

PENGELOMPOKKAN JUMLAH PENGUNJUNG PERPUSTAKAAN DI DKI JAKARTA MENGGUNAKAN METODE K-MEANS

Esi Farisan Kaisi. R¹, Rizki Tri Prasetyo²

¹Universitas Adhirajasa Reswara Sanjaya
Jl Sekolah Internasional No 1-2, Antapani, Bandung, (022) 7100124
e-mail: esifarisankaisirahmat@gmail.com

²Universitas Adhirajasa Reswara Sanjaya
Jl Sekolah Internasional No 1-2, Antapani, Bandung, (022) 7100124
e-mail: rizki@ars.ac.id

Abstrak

Perpustakaan adalah suatu ruangan pengoleksian buku yang berfungsi sebagai wahana pendidikan, penelitian, informasi dan rekreasi untuk meningkatkan kecerdasan dan keberdayaan bangsa. Salah satu jenis perpustakaan adalah perpustakaan daerah yang berkedudukan di tiap provinsi di Indonesia misalnya di DKI Jakarta. Banyaknya pengunjung perpustakaan yang tercatat pada daftar buku pengunjung perpustakaan dapat diolah menjadi sebuah data yang menghasilkan informasi profil pengunjung perpustakaan dan alamat pengunjung perpustakaan. Apabila dilakukan pengolahan data pada sumber data tersebut maka dapat diketahui berbagai informasi yang bermanfaat dalam membantu pengelompokan jumlah pengunjung perpustakaan. Oleh sebab itu, diperlukan pengelompokan jumlah pengunjung perpustakaan dengan menggunakan teknik data mining *clustering* dan metode *K-Means*. *K-Means* merupakan salah satu metode *Clustering nonhierarki* yang mempartisi data ke dalam bentuk satu atau lebih *cluster* sehingga data dengan karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama pula. Dari 276 data dibentuk 2 *cluster* yang telah dievaluasi menggunakan *Davies Bouldin Index* dengan nilai 0.363 yang mana *cluster* 0 menjadi *cluster* rendah dengan jumlah anggota sebanyak 74 data sedangkan *cluster* 1 menjadi *cluster* tinggi dengan jumlah anggota 202 data

Kata Kunci: *Clustering, K-Means, Perpustakaan*

Abstract

A library is a room for collecting books that functions as a vehicle for education, research, information and recreation to increase intelligence and empowerment of the nation. One type of library is a regional library which is located in each province in Indonesia, for example in DKI Jakarta. The number of library visitors who are recorded in the library visitor book list can be processed into data that produces information on the profile of library visitors and addresses of library visitors. If the data is processed at the data source, various information that can be useful in helping to group the number of library visitors can be identified. Therefore, it is necessary to group the number of library visitors using data mining clustering techniques and the K-Means method. K-Means is a non- hierarchical clustering method which partitions data into one or more clusters so that data with the same characteristics are grouped into the same cluster. From 276 data, 2 clusters were formed which were evaluated using the Davies Bouldin Index with a value of 0.363, in which cluster 0 became a low cluster with 74 data members while cluster 1 became a high cluster with 202 data members.

Keywords: *Clustering, K-Means, Library*

1. Pendahuluan

Perpustakaan adalah suatu ruangan pengoleksian buku, berfungsi sebagai wahana pendidikan, penelitian, informasi dan rekreasi untuk meningkatkan kecerdasan dan keberdayaan bangsa. (Arumsari & Krismayani, 2016). Perpustakaan sendiri memiliki beberapa jenis, salah satunya adalah perpustakaan daerah yang berkedudukan di tiap provinsi di Indonesia. DKI Jakarta juga memiliki perpustakaan daerah yaitu, Kantor Perpustakaan Umum dan Arsip Kota DKI Jakarta. Yang bertujuan untuk menyediakan fasilitas dan sumber informasi dan menjadi pusat pembelajaran.

Banyaknya pengunjung perpustakaan yang tercatat pada daftar buku pengunjung perpustakaan dapat diolah menjadi sebuah data. Apabila dilakukan pengolahan data pada sumber data tersebut maka dapat diketahui berbagai informasi yang bermanfaat dalam membantu pengelompokan jumlah pengunjung perpustakaan (Bakker & Penelitian, 2020).

Karena itu, dibutuhkan suatu teknik ataupun metode untuk mengolah data tersebut. Selain itu, pengolahan data ini dapat dijadikan sebagai bahan evaluasi pihak perpustakaan dalam merencanakan fasilitas baru, penambahan buku baru, dan lain sebagainya. Metode pengolahan data tersebut sering disebut dengan *data mining* yang cara kerjanya mencari dan mengelompokkan data sehingga proses pencarian bentuk pola yang tersembunyi dapat di ambil (Dan, 2019). Salah satu teknik yang dikenal dalam data mining yaitu *clustering*. *Clustering* dalam data mining adalah pengelompokan sejumlah data atau objek ke dalam cluster (group).

Algoritma yang digunakan pada *clustering* ini adalah algoritma *K-Means*, teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. Metode Algoritma *K-Means* adalah metode *Clustering* yang paling populer digunakan karena beberapa kelebihan yang dimiliki antara lain algoritma ini sederhana dan mudah diterapkan (Yanto et al., 2019). Algoritma *K-Means* yang didasarkan pada penelitian terdahulu yang dilakukan oleh (Mahmuda et al., 2017) mengenai *Clustering* Profil Pengunjung Perpustakaan.

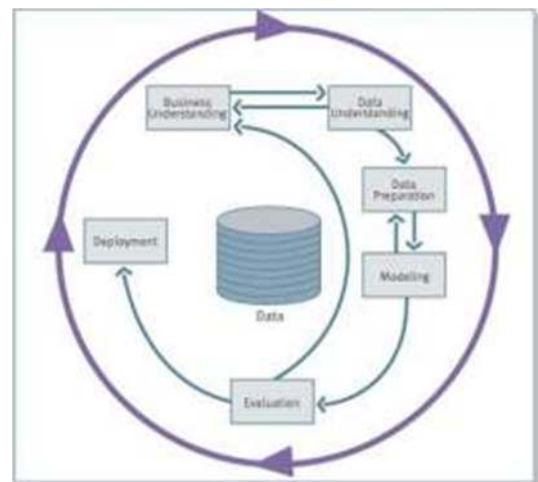
Berdasarkan penelitiannya menunjukkan bahwa algoritma *K-Means* dapat melakukan pengelompokan jumlah pengunjung Perpustakaan BP Batam.

2. Metode Penelitian

2.1. Perancangan Penelitian

Pada penelitian eksperimen ini, digunakan model CRISP-DM (*Cross-Standard Industry Process for Data Mining*) yang terdiri dari 6 fase atau tahapan. Berikut 6 fase CRISP-DM menurut Larose dalam (Setiawan & Tes, 2016), yaitu :

Gambar 1. Tahapan Penelitian



2.2. Business Understanding

Perpustakaan adalah suatu instansi yang mengumpulkan berbagai pengetahuan, selain itu perpustakaan merupakan salah satu pusat informasi, sumber ilmu pengetahuan, penelitian, dan rekreasi. Di daerah DKI Jakarta jumlah pengunjung perpustakaan masih cukup tinggi sehingga di butuhkan suatu teknik pengelompokan daerah mana yang paling banyak di kunjungi.

2.3. Data Understanding Phase

Data pengunjung perpustakaan diperoleh dari *website* resmi pemerintah DKI Jakarta <http://data.jakarta.go.id>. Data terdiri dari jumlah pengunjung perpustakaan di daerah DKI Jakarta pada tahun 2019. Dari data jumlah pengunjung perpustakaan terdapat atribut jenis, perpustakaan, jumlah pengunjung.

2.4. Data Preparation Phase

Data yang telah diperoleh akan diolah terlebih dahulu agar dapat dilakukan proses *clustering*. Pada tahap ini dilakukan tahap *integration* atau penggabungan data dan pembersihan data, proses pengolahan data tersebut

dilakukan pada *Microsoft Excel*.

A. Transformasi Data

Tahap transformasi adalah tahapan mengubah format data asli menjadi bentuk yang sesuai untuk mempermudah proses penambangan data. Data yang berjenis nominal di olah menjadi numerik. Perubahan ini dilakukan pada *Micorosoft Excel*.

Tabel 1. Data Transformasi

tahun	bulan	jenis	perpustakaan	jumlah
2019	1	1	1	5797
2019	1	1	4	5191
2019	1	1	12	2459
2019	1	1	10	1509
2019	1	1	11	32754
2019	1	1	18	582
2019	1	1	16	27078
2019	1	1	17	32081
2019	1	1	9	4523
2019	1	1	7	19634
2019	1	1	8	43753
2019	1	1	15	2155
2019	1	1	13	8259
2019	1	1	14	39

B. Data Cleaning

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang akan diolah. Proses *cleaning* memeriksa data yang inkonsisten, menghapus atau memperbaiki kesalahan pada data, seperti kesalahan penulisan. Pada tahap ini dilakukan pemilihan atribut yang tidak digunakan, maka atribut yang tidak digunakan yaitu tahun dan bulan.

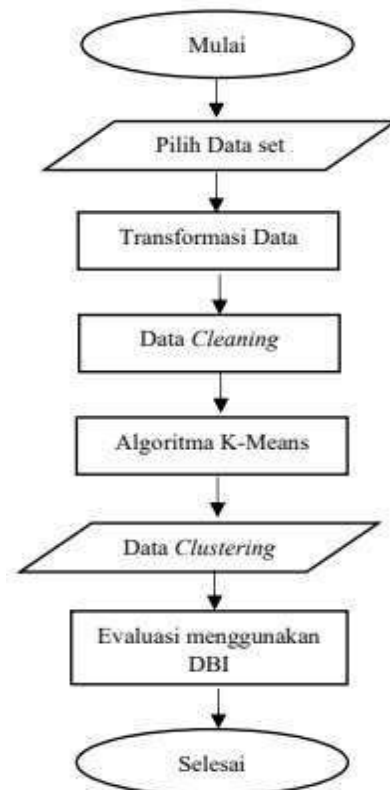
C. Populasi

Populasi adalah wilayah generalisasi yang terdiri dari objek/subjek yang memiliki kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulannya. Dalam penelitian ini, penulis mengambil data dari situs <http://data.jakarta.go.id>, dengan data sebanyak 276.

2.5. Modeling Phase

Pada tahap ini ditentukan model yang akan digunakan dalam penelitian. Pemodelan penelitian ini menggunakan Algoritma *K-*

Means Clustering. Berikut *flowchart* model yang di usulkan :



Gambar 2. Model Yang Diusulkan

2.6. Evaluation Phase

Dalam tahap ini akan dilakukan tahap evaluasi menggunakan model *davies-bouldin index* (DBI). Pengukuran dengan *Davies-Bouldin Index* ini memaksimalkan jarak inter-cluster dan pada waktu yang sama mencoba untuk meminimalkan jarak antar titik dalam sebuah cluster. Semakin kecil nilai DBI yang diperoleh atau semakin mendekati nilai 0, maka semakin baik *cluster* yang diperoleh dari pengelompokan algoritma yang digunakan.

3. Hasil dan Pembahasan

3.1. Perhitungan Manual Algoritma *K-Means*

Dalam penelitian ini dilakukan perhitungan manual algoritma *K-Means* dengan menggunakan *Microsoft Excel*. Adapun langkah-langkah perhitungan algoritma *K-Means* adalah sebagai berikut:

1. Tentukan banyaknya *K*, dimana *K* adalah jumlah *Cluster* yang akan dibentuk. Pada penelitian ini dibentuk *cluster* dengan jumlah *k=2*.
2. Tetapkan titik pusat *Cluster* secara acak atau *random*, titik pusat *Cluster*

sering disebut dengan nama centroid. Berikut adalah titik pusat *cluster* atau centroid awal yang dipilih secara *random*:

Tabel 2. Centroid Awal

Cluster	Jenis	Perpustakaan	Jumlah
C0	1	14	40945
C1	1	9	2921

3. Hitung setiap jarak data yang ada terhadap masing-masing centroid dengan menggunakan rumus *Euclidean Distance*

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Keterangan :

De= Euclidean Distance

i= banyaknya objek,

(x,y)=merupakan koordinat objek, dan

(s,t)=merupakan koordinat centroid (titik pusat *Cluster*).

Setelah menentukan centroid awal, hitung setiap data ke masing-masing centroid dengan menggunakan rumus *Euclidean Distance*. Berikut merupakan contoh perhitungan pada data ke-1:

Data pertama ke *cluster* 0:

$$d(x_1c_0)$$

$$= \sqrt{(1 - 1)^2 + (1 - 14)^2 + (5797 - 40945)^2}$$

$$= 35148$$

Data pertama ke *cluster* 1:

$$d(x_1c_1)$$

$$= \sqrt{(1 - 1)^2 + (1 - 9)^2 + (5797 - 2921)^2}$$

$$= 2876.01$$

Dilakukan pada semua data sehingga setiap data memiliki nilai centroid masing-masing.

4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan masing-masing centroid.

Pada tahap ini, cari jarak terdekat untuk menentukan cluster apa yang diikuti oleh data tersebut. Berikut merupakan tabel pengelompokan data pada iterasi 1 berdasarkan jarak minimum untuk menentukan cluster yang diikuti:

Tabel 3. Hasil Perhitungan Iterasi-1

Data Ke-	C0	C1	Jarak Minimum	Cluster
1	35148	2876.01	2876.011	1
2	35754	2270.01	2270.006	1
3	38486	462.01	462.0097	1
4	39436	1412	1412	1
5	8191	29833	8191.001	0
6	40363	2339.02	2339.017	1
7	13867	24157	13867	0
8	8864	29160	8864.001	0
....
276	39256	1232.02	1232.015	1

5. Tentukan *centroid* baru dengan menggunakan persamaan 2.2

Berdasarkan perhitungan pada iterasi ke-1, terdapat 68 data pada *cluster* 0 dan 208 data pada *cluster* 1. Berikut tabel anggota pada *cluster* 0 dan *cluster* 1:

Tabel 4. Anggota *Cluster* 0 Iterasi ke-1

Data Ke-	Jenis	Perpusatakaan	Jumlah
6	1	11	32754
8	1	16	27078
9	1	17	32081
.....
272	1	37	34115
Jumlah Data : 68			
Rata-rata	16.30882	32889.57	1.057692

Tabel 5. Anggota *Cluster* 1 Iterasi ke-1

Data Ke-	Jenis	Perpusatakaan	Jumlah
1	1	1	5797
2	1	4	5191
3	1	12	2459
.....
276	2	3	1689
Jumlah Data : 208			
Rata-rata	1.057692	13.37019	4771.822

Dari hasil perhitungan diatas, didapatkan centroid baru yang akan dijadikan sebagai titik pusat perhitungan berikutnya. Berikut merupakan tabel centroid baru:

Tabel 6. Centroid Baru

Cluster	Jenis	Perpustakaan	Jumlah
C0	16.30882	32889.57	1.057692
C1	1.057692	13.37019	4771.822

6. Kembali ke langkah no 3, 4 dan 5.

Jika pada iterasi kedua, anggota *Cluster* tidak ada yang berpindah ke *Cluster* lain maka iterasi berhenti tetapi jika ada anggota *Cluster* yang berpindah ke *Cluster* lain maka kembali ke langkah nomor 3, 4 dan 5. Lakukan iterasi berikutnya sampai tidak ada anggota *Cluster* yang berpindah ke *Cluster* lain.

Pada penelitian ini, iterasi berhenti sampai iterasi 3. Berikut merupakan hasil akhir dari perhitungan cluster pada iterasi ke-3:

Tabel 7. Hasil Cluster Iterasi ke-3

Data Ke-	C0	C1	Jarak Minimum	Cluster
1	26503.6	1644.626	1644.626	1
2	27109.6	1038.621	1038.621	1
3	29841.6	1693.42	1693.42	1
4	30791.6	2643.422	2643.422	1
5	453.4324	28601.58	453.4324	0
6	31718.6	3570.423	3570.423	1
7	5222.6	22925.58	5222.6	0
8	219.6008	27928.58	219.6008	0
9	27777.6	370.6043	370.6043	1
.....
276	30611.6	2463.441	2463.441	1

Dari hasil perhitungan manual yang telah dilakukan, didapatkan hasil *clustering* dengan jumlah *cluster* 0 = 74 items dan *cluster* 1 = 202 items.

3.2. Perhitungan Algoritma K-Means dengan Rapidminer

Pada tahap ini pemodelan dilakukan dengan menggunakan *Rapidminer* versi 9.6 dengan langkah sebagai berikut:

1. Buka aplikasi *Rapidminer* lalu pilih *new process*
2. Pilih *read excel* pada panel operator lalu drag ke panel *process*
3. Pilih data *excel* yang akan diuji pada *excel file*
4. Pilih Metode *K-Means* pada panel operator lalu drag ke panel *process*
5. Tentukan jumlah *k* yang akan digunakan
6. Sambungkan *connector* pada masing-masing proses lalu klik *Run*

**Gambar 3. Sambungkan Connector**

7. Setelah proses selesai maka hasil *clustering* akan ditampilkan

**Gambar 4. Hasil Clustering**

Berdasarkan hasil perhitungan *rapidminer* yang telah dilakukan, didapatkan *cluster* 0 sebanyak 74 data dan *cluster* 1 sebanyak 202 data.

4. Kesimpulan

Berdasarkan hasil *clustering* dengan menggunakan metode *K-Means* yang telah dilakukan menggunakan *Rapidminer* dan perhitungan manual, maka dihasilkan kesimpulan sebagai berikut:

1. Jumlah *cluster* yang didapat ada 2 *cluster* yaitu, *cluster* 0 dan *cluster* 1.
2. *Cluster* 0 merupakan *cluster* rendah yang jumlah anggotanya sebanyak 74.
Perpustakaan yang paling rendah dikunjungi yaitu Sudin Pusip Jakbar (Pusling), Sudin Pusip Jakpus (Pusling) dan Sudin Pusip Jaktim (Pusling), yang didominasi oleh pengunjung. Rata-rata jumlah pengunjung yang datang sebanyak 25580.
3. *Cluster* 1 merupakan *cluster* tinggi yang jumlah anggotanya sebanyak 202.

Perpustakaan yang paling tinggi dikunjungi yaitu KPAK Pulau 1000 (RPTRA), KPAK Jakpus (Pusling) dan Cikini, yang didominasi oleh pengunjung dan pengguna. Rata-rata jumlah pengunjung yang datang sebanyak 87180.

Referensi

- Arumsari, R., & Krismayani, I. (2016). *Peran Perpustakaan Keliling Dalam Menumbuhkan Minat Baca Masyarakat Desa Kepek Kecamatan Saptosari Kabupaten Gunungkidul*. <https://ejournal3.undip.ac.id/index.php/jip/article/view/15333>
- Bakker, E., & Penelitian, A. O. (2020). *Implementasi Data Mining Clustering Data Perpustakaan Menggunakan Algoritma K-Means untuk Menentukan Penambahan Koleksi Buku di Perpustakaan UPY*. 22–25. <http://prosiding.senadi.upy.ac.id/index.php/senadi/article/view/118>
- Dan, P. (2019). *Pengelompokan dan pemetaan derajat kesehatan kota bengkulu dengan metode k-means clustering*. 7(1), 91–97. <http://ejournal.unib.ac.id/index.php/rekursif/article/view/7097>
- Mahmuda, F., Sitorus, M. A. R., Widyastuti, H., & Kurniawan, D. E. (2017). Clustering Profil Pengunjung Perpustakaan (Studi Kasus Perpustakaan BP Batam). *Journal Of Applied Informatics and Computing*, 1(1), 14–21. <http://jurnal.polibatam.ac.id/index.php/JAIC/article/view/476>
- Setiawan, R. (2016). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik Lp3i Jakarta). *J. Lentera Ict*, 3(1), 76–92.
- Yanto, D., Probolinggo, L., Loyal, C., & Loyal, K. (2019). *Analisis RFM dan Algoritma K-Means untuk Clustering Loyalitas Customer*. 9(1), 0–8. <http://ejournal.upm.ac.id/index.php/energy/article/view/459>