

PENERAPAN K-MEANS CLUSTERING PADA PENDAFTARAN ANGGOTA PERPUSTAKAAN DI DINAS KEARSIPAN DAN PERPUSTAKAAN PROVINSI JAWA TENGAH

Tsurayya Nur Laila¹, Arief Jananto²

Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Stikubank

e-mail: [1tsurayya60@gmail.com](mailto:tsurayya60@gmail.com), [2ajananto09@edu.unisbank.ac.id](mailto:ajananto09@edu.unisbank.ac.id)

ABSTRAK

Dinas Kearsipan dan Perpustakaan Provinsi Jawa Tengah merupakan pusat pengolahan buku dan arsip yang ada di Kota Semarang. Dinas Kearsipan dan Perpustakaan Provinsi Jawa Tengah menyediakan berbagai jenis layanan dan juga koleksi, karena itu hampir setiap hari masyarakat umum dari berbagai latar belakang datang ke perpustakaan untuk menikmati layanan dan koleksi yang disediakan. Para pengunjung diwajibkan memiliki kartu anggota perpustakaan, jadi setiap hari banyak masyarakat yang ingin mendaftar menjadi anggota. Untuk mengelompokkan kalangan masyarakat yang mendaftar di perpustakaan menggunakan metode clustering dengan algoritma K-Means yang diharapkan dapat membantu untuk mengetahui kalangan apa yang mendominasi pendaftaran anggota di perpustakaan. Penelitian ini menggunakan 2007 data dengan jumlah cluster 3 cluster.

Hasil dari perhitungan untuk masing –masing cluster adalah untuk cluster 1 berjumlah 173 data, untuk yang mendominasi pendaftaran anggota pada cluster 1 adalah berusia 17 sampai 25 tahun, bertingkat pendidikan S1 yang merupakan Pegawai Swasta. Cluster 2 berjumlah 967 yang mendominasi adalah berusia 17 sampai 25 tahun, bertingkat pendidikan S1 yang merupakan Mahasiswa. Cluster 3 berjumlah 867 data yang mendominasi adalah berusia 6 sampai 11 tahun yang bertingkat pendidikan SD yang merupakan sebagai Pelajar.

Kata Kunci: Pendaftaran Anggota Perpustakaan, Data Mining, Clustering, K-Means

1. PENDAHULUAN

Perpustakaan merupakan tempat yang menyediakan berbagai koleksi seperti buku, bahan cetakan serta rekaman lain untuk kepentingan masyarakat umum. Perpustakaan juga dapat diartikan sebagai tempat pendidikan bagi masyarakat luas sebagai sarana pembelajaran karena perpustakaan menyediakan berbagai sumber informasi ilmu pengetahuan, budaya dan teknologi, serta koleksi yang dapat memudahkan masyarakat dalam memperoleh informasi yang dibutuhkan. Dinas Kearsipan dan Perpustakaan Provinsi Jawa Tengah merupakan instansi Pemerintahan, yang merupakan pusat pengolahan buku dan arsip yang ada di Kota Semarang yang menyediakan berbagai jenis layanan dan juga koleksi. Di Dinas Kearsipan dan Perpustakaan Provinsi Jawa Tengah mewajibkan untuk para Pengunjung memiliki kartu anggota perpustakaan yang bertujuan agar dapat menikmati layanan dan koleksi yang ada, jadi setiap hari banyak masyarakat yang ingin mendaftar menjadi anggota perpustakaan Provinsi Jawa Tengah.

Proses pendaftaran anggota menghasilkan data yang sangat berlimpah. Banyaknya data pendaftaran tersebut dapat dimanfaatkan untuk mencari informasi yang tersembunyi, untuk mengetahui informasi yang ada didalamnya dapat dilakukan dengan cara mengolah dan memproses data yang ada dengan menggunakan metode clustering dan algoritma K-Means untuk mengetahui kalangan apa yang mendominasi pendaftaran anggota di perpustakaan.

2. TINJAUAN PUSTAKA

Penelitian tentang penerapan data mining Dalam Pengelompokan Penderita Thalassaemia[1] menjelaskan bahwa penderita Thalassaemia berjumlah 203 pada tahun 2015 di wilayah Kota dan Kabupaten Tasikmalaya studi kasus RSUD dr. Soekardjo dan RS. Prasetya Bunda. Metode yang digunakan merupakan metode *Clustering* dan algoritma *K-means*. Berhasil mengelompokkan penderita Thalassaemia menjadi 3 kelompok, kelompok 1 berjumlah 214 data, kelompok 2 berjumlah 138 data, dan yang terakhir kelompok 3 berjumlah 23 data.

Penelitian K-Mean clustering[2] bertujuan untuk mengetahui seberapa besar hasil dari pengelompokan barang berpengaruh terhadap kebutuhan dari konsumen dengan menggunakan metode *Clustering* dengan algoritma *K-means*, Dari hasil akhir penelitian akan diketahui jumlah stok terbanyak dan sedikit dari pengelompokan jenis artikel.

Penelitian Clustering Profil Pengunjung Perpustakaan yang bertujuan untuk mengelompokan dan membangun suatu aplikasi yang dapat mengelompokan profil pengunjung berdasarkan status dan alamat. Metode yang dipilih adalah metode *Clusering* dengan algoritma *K-means* yang dikombinasikan dengan metode optimasi elbow.[3]

3. METODE PENELITIAN

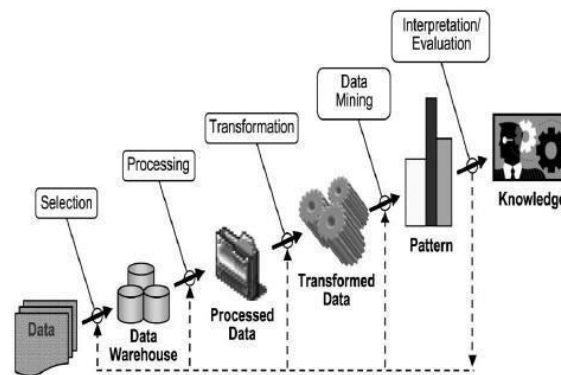
3.1 Objek Penelitian

Objek penelitian ini adalah data primer berupa data pendaftaran anggota perpustakaan dari bulan januari sampai maret 2020 yang didapatkan secara langsung dari Dinas Kearsipan dan Perpustakaan Provinsi Jawa Tengah yang beralamat di Jalan Sriwijaya No 29A Tegalsari, Kec Candisari, Kota Semarang, Jawa Tengah.

Data sekunder merupakan data yang diperoleh secara tidak langsung yang digunakan untuk melengkapi data primer seperti buku, jurnal, dan artikel yang berkaitan dengan penelitian yang dilakukan.

3.2 Knowledge Discovery in Database (KDD)

Istilah *data mining* dan Knowledge Discovery in Databases (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar [4]



Gambar 1. Tahapan Knowledge Discovery in Databases (KDD)

Proses-proses dalam Knowledge Discovery in Databases (KDD) sebagai tersebut:

1. Data Selection

Data yang didapat berupa data *excel* yang berjumlah 2827 record. Berikut merupakan potongan dari dataset pendaftaran anggota.

Tabel 1. Potongan Data Mentah Pendaftaran Anggota

MemberNo	Fullname	Tanggal_Jahir	umur	Kota	Alamat	Tingkat_Pendidikan	InstitutionName	Pekerjaan	Jenis_Kelamin
20010200001	LUSITANIA DIAN PEF	1997 November 22	22	GROBOGAN	GUBUG	D3	AKPER KESDAM IV/D	Mahasiswa	Wanita
20010200002	INDRIANA SAFITRI	1999 June 21	21	LAMPUNG	MULYOREJO II R'	D3	akper kesdam IV/Dip	Mahasiswa	Wanita
20010200003	ATIKA FEBRI DAMAY	1999 February 21	21	Kabupaten Boyolali	JEMBUNGAN RT D3		Akademi Keperawat	Mahasiswa	Wanita
20010200005	DEVIA RAHMATIKA	2001 November 18	18	Kabupaten Demak	wringinajar	S1	UIN Walisongo	Mahasiswa	Wanita
20010200006	M. ABDUL NAAJIB	1997 September 22	22	Kabupaten Rembang	Desa Trembes	SMA	UIN Walisongo Sema	Mahasiswa	Pria
20010200007	AHMAD FAISAL NAJJ	1996 May 31	24	Kabupaten Lamongan	TAMBAR. RT 003	S1	UIN Walisongo Sema	Mahasiswa	Pria
20010200008	LINDA AIDA SAFITRI	1997 April 10	23	brebes	dukuh penjalin	S1	uin walisongo	Mahasiswa	Wanita
20010200009	TANIA MARDIATI YU	1999 March 13	21	semarang	nolokerto	D3	AKPER KESDAM 4 DIF	Mahasiswa	Wanita
20010200010	KASNADI	1972 February 11	48	Kota Semarang	MARGOYOSO N	S1		Pegawai Sw	Pria
20010200011	HERI SUSANTO	1960 March 17	60	Kota Semarang	JL. TAMAN SEKA	S1	STTIH SEMARANG	Lainnya	Pria
20010200012	MUH. NURROHMAN	1990 February 23	30	Kabupaten Pemalan	Desa Susukan D	S2	UIN WALISONGO SEN	Mahasiswa	Pria

Dari dataset tersebut perlu dilakukan penyeleksian terhadap atribut karena tidak semua atribut digunakan jadi akan dilakuakn proses seleksi terhadap atribut. Atribut yang akan digunakan dalam proses data mining adalah Umur, Kota, Tingkat Pendidikan, Pekerjaan, dan Jenis Kelamin.

2. Pre-processing/Cleaning

Proses cleaning merupakan proses yang dimana membuang duplikasi data, memeriksa data yang tidak konsisten, memeriksa data yang kosong. Dalam penelitian ini jumlah data awal sebelum dibersihkan 2827 record dan setelah dibersihkan menjadi 2007 record.

3. Transformation

Transformasi merupakan Tahap perubahan data, setelah data di *transformasi* selanjutnya data akan dikonversi, data yang semula berupa kata diubah menjadi format angka agar lebih mudah dalam perhitungan nantinya. Berikut merupakan potongan dari hasil konversi data.

Tabel 2. Hasil Dataset Setelah Dikonversi

Umur	Kota	Tingkat_Pendidikan	Pekerjaan	Jenis_Kelamin
3	2	6	6	1
3	2	6	6	1
3	2	6	6	1
3	2	7	6	1
3	2	7	6	2
3	2	7	6	1
3	1	6	6	1
6	1	7	3	2
7	1	7	8	2

4. Data Mining

Pada penelitian ini menggunakan metode *clustering* dengan *algoritma K-Means* untuk menemukan informasi mengenai data pendaftaran anggota perpustakaan. Berikut merupakan tahapan dari K-Means.

- Menentukan banyaknyakluster
- Menentukan pusa kluster (centroid) awal secaraacak
- Setelah menentukan centroid, hitung jarak objek ke centroid menggunakan rumus *Euclidean Distance*. Dengan persamaan pada rumus1.

$$d(x,y)=\sqrt{\sum_{i=1}^n(x_i - y_i)^2} \quad (1)$$

Keterangan :

x_i : objek xke- i

y_i : objek yke- i

n : banyaknya objek

- Mengelompokan jarak terdekat suatu objek ke dalam masing-masingkluster
- Menentukan pusat kluster baru dengan menghitung rata-rata dari setiap anggota masing-masing kluster. Dengan persamaan pada rumus2.

$$C_k = \frac{1}{n_k} \sum d_i \quad (2)$$

Keterangan :

n_k : jumlah data dalam kluster k

d_i : jumlah dari nilai jarak dalam masing-masingkluster

- Ulangi langkah ke-3, jika perhitungan rata- rata suatu objek masih ada yang pindah atau belum stabil dan Perulangan berhenti jika objek sudahstabil.

5. Interpretation/Evaluation

Hasil dari analisa tersebut dapat memberikan kesimpulan mengenai kalangan masyarakat yang mendominasi pendaftaran tersebut dengan menggunakan metode ini diharapkan dapat mencapai suatu tujuan yang telah ditentukan.

4. HASIL DANPEMBAHASAN

4.1 PerhitunganManual

- Proses data mining menggunakan metode *Clustering* dengan algoritma *K-means*. Dalam penelitian ini jumlah *cluster* yang akan digunakan adalah 3 *cluster*. Berikut merupakan pusat cluster yang telah ditentukan secara acak.

Tabel 3. Pusat Cluster

Atribut	Cluster 1	Cluster 2	Cluster 3
Umur	2	3	5
Kota	1	2	1

Tingkat Pendidikan	3	7	8
Pekerjaan	7	6	2
Jenis Kelamin	1	1	2

2. Setelah menentukan titik pusat cluster, selanjutnya adalah menghitung jarak objek ke centroid dengan menggunakan rumus *Euclidean Distance*. Sebagai contoh perhitungan jarak pada datapertama

$$d(1,1) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + (x_4 - y_4)^2 + (x_5 - y_5)^2}$$

$$d(1,1) = \sqrt{(3 - 2)^2 + (2 - 1)^2 + (6 - 3)^2 + (6 - 7)^2 + (1 - 1)^2}$$

$$d(1,1) = \sqrt{12}$$

$$d(1,1) = 3,4641$$

$$d(1,2) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + (x_4 - y_4)^2 + (x_5 - y_5)^2}$$

$$d(1,2) = \sqrt{(3 - 3)^2 + (2 - 2)^2 + (6 - 7)^2 + (6 - 6)^2 + (1 - 1)^2}$$

$$d(1,2) = \sqrt{1}$$

$$d(1,2) = 1$$

$$d(1,3) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + (x_4 - y_4)^2 + (x_5 - y_5)^2}$$

$$d(1,3) = \sqrt{(3 - 5)^2 + (2 - 1)^2 + (6 - 8)^2 + (6 - 2)^2 + (1 - 2)^2}$$

$$d(1,3) = \sqrt{26}$$

$$d(1,3) = 5,09902$$

Perhitungan terus dilakukan pada seluruh data yang digunakan dalam penelitian.

3. Setelah menghitung jarak objek ke *centroid* menggunakan rumus *Euclidean Distance* selanjutnya mengelompokan jarak terdekat suatu objek kedalam masing-masing kluster, jadi objek yang memiliki jarak terdekat dengan titik pusat (*centroid*) maka akan masuk kedalam kelompok anggota *centroid* tersebut. Berikut ini hasil dari pengelompokan *cluster* berdasarkan jarakterdekat.

Tabel 4. Hasil Perhitungan Jarak *Euclidean Distance*

Data ke-	Umur	Kota	Tingkat Pendidikan	Pekerjaan	Jenis Kelamin	Jrk Ke K1	Jrk Ke K2	Jrk Ke K3	Cluster
1	3	2	6	6	1	3.464102	1	5.09902	C2
2	3	2	6	6	1	3.464102	1	5.09902	C2
3	3	2	6	6	1	3.464102	1	5.09902	C2
4	3	2	7	6	1	4.358899	0	4.795832	C2
5	3	2	7	6	2	4.472136	1	4.690416	C2
6	3	2	7	6	1	4.358899	0	4.795832	C2
7	3	1	6	6	1	3.316625	1.414214	5	C2
8	6	1	7	3	2	7	4.472136	1.732051	C3
9	7	1	7	8	2	6.557439	4.690416	6.403124	C2
10	4	2	8	6	2	5.656854	1.732051	4.242641	C2
...
710	3	2	4	4	1	3.464102	3.605551	5.09902	C1
...
2007	3	2	7	6	1	4.358899	0	4.795832	C2

4. Setelah mengelompokan jarak terdekat suatu objek kedalam masing-masing kluster, selanjutnya menghitung nilai rata-rata dari setiap anggota masing-masing kluster yang nantinya hasil rata-rata tersebut digunakan untuk pusat kluster baru. Berikut ini hasil dari perhitungan rata-rata *cluster* pada

iterasi pertama.

Tabel 5. Hasil Perhitungan Rata-rata Iterasi Pertama

Atribut	K1	K2	K3
Umur	1.77199	3.09053	4.47706
Kota	1.08360	1.74383	1.26606
Tingkat_Pendidikan	1.62758	6.83848	6.89908
Pekerjaan	6.80782	6.06173	2.6789
Jenis_Kelamin	1.40717	1.24897	1.48624

5. selanjutnya adalah menghitung jarak *euclidean* dengan cara yang sama pada perhitungan sebelumnya dengan menggunakan hasil dari perhitungan rata – rata iterasi pertama untuk pusat *cluster* baru. Jika terdapat data yang pindah atau belum stabil maka ulangi tahapan – tahapan diatas sampai data stabil dan tidak ada yang pindah. Pada penelitian ini perhitungan berhenti pada iterasi ke 4, karena tidak ada lagi data yang pindah atau sudah stabil.

Tabel 6. Hasil Akhir Dari perhitungan Cluster

Data ke-	Umur	Kota	Tingkat Pendidikan	Pekerjaan	Jenis Kelamin	Hasil cluster iterasi 1	Hasil cluster iterasi 2	Hasil cluster iterasi 3	Hasil cluster iterasi 4
1	3	2	6	6	1	C2	C2	C2	C2
2	3	2	6	6	1	C2	C2	C2	C2
3	3	2	6	6	1	C2	C2	C2	C2
4	3	2	7	6	1	C2	C2	C2	C2
5	3	2	7	6	2	C2	C2	C2	C2
6	3	2	7	6	1	C2	C2	C2	C2
7	3	1	6	6	1	C2	C2	C2	C2
8	6	1	7	3	2	C3	C3	C3	C3
9	7	1	7	8	2	C2	C2	C2	C2
10	4	2	8	6	2	C2	C2	C2	C2
...
710	3	2	4	4	1	C1	C2	C3	C3
...
2007	3	2	7	6	1	C2	C2	C2	C2

Dari perhitungan jarak *euclidean* pada iterasi keempat menunjukkan bahwa data sudah stabil. Berikut merupakan hasil dari perhitungan rata-rata pada iterasi keempat dari setiap anggota masing-masing *cluster*

Tabel 7. Hasil Perhitungan Rata- rata Iterasi Keempat

Atribut	K1	K2	K3
Umur	1.67128	3.08893	4.12139
Kota	1.06228	1.74354	1.32370
Tingkat_Pendidikan	1.54095	6.84074	5.56069
Pekerjaan	7.02999	6.07859	2.86705
Jenis_Kelamin	1.40830	1.25026	1.45665

Dari tabel diatas telah didapatkan hasil akhir dari proses *clustering*. Hasil yang didapatkan dari perhitungan *cluster* 1, 2, dan 3 sebagai berikut :

Tabel 8. Hasil Dari Perhitungan Cluster 1, 2, dan 3

KLUSTER 1			KLUSTER 2			KLUSTER 3		
ATRIBUT	BANYAK	JUMLAH	ATRIBUT	BANYAK	JUMLAH	ATRIBUT	BANYAK	JUMLAH
UMUR	6-11	460	UMUR	6-11	0	UMUR	6-11	0
	12-16	242		12-16	0		12-16	0
	17-25	160		17-25	912		17-25	61
	26-35	2867		26-35	37		26-35	56173
	36-45	1		36-45	8		36-45	33
	46-55	2		46-55	7		46-55	20
56-65	0	56-65	3	56-65	3			
KOTA	Dalam Kota	813	KOTA	Dalam Kota	248	KOTA	Dalam Kota	117
	Luar kota	54		Luar kota	719		Luar kota	56173
TINGKAT_PENDIDIKAN	SD	599	TINGKAT_PENDIDIKAN	SD	0	TINGKAT_PENDIDIKAN	SD	0
	SMP	68		SMP	0		SMP	1
	SMA	199		SMA	0		SMA	56
	D1	1		D1	1		D1	3
	D2	0867		D2	3		D2	0173
	D3	0		D3	166		D3	21
	S1	0		S1	777		S1	82
	S2	0		S2	19		S2	10
	S3	0		S3	1		S3	0
PEKERJAAN	PNS	0	PEKERJAAN	PNS	0	PEKERJAAN	PNS	10
	Pendidik	0		Pendidik	0		Pendidik	32
	Pegawai Swasta	0		Pegawai Swasta	0		Pegawai Swasta	102
	Wiraswasta	0		Wiraswasta	2		Wiraswasta	29
	Peneliti	2		Peneliti	2		Peneliti	0173
	Mahasiswa	0		Mahasiswa	922		Mahasiswa	0
	Pelajar	835		Pelajar	0		Pelajar	0
Lainnya	30	Lainnya	41	Lainnya	0			
JENIS_KELAMIN	Wanita	513	JENIS_KELAMIN	Wanita	725	JENIS_KELAMIN	Wanita	94
	Pria	354		Pria	242		Pria	79173

4.2 Implementasi Menggunakan RStudio

1. Eksekusi K-Means

Pada perhitungan K-Means dengan RStudio, dalam perhitungan ini ditentukan jumlah clusternya adalah tiga cluster.

```
final<-kmeans(datakmeans2, 3, nstart=25) final
```

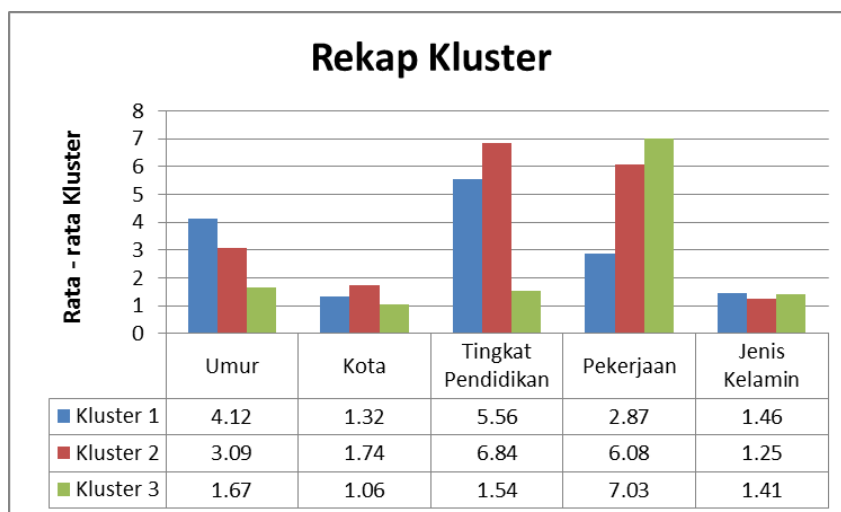
dari source code diatas menampilkan cluster membership, cluster means yang bisa dilihat pada gambar 2. Untuk clustering vector dan sum of squares bisa dilihat pada gambar 3.

```
K-means clustering with 3 clusters of sizes 173, 967, 867

Cluster means:
      umur      Kota Tingkat_Pendidikan Pekerjaan Jenis_Kelamin
1 4.121387 1.323699      5.560694  2.867052      1.456647
2 3.088935 1.743537      6.840745  6.078594      1.250259
3 1.671280 1.062284      1.540946  7.029988      1.408304
```

Gambar 2. Hasil Jumlah Pengelompokan Cluster dan Rata-rata Cluster

2. Analisa Rekap Cluster



5. KESIMPULAN

Hasil dari penelitian dengan menggunakan Clustering K-Means pada pendaftaran anggota Di Dinas Kearsipan dan Perpustakaan Provinsi Jawa Tengah dapat ditarik kesimpulan sebagai berikut :

1. Algoritma K-Means berhasil diimplementasikan pada pendaftaran anggota perpustakaan Di Dinas Kearsipan dan Perpustakaan Provinsi Jawa Tengah dengan jumlah data 2007 data dan dengan menggunakan 3 cluster. Untuk cluster 1 berjumlah 173 data, cluster 2 berjumlah 967 data dan cluster 3 berjumlah 867 data.
2. Dari hasil perhitungan tersebut dapat diketahui profil yang mendominasi pendaftaran anggota pada cluster 1 adalah berusia 17 sampai 25 tahun, bertingkat pendidikan S1, merupakan pegawai swasta, untuk cluster 2 yang mendominasi adalah berusia 17 sampai 25 tahun, bertingkat pendidikan S1, merupakan Mahasiswa. Dan untuk cluster 3 yang didominasi berusia 6 sampai 11 tahun, bertingkat pendidikan SD, yang merupakan pelajar.
3. Dari hasil keseluruhan dapat disimpulkan kembali bahwa profil kalangan masyarakat yang paling banyak mendaftar sebagai anggota adalah berusia 17 sampai 25 tahun, bertingkat pendidikan S1, yang merupakan Mahasiswa.

6. SARAN

Saran yang dapat peneliti berikan kepada penelitian selanjutnya adalah sebagai berikut :

1. Untuk penelitian selanjutnya dapat menggunakan variasi data yang lebih banyak agar hasil dari pengelompokannya tidak saling berdekatan antarcluster.
2. Hasil dari penelitian ini dapat digunakan oleh pihak Dinas Kearsipan dan Perpustakaan Provinsi Jawa Tengah untuk mengetahui profil dari kalangan yang mendominasi pendaftaran anggotaperpustakaan.
3. Pada penelitian selanjutnya dapat menggunakan metode optimasi untuk menentukan jumlah cluster terbaik seperti metode *silhouette*, metode *elbow* atau metode *gapstatistic*
4. Untuk penelitian selanjutnya juga dapat menggunakan metode dan algoritma yang lain seperti algoritma *AHC* atau *K-Medoid*

DAFTAR PUSTAKA

- [1] Sulastrri, H., & Gufron, A. I. (2017). Penerapan Data Mining Dalam Pengelompokan PenderitaThalassaemia. *Jurnal Nasional Teknologi dan Sistem Informasi*, 299-305
- [2] Sani, A. (2018). Penerapan Metode K- Means Clustering Pada Perusahaan.1-7
- [3] Mahmuda, F. M., & Kurniawan, D. E. (2017). Clustering Profil Pengunjung Perpustakaan. *Journal of Applied Informatics and Computing (JAIC)*, 14-21.
- [4] Kusrini, & Lutfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta:Andi..