

ANALISIS CLUSTER DENGAN ALGORITMA K-MEANS UNTUK CLUSTERING DATA DONATUR LEMBAGA AMIL ZAKAT

Oleh :

¹Hotmaida Lestari Siregar

¹Program Studi Pendidikan Vokasional Informatika

¹FPMIPA, Institut Pendidikan Tapanuli Selatan

Email : hotmaidalearisiregar@gmail.com

Abstrak

K-Means *Clustering* merupakan metode untuk menghasilkan k *Cluster*, yang k didefinisikan oleh pengguna. Analisis *cluster* adalah salah satu metode analisis multivariat yang tujuannya untuk mengelompokkan suatu objek ke dalam suatu kelompok berdasarkan karakteristik tertentu. Dalam analisis *cluster* menentukan jumlah *cluster* awal sangat penting agar *cluster* yang dihasilkan juga optimal. Dalam penelitian ini, akan dilakukan analisis jumlah *cluster* yang optimal terhadap klasifikasi data menggunakan metode K-Means. Data dianalisis menggunakan model RFM dan dilakukan analisis perbandingan berdasarkan nilai DBI dan *cluster compactness* yang dinilai dari *average silhouette score*. Metode K-Means menghasilkan nilai DBI terkecil sebesar 0,485 dan nilai *average silhouette score* tertinggi 0,781 berada pada jumlah $k=6$. Hasil menunjukkan metode terbaik untuk *clustering* data donatur adalah menggunakan metode K-Means dengan jumlah *cluster* optimal sebanyak 6 *cluster*.

Kata Kunci: K-Means; Clustering; RFM Model; DBI; *Average Silhouette Score*.

1. PENDAHULUAN

Clustering merupakan proses pengelompokan sehingga semua anggota kelompok dari setiap partisi memiliki persamaan berdasarkan matriks tertentu, yang tujuannya adalah untuk mengelompokkan objek ke dalam kelompok yang memiliki sifat berbeda antar kelompok, sehingga objek yang terletak dalam satu kelompok tersebut memiliki sifat yang relatif homogen (Talaku et al., 2). Konsep dasar *Clustering* yaitu mengelompokkan objek ke dalam *Cluster*, yang dimana *Cluster* yang paling baik merupakan *Cluster* yang memiliki tingkat kesamaan yang tinggi antar objek dalam satu *Cluster* dan tingkat ketidaksamaan yang tinggi pada objek *Cluster* lainnya (Ali, 2019). Analisis klaster merupakan salah satu metode analisis multivariat yang bertujuan untuk mengelompokkan objek-objek ke dalam suatu kelompok berdasarkan karakteristik tertentu. Dalam pengelompokannya, ukuran jarak diperlukan untuk mengukur kemiripan antar objek. Ukuran jarak yang biasa digunakan adalah jarak Euclidien (Johnson & Dean, 2007). *Clustering* juga diartikan sebagai teknik analisis data eksploratif yang digambarkan sebagai pengelompokan objek, dimana objek-objek tersebut memiliki kesamaan karakteristik. Analisis *Cluster* atau biasa disebut sebagai “pengelompokan” merupakan teknik yang digunakan untuk mengelompokkan pengamatan, titik data atau vektor fitur yang serupa berdasarkan kesamaan karakteristiknya (Govender & Sivakumar,

2020).

Penelitian ini akan dilakukan untuk menganalisis jumlah *Cluster* yang paling optimal terhadap klasifikasi data. Menggunakan algoritma data mining untuk melakukan *Clustering* (pengelompokan data) menggunakan variabel yang digunakan sebagai parameter penelitian. Metode yang digunakan pada penelitian ini adalah metode K-means *Clustering*. *Clustering* adalah metode untuk mengelompokkan objek sedemikian rupa sehingga objek yang serupa ditempatkan di dalam satu kelompok dan objek yang berbeda ditempatkan ke dalam kelompok lain. Salah satu pendekatan yang populer dalam *Clustering* adalah k-means. K-means *Clustering* merupakan metode untuk menghasilkan k *Cluster*, yang k didefinisikan oleh pengguna. Grup dipilih dengan mencari jumlah terkecil dari kesalahan kuadrat (SSE) dalam perbandingan rata-rata (Adi et al., 2018). Selanjutnya, setelah mendapatkan pengelompokan donatur dari hasil *Clustering*, akan dilakukan analisis *Cluster* dari metode tersebut. Pada penelitian ini, model yang akan digunakan adalah model RFM (*Recency, Frequency, Monetary*) yang digunakan dalam pengelompokan waktu transaksi terakhir, frekuensi transaksi dan jumlah pendapatan yang diperoleh oleh perusahaan. Model ini lebih mudah digunakan dan juga mudah dimengerti dalam pengambilan keputusan oleh perusahaan (Astuti, 2019).

2. METODE PENELITIAN

Data yang akan digunakan dalam penelitian ini

adalah data yang diambil secara primer dari lembaga zakat NU Care LAZISNU Padangsidempuan-SUMUT. Pengambilan data donatur dari lembaga amal zakat berupa file Microsoft excel yang berisi atribut Nomor kotak, nama rumah makan, alamat, kontak, status kontak, tanggal dan total. Setelah pengambilan data, akan dilakukan pembersihan data yang *redundant*, kurang jelas dan tidak lengkap. Data pra proses tersebut tipe data mentah untuk dilanjutkan pada proses berikutnya. Tujuannya adalah menyaring data dari atribut yang tidak digunakan dan membuat data menjadi suatu format yang lebih mudah dan efektif. Selanjutnya data yang sudah dilakukan pembersihan pada pra-proses data, akan diubah ke bentuk model RFM. Atribut yang akan digunakan menjadi 3 atribut yaitu *recency*, *frequency* dan *monetary*. Nilai *Recency* didapat dari selisih jumlah hari antara tanggal penelitian dan tanggal terakhir pengumpulan donasi. Nilai *frequency* dihasilkan dari seberapa sering pengumpulan donasi selama periode analisis. Nilai *Monetary* dihasilkan dari jumlah besaran uang yang dihasilkan selama periode analisis. Berikut hasil dari analisis RFM seperti di bawah ini.

N	F	P
REGENCY	FREQUENCY	MONETARY
184	6	12100
184	6	75100
184	6	464900
184	6	913000
184	6	169400
184	6	24600
184	6	574700
184	6	58300
184	6	0
184	6	41000
184	6	10000
184	6	28700
184	6	49500
184	6	8500
184	6	14800
184	6	28000
184	6	29600
184	6	25000
184	6	147900
184	6	10700
184	6	171300
184	6	44600

Proses melakukan *Clustering* dengan menggunakan K-Means rumus sebagai berikut:

- 1) Tentukan k sebagai jumlah kluster yang akan dibentuk.
- 2) Menentukan nilai random untuk pusat *Cluster* awal (*Centroid*)
- 3) Menghitung jarak setiap input terhadap *Centroid* menggunakan rumus jarak *Euclidean* hingga ditemukan jarak yang paling dekat dari setiap data *Centroid*.
- 4) Mengklasifikasikan data berdasarkan kedekatan dengan *Centroid*
- 5) Memperbarui nilai *Centroid*. Ulangi langkah 2 sampai 5, hingga anggota tiap *Cluster* tidak ada yang berubah. Jika langkah 6 telah selesai, maka nilai pusat

Cluster pada iterasi terakhir akan digunakan sebagai parameter untuk klasifikasi data

3. HASIL DAN PEMBAHASAN

Setelah dilakukan proses clustering, dihasilkan anggota- anggota dari tiap cluster, yang akan ditunjukkan pada tabel berikut ini.

Cluster	Jumlah Anggota
0	171
1	818
2	24
3	94
4	82
5	72

Pada tabel di atas, jumlah anggota dalam cluster 0 ada 171 donatur, pada cluster 1 berjumlah 818 donatur, cluster 2 berjumlah 24 donatur, cluster 3 berjumlah 94 donatur, cluster 4 berjumlah 82 donatur dan pada cluster 5 berjumlah 72 donatur.

Proses pengujian cluster akan dilakukan validasi cluster menggunakan nilai *davies ibouldin index* dan *average silhouete coefficient* yang akan digunakan untuk mengetahui kekuatan dari suatu cluster. Semakin nilai DBI mendekati 0, maka semakin baik pula cluster yang dihasilkan. *Silhouette score* berada pada kisaran -1 maka hasil pengelompokan tersebut sangat buruk dan angka 1 menunjukkan bahwa kualitas hasil pengelompokan cluster adalah Kuat. Nilai DBI dan *average silhouete score* pada pengujian menggunakan k-means akan ditunjukkan pada tabel berikut.

K-Means		
K	DBI	Average silhouete score
2	0,551	0,676
3	0,513	0,683
4	0,540	0,712
5	0,514	0,797
6	0,464	0,815

Berdasarkan tabel di atas dapat disimpulkan bahwa nilai DBI yang paling rendah dan mendekati 0 (bernilai positif) adalah 0,464 dan nilai *Average silhouete score* yang paling tinggi adalah 0,815. Dengan demikian, jumlah cluster paling optimal pada metode k-means untuk clustering data donatur adalah jumlah k= 6.

4. KESIMPULAN

Bagian Berdasarkan hasil penelitian yang dilakukan, penulis menarik kesimpulan yang didasarkan pada pengujian *Cluster* menggunakan K-Means *Clustering* adalah sebagai berikut:

1. Berdasarkan hasil *clustering*, jumlah *cluster* yang terbentuk adalah sebanyak 6 *cluster* dari keseluruhan data sebanyak 1261 donatur

2. Untuk pengujian nilai DBI, dihasilkan masing-masing nilai DBI dari metode K-Means sebesar 0,4641. *Cluster compactness* dilakukan dengan menilai dari *average silhouette score*. Dalam hal ini *average silhouette score* metode K-Means adalah 0,815
3. *Cluster* yang terbentuk adalah *Cluster* pertama berjumlah 171 donatur, *Cluster* kedua berjumlah 818 donatur, *Cluster* ketiga berjumlah 24 donatur, *Cluster* keempat berjumlah 94 donatur, *Cluster* 5 berjumlah 82 donatur dan *Cluster* 6 berjumlah 72 donatur.
4. Hasil analisis *Clustering* berdasarkan karakteristiknya membentuk 6 *Cluster* yaitu:
 - a. *Cluster 0 : Everyday Shopper*
 - b. *Cluster 1 : Golden Customer*
 - c. *Cluster 2 : Superstar*
 - d. *Cluster 3 : Occasional Customer*
 - e. *Cluster 4 : Dormant Customer*
 - Cluster 5 : Typical Customer*

Terapan, 11(2), 119-128.

<https://doi.org/10.30598/barekengvol11iss2pp119-1>

5. REFERENSI

- Adi, S., Er, M., & Nugroho, B. (2018). ScienceDirect ScienceDirect Curriculum Assessment of Higher Educational Institution Using Aggregate Profile Clustering. *Procedia Computer Science*, 124, 264-273.
<https://doi.org/10.1016/j.procs.2017.12.155>
- Ali, A. (2019). Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 19(1), 186-195.
<https://doi.org/10.30812/matrik.v19i1.529>
- Astuti, R. D. (2019). Analisis Perbandingan Algoritma K-Means Dan K-Medoids Untuk Menerapkan Segmentasi Pelanggan (Issue 14002140).
- Govender, P., & Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980-2019). In *Atmospheric Pollution Research* (Vol. 11, Issue 1). Turkish National Committee for Air Pollution Research and Control.
<https://doi.org/10.1016/j.apr.2019.09.009>
- Johnson, W., & Dean, R. (2007). Clustering, Distance Methods, and Ordination. In *Applied Multivariate Statistical Analysis* (pp. 671-757).
- Talakua, M. W., Leleury, Z. A., & Taluta, A. W. (2017). Analisis Cluster Dengan Menggunakan Metode K-Means Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014. *BAREKENG: Jurnal Ilmu Matematika Dan*