Cluster Rendah	4915	1631

Dengan menggunakan langkah – langkah yang sama seperti sebelumnya untuk menentukan Jarak dari *Centroid* dengan menggunakan *Centroid* baru Iterasi ke-1, maka berikut hasil Jarak dari *Centroid*:

**Tabel 6.** Jarak *Centroid* Iterasi ke-2

C1	C2	C3	Jarak <i>Centroid</i>
459,03	556,02	3807,12	459,03
624,18	124,56	3421,33	124,56
349,02	307,73	3366,66	307,73
233,39	671,19	3771,13	233,39
385,16	585,17	3192,91	385,16
494,18	661,21	3091,22	494,18
746,73	99,30	3257,05	99,30
2971,72	2812,32	830,90	830,90
3774,94	3546,38	380,20	380,20
4133,38	3728,67	970,14	970,14
673,76	1103,53	3223,89	673,76
1979,50	1401,33	3832,82	1401,33

Dari tabel Jarak *Centroid* diatas, maka *Cluster* atau pengelompokan Iterasi ke-2 adalah sebagai berikut :

**Tabel 7.** *Cluster* Iterasi ke-2

C1	C2	C3
1		
	1	
	1	
1		
1		
1		
	1	
		1
		1
		1
1		
	1	

Dari tabel *Cluster* Iterasi ke-1 dan *Cluster* Iterasi ke-2 berbeda, maka perhitungan dilanjut sampai data *Cluster* tersebut tidak ada yang berubah. Selanjutnya mencari *Centroid* dari Iterasi ke-2. Maka berikut table *Centroid* baru Iterasi ke 2 :

**Tabel 8.** *Centroid* Baru Iterasi Ke-2

Data Cluster	2016	2017
(C1) Cluster Tinggi	1647	681
(C2) Cluster Normal	1501	1561
(C3) Cluster Rendah	4915	1631



**Tabel 9.** Jarak *Centroid* Iterasi ke-3

<b>C1</b>	<b>C2</b>	<b>C3</b>	<b>Jarak <i>Centroid</i></b>
575,78	662,76	3807,12	575,78
611,14	283,71	3421,33	283,71
294,99	597,10	3366,66	294,99
397,23	866,74	3771,13	397,23
210,13	895,32	3192,91	210,13
318,86	971,47	3091,22	318,86
694,31	249,60	3257,05	249,60
2796,35	2990,73	830,90	830,90
3599,42	3686,46	380,20	380,20
3967,60	3771,29	970,14	970,14
567,75	1412,84	3223,89	567,75
1983,56	1094,61	3832,82	1094,61

**Tabel 10.** *Cluster* Iterasi ke-3

<b>C1</b>	<b>C2</b>	<b>C3</b>
1		
	1	
1		
1		
1		
1		
	1	
		1
		1
		1
1		
	1	

Dari tabel *Cluster* Iterasi ke-3 diatas dengan table *Cluster* Iterasi ke-2 sebelumnya berubah nilai *Cluster*-nya maka selanjutnya mencari Iterasi ke-4 atau seterusnya sampai nilai *Cluster* tersebut tidak ada berubah atau sama. Berikut perhitungan untuk mencari *Cluster* Iterasi ke-4 :

**Tabel 11.** *Centroid* Baru Iterasi Ke-3

<b>Data Cluster</b>	<b>2016</b>	<b>2017</b>
(C1) Cluster Tinggi	1642	730
(C2) Cluster Normal	1464	1756
(C3) Cluster Rendah	4915	1631

**Tabel 12.** Jarak *Centroid* Iterasi ke-4

<b>C1</b>	<b>C2</b>	<b>C3</b>	<b>Jarak <i>Centroid</i></b>
546,01	822,33	3807,12	546,01
562,28	481,43	3421,33	481,43
245,82	796,13	3366,66	245,82
388,50	1047,40	3771,13	388,50
208,59	1090,71	3192,91	208,59
324,30	1162,64	3091,22	324,30
645,71	432,33	3257,05	432,33
2797,66	3072,32	830,90	830,90
3596,14	3739,36	380,20	380,20
3949,54	3759,84	970,14	970,14
609,77	1609,15	3223,89	609,77
1934,67	896,35	3832,82	896,35

**Tabel 13.** *Cluster* Iterasi ke-4

<b>C1</b>	<b>C2</b>	<b>C3</b>

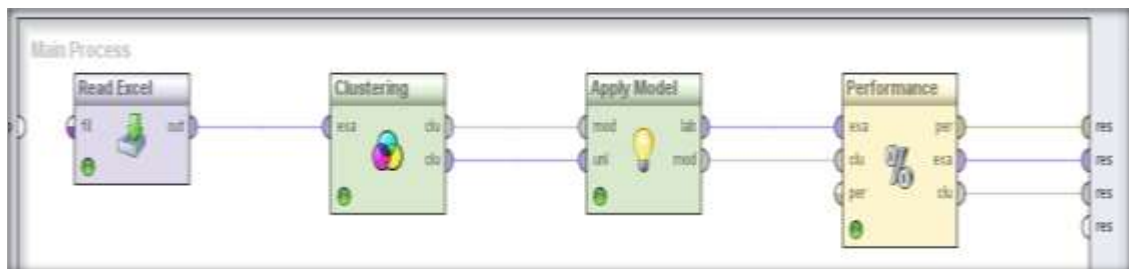
C1	C2	C3
1		
	1	
1		
1		
1		
1		
	1	
		1
		1
		1
1		
	1	

Dari tabel *Cluster* Iterasi ke-3 dan tabel *Cluster* Iterasi ke-4 memiliki nilai *Cluster* yang sama atau tidak berubah pada *Cluster* Iterasi ke-4 maka perhitungan dihentikan dan hasil yang diperoleh yaitu :

1. *Cluster* 0 (*C1*) memiliki 6 data yang diartikan bahwa kelompok pertama adalah kategori angka kelahiran rendah pada tahun 2016-2017.
2. *Cluster* 1 (*C2*) memiliki 3 data yang diartikan bahwa kelompok kedua adalah kategori angka kelahiran sedang padat ahun 2016-2017.
3. *Cluster* 3 (*C3*) memiliki 3 data yang diartikan bahwa kelompok ketiga adalah kategori angka kelahiran tinggi pada tahun 2016-2017.

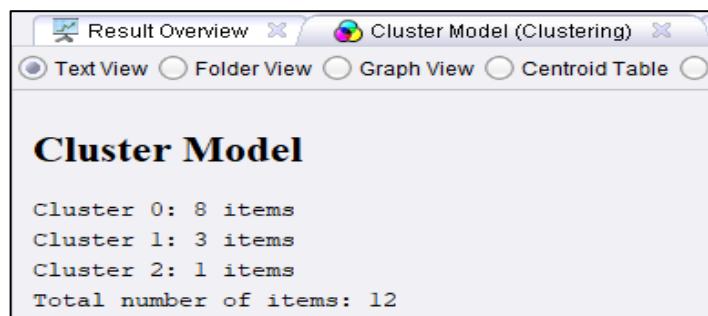
#### 4. IMPLEMENTASI

Berdasarkan data pada tabel 1, pengelompokan data Akta Kelahiran pada Tiap Kecamatan di Kabupaten Simalungun akan dilakukan dengan tools RapidMiner. Proses pengelompokan dengan algoritma K-Means menggunakan RapidMiner dapat dilihat pada gambar berikut:



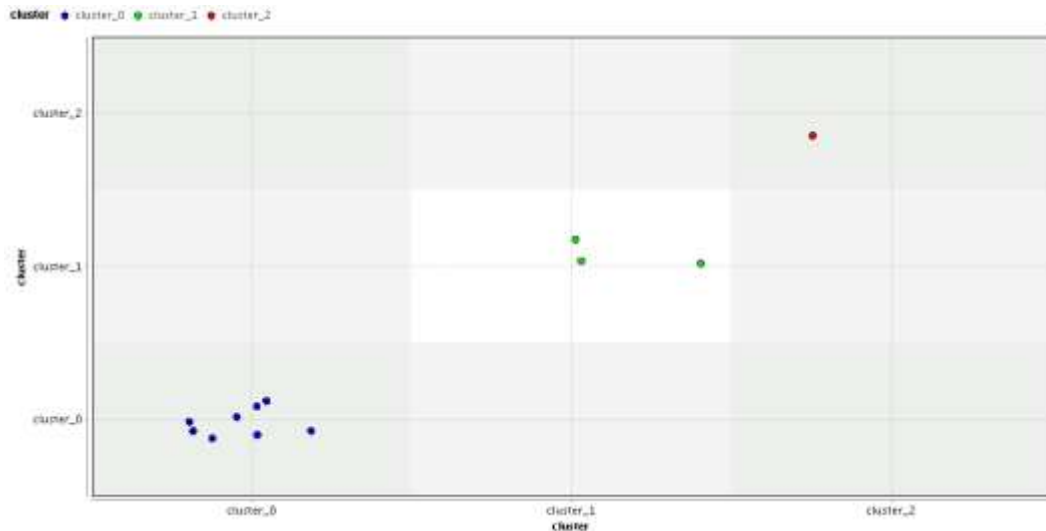
**Gambar 2.** Proses Pengelompokan Algoritma K-Means dengan Nilai K = 3

Gambar 2 adalah proses menghubungkan read excel dengan K-Means dan output yang akan dieksekusi dengan nilai K = 3. Semua parameter yang diperlukan disimpan dalam model objek. Dalam hal ini para peneliti menggunakan sampel data Akta Kelahiran (Berdasarkan Tabel 1) dengan 2 atribut, yaitu: tahun 2016 dan tahun 2017. Hasil pengelompokan data akhir dapat dilihat pada gambar berikut:



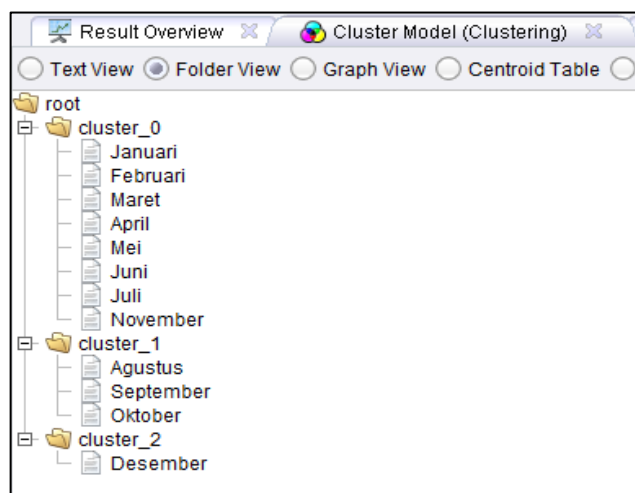
**Gambar 3.** Hasil Clustering menggunakan K-Means

Berdasarkan gambar 3 dapat dijelaskan bahwa terdapat 3 kluster mulai dari kluster 0, kluster 1, dan kluster 1. Kluster 0 adalah kluster rendah, kluster 1 adalah kluster normal, dan kluster 2 adalah kluster tinggi. Jadi kami mendapatkan grafik tampilan plot dari pengujian dengan RapidMiner sebagai berikut:



**Gambar 4.** Grafik Clustering dalam Tampilan Plot

Sedangkan hasil pengelompokan data akta dapat dilihat pada gambar 5 berikut:



**Gambar 5.** Pengelompokan Dokumen Akta Kelahiran pada Tiap Kecamatan di Kab. Simalungun

Pada gambar 5 dapat dijelaskan bahwa bulan Desember masuk kedalam klaster level tinggi (C1), bulan Agustus, September dan Oktober masuk kedalam klaster sedang (C2), dan bulan Januari, Februari, Maret, April, Mei, Juni, Juli dan November masuk kedalam klaster rendah (C3)

## 5. KESIMPULAN

Diperoleh suatu model aturan yang dapat memperlihatkan model aturan dalam menganalisa data kelahiran, hasil penelitian berisi tentang analisa dan rancangan yang digunakan dalam Clustering data akta kelahiran terbanyak pada kecamatan Kabupaten Simalungun. Adapun hasil eksperimen yang dilakukan pada penelitian ini menunjukkan bahwa algoritma clustering K-Means melakukan pengelompokan tweet dengan cara membagi data ke dalam jumlah cluster k yang ditentukan, dan memanfaatkan perhitungan jarak untuk mengukur kemiripan antar data. Pemilihan jumlah cluster (k) sangat mempengaruhi hasil clustering dan nilai purity. Semakin besar jumlah cluster (k) mengakibatkan nilai purity yang dihasilkan pun semakin baik.

## REFERENCES

- [1] S. Liem, K. Gunadi, and A. N. Tjondrowiguno, "Penggunaan Convolutional Recurrent Neural Network dan RLSA untuk Mengambil Data pada Akta Kelahiran," *Jurnal Infra Petra*, vol. 7, no. 1, pp. 8–14, 2019.
- [2] A. M. M. Anwar, P. Harsani, and A. Maesya, "Penentuan Daerah Prioritas Pelayanan Akta Kelahiran Dengan Metode K-NN dan K-Means," *Komputasi: Jurnal Ilmiah Ilmu Komputer dan Matematika*, vol. 17, no. 1, pp. 319–328, 2020.
- [3] P. P. P. A. N. . F. I. R.H Zer, Masitha, A. P. Windarto, and A. Wanto, "Analysis of the ELECTRE Method on the Selection of Student





- Creativity Program Proposals,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
- [4] A. P. W. Budiharjo and A. Muhammad, “Comparison of Weighted Sum Model and Multi Attribute Decision Making Weighted Product Methods in Selecting the Best Elementary School in Indonesia,” *International Journal of Software Engineering and Its Applications*, vol. 11, no. 4, pp. 69–90, 2017.
  - [5] D. R. Sari, N. Rofiqo, D. Hartama, A. P. Windarto, and A. Wanto, “Analysis of the Factors Causing Lazy Students to Study Using the ELECTRE II Algorithm,” *Journal of Physics: Conference Series*, vol. 1255, no. 012007, pp. 1–6, 2019.
  - [6] K. Fatmawati *et al.*, “Analysis of Promethee II Method in the Selection of the Best Formula for Infants Under Three Years,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
  - [7] P. Alkhairi, L. P. Purba, A. Eryzha, A. P. Windarto, and A. Wanto, “The Analysis of the ELECTREE II Algorithm in Determining the Doubts of the Community Doing Business Online,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
  - [8] S. Sundari, Karmila, M. N. Fadli, D. Hartama, A. P. Windarto, and A. Wanto, “Decision Support System on Selection of Lecturer Research Grant Proposals using Preferences Selection Index,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
  - [9] S. R. Ningsih, R. Wulansari, D. Hartama, A. P. Windarto, and A. Wanto, “Analysis of PROMETHEE II Method on Selection of Lecturer Community Service Grant Proposals,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
  - [10] T. Imandasari, M. G. Sadewo, A. P. Windarto, A. Wanto, H. O. Lingga Wijaya, and R. Kurniawan, “Analysis of the Selection Factor of Online Transportation in the VIKOR Method in Pematangsiantar City,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
  - [11] E. Siregar, H. Mawengkang, E. B. Nababan, and A. Wanto, “Analysis of Backpropagation Method with Sigmoid Bipolar and Linear Function in Prediction of Population Growth,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
  - [12] G. W. Bhawika *et al.*, “Implementation of ANN for Predicting the Percentage of Illiteracy in Indonesia by Age Group,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
  - [13] W. Saputra, J. T. Hardinata, and A. Wanto, “Resilient method in determining the best architectural model for predicting open unemployment in Indonesia,” *IOP Conference Series: Materials Science and Engineering*, vol. 725, no. 1, pp. 1–7, 2020.
  - [14] A. Wanto and J. T. Hardinata, “Estimations of Indonesian poor people as poverty reduction efforts facing industrial revolution 4.0,” *IOP Conference Series: Materials Science and Engineering*, vol. 725, no. 1, pp. 1–8, 2020.
  - [15] A. Wanto *et al.*, “Model of Artificial Neural Networks in Predictions of Corn Productivity in an Effort to Overcome Imports in Indonesia,” *Journal of Physics: Conference Series*, vol. 1339, no. 1, pp. 1–6, 2019.
  - [16] A. Wanto *et al.*, “Analysis of the Accuracy Batch Training Method in Viewing Indonesian Fisheries Cultivation Company Development,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
  - [17] I. S. Purba *et al.*, “Accuracy Level of Backpropagation Algorithm to Predict Livestock Population of Simalungun Regency in Indonesia Accuracy Level of Backpropagation Algorithm to Predict Livestock Population of Simalungun Regency in Indonesia,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
  - [18] P. Parulian *et al.*, “Analysis of Sequential Order Incremental Methods in Predicting the Number of Victims Affected by Disasters,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
  - [19] A. Wanto *et al.*, “Analysis of the Backpropagation Algorithm in Viewing Import Value Development Levels Based on Main Country of Origin,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
  - [20] S. Setti, A. Wanto, M. Syafiq, A. Andriano, and B. K. Sihotang, “Analysis of Backpropagation Algorithms in Predicting World Internet Users,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
  - [21] T. Afriliansyah *et al.*, “Implementation of Bayesian Regulation Algorithm for Estimation of Production Index Level Micro and Small Industry,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
  - [22] A. Wanto *et al.*, “Forecasting the Export and Import Volume of Crude Oil, Oil Products and Gas Using ANN,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
  - [23] M. K. Z. Sormin, P. Sihombing, A. Amalia, A. Wanto, D. Hartama, and D. M. Chan, “Predictions of World Population Life Expectancy Using Cyclical Order Weight / Bias,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
  - [24] R. Rahim *et al.*, “C4.5 Classification Data Mining for Inventory Control,” *International Journal of Engineering & Technology*, vol. 7, pp. 68–72, 2018.
  - [25] I. Parlina *et al.*, “Naive Bayes Algorithm Analysis to Determine the Percentage Level of visitors the Most Dominant Zoo Visit by Age Category,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, 2019.
  - [26] I. S. Damanik, A. P. Windarto, A. Wanto, Poningsih, S. R. Andani, and W. Saputra, “Decision Tree Optimization in C4.5 Algorithm Using Genetic Algorithm,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
  - [27] H. Siahaan, H. Mawengkang, S. Efendi, A. Wanto, and A. Perdana Windarto, “Application of Classification Method C4.5 on Selection of Exemplary Teachers,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, 2019.
  - [28] D. Hartama, A. Perdana Windarto, and A. Wanto, “The Application of Data Mining in Determining Patterns of Interest of High School Graduates,” *Journal of Physics: Conference Series*, vol. 1339, no. 1, pp. 1–6, 2019.
  - [29] M. Widyastuti, A. G. Fepdiani Simanjuntak, D. Hartama, A. P. Windarto, and A. Wanto, “Classification Model C.45 on Determining the Quality of Customer Service in Bank BTN Pematangsiantar Branch,” *Journal of Physics: Conference Series*, vol. 1255, no. 012002, pp. 1–6, 2019.
  - [30] W. Katrina, H. J. Damanik, F. Parhusip, D. Hartama, A. P. Windarto, and A. Wanto, “C.45 Classification Rules Model for Determining Students Level of Understanding of the Subject,” *Journal of Physics: Conference Series*, vol. 1255, no. 012005, pp. 1–7, 2019.
  - [31] S. Sudirman, A. P. Windarto, and A. Wanto, “Data Mining Tools | RapidMiner : K-Means Method on Clustering of Rice Crops by Province as Efforts to Stabilize Food Crops In Indonesia,” *IOP Conference Series: Materials Science and Engineering*, vol. 420, no. 012089, pp. 1–8, 2018.
  - [32] Z. S. Younus *et al.*, “Content-based image retrieval using PSO and k-means clustering algorithm,” *Arabian Journal of Geosciences*, vol. 8, no. 8, pp. 6211–6224, 2015.
  - [33] A. Wanto *et al.*, *Data Mining : Algoritma dan Implementasi*. Yayasan Kita Menulis, 2020.
  - [34] E. Prasetyo, *Data Mining: Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: Andi Offset, 2012.
  - [35] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*. New Jersey: John Wiley & Sons, 2005.
  - [36] R. Primartha, *Belajar Machine Learning Teori dan Praktik*. Bandung: Informatika Bandung, 2018.
  - [37] T. Khotimah, “Pengelompokan Surat dalam Al Qur’an menggunakan Algoritma K-Means,” *Jurnal Simetris*, vol. 5, no. 1, pp. 83–88, 2014.
  - [38] I. Parlina, A. P. Windarto, A. Wanto, and M. R. Lubis, “Memanfaatkan Algoritma K-Means dalam Menentukan Pegawai yang Layak Mengikuti Assessment Center untuk Clustering Program SDP,” *CESS (Journal of Computer Engineering System and Science)*, vol. 3, no. 1, pp. 87–93, 2018.