

KLASTERISASI DATA JAMAAH UMROH PADA AULIYA TOUR & TRAVEL MENGUNAKAN METODE *K-MEANS CLUSTERING*

Muhammad Iqbal

Sistem Informasi, STMIK Royal Kisaran

email: muhammad.iqbal@gmail.com

Abstract: Religious tourism especially for Hajj and Umroh is increasingly in demand by the public. Auliya Tour & Travel, which is engaged in the travel agency, has diverse pilgrimage data so that the data collection is used to find new knowledge as a marketing strategy using Data Mining techniques. Data Mining is one of the KDD processes that has the function for grouping. *K-Means Clustering* is a data mining technique that aims to group data into a data subset. The grouping of data on Umroh pilgrims is conducted to find out the interest groups of pilgrims based on age. This study categorizes pilgrim data into three *clusters*, which are very popular, in high demand and less desirable. Attributes used in processing data include gender, age, and congregation package. Before the data calculation process is carried out, the data transformation process is carried out before data calculation process. Based on data calculations that have been done through the RapidMiner software, the members of the group were very interested, ranging in age from 56 to 83 years, interested groups ranging in age from 29 to 55 years and the group was less interested with an age range 2 to 22 years from 170 records.

Keywords: *Clustering*, Data Mining, K-Means, RapidMiner, Umroh.

Abstrak: Perjalanan wisata religi khususnya untuk ibadah haji dan umroh semakin diminati masyarakat. Auliya Tour & Travel yang bergerak pada bidang biro perjalanan memiliki data jamaah yang beragam sehingga kumpulan data tersebut dimanfaatkan untuk menemukan pengetahuan yang baru sebagai strategi pemasaran dengan menggunakan teknik Data Mining. Data Mining merupakan salah satu proses dari KDD yang fungsi salah satunya untuk pengelompokan. *K-Means Clustering* merupakan salah satu teknik Data Mining yang bertujuan untuk mengelompokkan data ke dalam subset data. Pengelompokan data jamaah umroh dilakukan bertujuan untuk mengetahui kelompok minat jamaah berdasarkan usia. Penelitian ini mengelompokkan data jamaah menjadi tiga *cluster* yaitu sangat diminati, diminati dan kurang diminati. Atribut yang digunakan dalam pengolahan data meliputi jenis kelamin, usia, dan paket jamaah. Sebelum proses perhitungan data dilakukan terlebih dahulu dilakukan proses transformasi data. Berdasarkan perhitungan data yang telah dilakukan melalui software RapidMiner diperoleh anggota kelompok sangat diminati dari rentang usia mulai 56 sampai 83 tahun, kelompok diminati dengan rentang usia mulai 29 sampai 55 tahun dan kelompok kurang diminati mulai usia 2 sampai 22 tahun dari 170 record.

Kata Kunci: *Clustering*, Data Mining, K-Means, RapidMiner, Umroh.

PENDAHULUAN

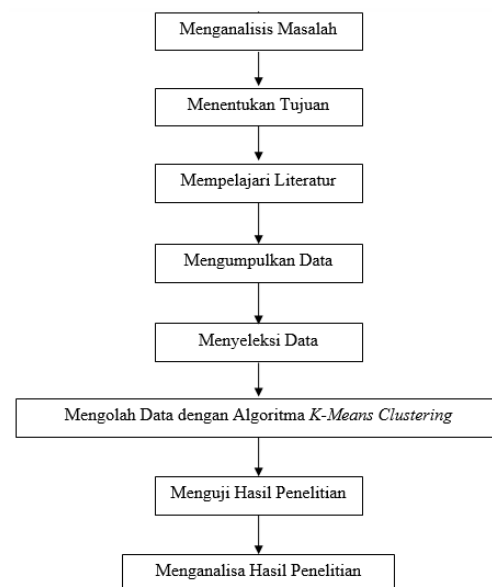
Perkembangan teknologi saat ini semakin pesat dan tidak terbendung. Semakin mudahnya mendapatkan informasi dan proses transaksi data yang cepat menjadi volume data semakin besar. Pertumbuhan data yang sangat cepat dalam waktu yang relatif singkat menjadikan penumpukan data di dalam *database*. Perkembangan pengetahuan dan teknologi memanfaatkan data yang besar tersebut menjadi suatu pengetahuan baru yang dikenal dengan teknik *Data Mining*.

Meningkatnya jumlah data jamaah umroh pada Auliya Tour & Travel tentunya memiliki variasi usia yang berbeda-beda. Pengelompokan usia jamaah merupakan salah satu tujuan untuk meningkatkan kualitas pelayanan. Pengelompokan usia ini bertujuan untuk mengetahui minat keberangkatan umroh pada Auliya Tour & Travel berdasarkan usia sehingga dibentuk tiga *cluster*, yaitu Sangat Diminati, Diminati dan Kurang Diminati berdasarkan usia jamaah. Permasalahan yang muncul adalah pihak perusahaan kesulitan dalam mengelompokkan usia jamaah umroh. Berdasarkan latar belakang ini penulis tertarik untuk melakukan penelitian menggunakan *K-Means Clustering* sebagai metode pengelompokan dengan menghitung jarak menggunakan rumus *Euclidean* di mana data dikelompokkan berdasarkan karakteristik yang sama akan dimasukkan pada kelompok yang sama sehingga data usia jamaah terkelompok dan dimasukkan ke dalam kelompok tidak tumpang tindih.

METODE

Metodologi penelitian merupakan sistematis keseluruhan tahapan yang akan dilaksanakan selama penelitian. Metodologi berperan penting untuk menjadikan baik atau tidaknya suatu penulisan karya ilmiah. Untuk memperoleh hasil penelitian yang diharapkan maka diperlukan melakukan studi pustaka sebagai referensi dalam penelitian.

Adapun kerangka kerja penelitian ini dapat digambarkan pada Gambar 1.



Gambar 1. Kerangka Kerja Penelitian

Knowledge Discovery in Database

KDD merupakan proses mengekstrak pola atau model dari data dengan menggunakan suatu algoritma yang spesifik. Adapun proses KDD sebagai berikut [1].

1. Data Selection

Data operasional perlu dilakukan sebelum tahap penggalian informasi

dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses Data Mining disimpan dalam suatu berkas, terpisah dari basis data operasional [1]

2. Pre-processing

Sebelum Data Mining dapat dilaksanakan, perlu dilakukan proses cleaning duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan enrichment data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

3. Transformation

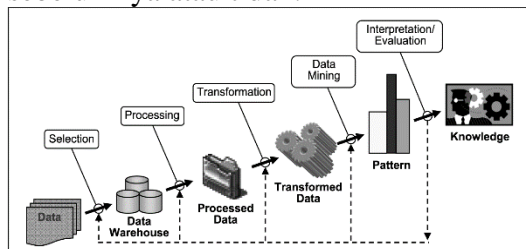
Merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam database.

4. Data Mining

Sebagai rangkaian kegiatan untuk menemukan pola yang menarik dari data dalam jumlah besar, kemudian data-data tersebut dapat disimpan dalam database, data warehouse atau penyimpanan informasi.

5. Interpretation / Evaluation

Data Mining ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD interpretation mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya atau tidak.



Gambar 2. Proses Knowledge Discovery in Database

Data Mining

Data Mining merupakan inti dari proses Knowledge Discovery in Database (KDD) yang melibatkan algoritma untuk mengeksplorasi data, mengembangkan model dan menemukan pola yang sebelumnya tidak diketahui juga dikenal sebagai pattern recognition yang digunakan untuk menemukan pola yang tersembunyi dari data yang diolah Metode Data Mining telah dikenal sejak tahun 1990 [3] sebagai salah satu alat untuk pengekstraksian suatu basis data yang berukuran besar guna memperoleh suatu pengetahuan baru.

Clustering

Clustering merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar daripada kesamaan data tersebut dengan data dalam kelompok lain Pada dasarnya clustering merupakan metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (similarity) antara satu data dengan data yang lain [5]. Data-data yang memiliki kemiripan karakteristik akan berkumpul dalam kelompok atau cluster yang sama. Data-data yang memiliki perbedaan karakteristik, akan berkumpul dalam kelompok atau cluster yang berbeda. Tujuan utama dari metode clustering adalah pengelompokan sejumlah data atau obyek ke dalam cluster (group) sehingga dalam setiap cluster akan berisi data yang semirip mungkin[6].

Algoritma K-Means

Algoritma K-Means merupakan algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah k cluster yang sudah ditetapkan

di awal [3] Algoritma K-Means memiliki ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk pengolahan objek dalam jumlah besar. Selain itu, algoritma K-Means ini tidak terpengaruh terhadap urutan objek. Salah satu tahapan penting dalam menerapkan K-Means *Cluster* adalah menentukan centroid, banyaknya *cluster* dan jarak centroid [4] pada penelitiannya memaparkan tahapan melakukan *clustering* atau pengelompokan dengan metode K-Means sebagai berikut:

1. Menentukan berapa banyak *cluster* yang ingin dibentuk, di mana nilai K adalah banyaknya *cluster* / jumlah *cluster*.
2. Menentukan pusat *cluster* (centroid) awal. Centroid awal ditentukan secara acak dari data yang ada dan jumlah centroid awal sama dengan jumlah *cluster*.
3. Setelah menentukan centroid awal, maka setiap data akan menemukan centroid terdekatnya yaitu dengan menghitung jarak setiap data ke masing-masing centroid menggunakan rumus korelasi antar dua obyek yaitu Euclidean Distance.

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

Di mana:

$d(x,y)$ = Jarak data ke x ke pusat *cluster*

y_{xi} = Data ke-i pada atribut data ke n

y_j = Data ke-j pada atribut data ke n

4. Setelah menghitung jarak data ke centroid-nya, maka langkah berikutnya adalah mengelompokkan data berdasarkan jarak minimumnya.

Suatu data akan menjadi anggota dari suatu *cluster* yang memiliki jarak terdekat (terkecil) dari pusat *cluster*-nya.

5. Berdasarkan pengelompokan tersebut, selanjutnya adalah mencari centroid baru berdasarkan membership dari masing-masing *cluster* yaitu dengan menghitung rata-rata dari data masing-masing *cluster*.
6. Kembali ke tahap 3.
7. Perulangan berhenti apabila tidak ada data lagi yang berpindah.

HASIL DAN PEMBAHASAN

Dimulai dengan mengumpulkan data di mana data tersebut diperoleh dari Auliya Tour & Travel Kisaran. Kemudian data tersebut diseleksi berdasarkan keperluan penelitian. Data hasil seleksi dilakukan proses transformasi data agar dapat diolah pada metode K-Means.

Data yang telah ditransformasi tersebut siap untuk diolah dengan metode K-Means yang dimulai dari menentukan jumlah *cluster*.

Pada penelitian ini menggunakan tiga *cluster*, yaitu Sangat Diminati, Diminati dan Kurang Diminati. Kemudian menentukan nilai titik pusat *cluster* atau disebut dengan centroid yang diambil secara acak dari dataset.

Setelah titik *cluster* ditentukan hitung jarak terdekat *cluster* lalu dikelompokkan berdasarkan jarak terdekat dengan *cluster*. Hitung kembali nilai centroid baru dengan cara menjumlahkan anggota kelompok dibagi total anggota.

Tabel 1. Data Sampel Hasil Transformasi

NO	NAMA	JK	USIA	PAKET
1	ABDUL HALIM MUHAMMAD NUH	1	8	1
2	ARMİYANTI ANSARI ITAM	0	6	1
3	ENNY HALAWATI	0	6	1
4	ERFINA LEGIMIN YOSO	0	6	1
5	HENNY SYARIF MUHAMMAD	0	7	2
6	LOKOT UDIN PANJAITAN	1	8	1
7	MAIMUNAH ABDUL MUNIR SAINAL	0	8	3
8	MARTINI DARMAJI SENGUT	0	8	1
9	MASNUN MUSA HASIBUAN	0	7	2
10	MISRAN RASUL HASAN	1	6	2
11	MUHAMMAD JUNAIDI SADINO	1	8	1
12	NANING WINARSIH SUTARNO	0	5	3
13	NEZA APRIDILLA	0	2	2
14	PONI WIRYO SUMARTO	0	8	3
15	RITAWATI NASIB GURUSINGA	0	7	2
16	ROSDIANA MUCHTARAM NASUTION	0	8	2
17	ROSLAINI ISMAIL FIRMAN	0	8	1
18	SUHARTINI LASIMIN KROMO	0	8	1
19	SUMARNI SUMARTO ABDULLAH	0	8	2
20	SUNARMI OSERI ABDULLAH	0	7	3
21	SUSILAWATI HUSIN TUMAN	0	6	2
22	TAING MUHAMMAD SARI SITOMPUL	0	7	1
23	TODUH KETEK ABDULLAH	0	7	2
24	ZAHIR NAUFAL AR RAZAQ	1	2	1
25	ZULIJAWATI NGATIMAN MIKAR	0	6	2
26	ZULISMAR SRI HARTATI	0	6	1

Tabel 2. Centroid Awal

NO.	NAMA	JK	USIA	PAKET
13	NEZA APRIDILLA	0	2	2
16	ROSDIANA MUCHTARA M NASUTION	0	8	2
25	ZULIJAWATI NGATIMAN MIKAR	0	6	2

Data sampel baris ke-1 atas nama Abdul Halim Muhammad Nuh dihitung dengan titik centroid ke-1 sebagai DC1, centroid ke-2 sebagai DC2 dan centroid ke-3 sebagai DC3,

di mana:

$$DC1 = \sqrt{(1 - 0)^2 + (8 - 2)^2 + (1 - 2)^2} = 6.164414003$$

$$DC2 = \sqrt{(1 - 0)^2 + (8 - 8)^2 + (1 - 2)^2} = 1.414213562$$

$$DC3 = \sqrt{(1 - 0)^2 + (8 - 6)^2 + (1 - 2)^2} = 2.449489743$$

Jadi, nilai terkecil baris ke-1 terdapat pada DC2 maka data jamaah baris ke-1 ditempatkan pada kelompok C2.

Perhitungan dilanjutkan sampai baris data ke-26 dan mendapatkan hasil penempatan anggota *cluster* berdasarkan kelompok

Lakukan iterasi untuk menghitung jarak minimum dengan nilai centroid baru dan mengelompokkan anggota *cluster*, apabila anggota *cluster* sudah tidak mengalami perubahan tempat maka iterasi dihentikan dan anggota kelompok telah ditemukan.

Tabel 3. Hasil Perhitungan Jarak dan Pengelompokan Pada Iterasi Ke-1

NO	NAMA	DC1	DC2	DC3	C1	C2	C3
1	ABDUL HALIM MUHAMMAD NUH	6.164414003	1.414213562	2.449489743		1	
2	ARMIYANTI ANSARI ITAM	4.123105626	2.236067977	1			1
3	ENNY HALAWATI	4.123105626	2.236067977	1			1
4	ERFINA LEGIMIN YOSO	4.123105626	2.236067977	1			1
5	HENNY SYARIF MUHAMMAD	5	1	1		1	
6	LOKOT UDIN PANJAITAN	6.164414003	1.414213562	2.449489743		1	
7	MAIMUNAH ABDUL MUNIR SAINAL	6.08276253	1	2.236067977		1	
8	MARTINI DARMAJI SENGUT	6.08276253	1	2.236067977		1	
9	MASNUN MUSA HASIBUAN	5	1	1		1	
10	MISRAN RASUL HASAN	4.123105626	2.236067977	1			1
11	MUHAMMAD JUNAIDI SADINO	6.164414003	1.414213562	2.449489743		1	
12	NANING WINARSIH SUTARNO	3.16227766	3.16227766	1.414213562			1
13	NEZA APRIDILLA	0	6	4	1		
14	PONI WIRYO SUMARTO	6.08276253	1	2.236067977		1	
15	RITAWATI NASIB GURUSINGA	5	1	1		1	
16	ROSDIANA MUCHTARAM NASUTION	6	0	2		1	
17	ROSLAINI ISMAIL FIRMAN	6.08276253	1	2.236067977		1	
18	SUHARTINI LASIMIN KROMO	6.08276253	1	2.236067977		1	
19	SUMARNI SUMARTO ABDULLAH	6	0	2		1	
20	SUNARMI OSERI ABDULLAH	5.099019514	1.414213562	1.414213562		1	
21	SUSILAWATI HUSIN TUMAN	4	2	0			1
22	TAING MUHAMMAD SARI SITOMPUL	5.099019514	1.414213562	1.414213562		1	
23	TODUH KETEK ABDULLAH	5	1	1		1	
24	ZAHIR NAUFAL AR RAZAQ	1.414213562	6.164414003	4.242640687	1		
25	ZULIJAWATI NGATIMAN MIKAR	4	2	0			1
26	ZULISMAR SRI HARTATI	4.123105626	2.236067977	1			1

Data yang dihasilkan dalam bentuk *clustering* atau kelompok data jamaah umroh.

$$C1 = \frac{(1+0)}{2} = 0.5$$

$$\frac{(2+2)}{2} = 2$$

$$\frac{(1+2)}{2} = 1.5$$

C1 = 0.5, 2, 1.5

C2 =

$$\frac{(1+0+1+0+0+0+1+0+0+0+0+0+0+0+0+0)}{16} = 0.1875$$

$$\frac{(8+7+8+8+8+7+8+8+7+8+8+8+8+7+7+7)}{16} = 7.625$$

$$\frac{(1+2+1+3+1+2+1+3+2+2+1+1+2+3+1+2)}{16}$$

Iterasi selanjutnya dilakukan perhitungan kembali dengan menggunakan titik centroid yang baru. Proses perhitungan tetap dengan menggunakan rumus *Euclidean Distance*.

Setelah didapatkan hasil iterasi pertama dan kedua kemudian dibandingkan dan diperoleh posisi anggota *cluster* pada iterasi pertama dan kedua tidak mengalami perubahan sehingga proses iterasi dihentikan dan hasil akhir yang diperoleh yaitu:

1. Kelompok Sangat diminati terdiri dari 16 anggota dengan kategori umur 7 sampai 8 yaitu rentang usia dari 49 tahun sampai 65 tahun.
2. Kelompok Diminati terdiri dari 8 anggota dengan kategori umur 5 sampai 6 yaitu rentang usia dari 34 tahun sampai 45 tahun.
3. Kelompok Kurang Diminati terdiri dari 2 anggota dengan kategori

= 1.75

C2 = 0.1875, 7.625, 1.75

$$C3 = \frac{(0+0+0+1+0+0+0+0)}{8} = 0.125$$

$$\frac{(6+6+6+6+5+6+6+6)}{8} = 5.875$$

$$\frac{(1+1+1+2+3+2+2+1)}{8} = 1.625$$

C3 = 0.125, 5.875, 1.625

Dari hasil rata-rata di atas didapatkan nilai pusat *cluster* baru yang ditampilkan dalam bentuk Tabel 4 sebagai berikut:

Tabel 4 Nilai *Centroid* Baru Untuk Iterasi Ke-2

CLUSTER	NILAI CENTROID BARU		
C1	0.5	2	1.5
C2	0.1875	7.625	1.75
C3	0.125	5.875	1.625

umur 2 yaitu rentang usia 8 tahun sampai 11 tahun.

Hasil di atas merupakan hasil dari 26 sampel data yang diuji. Hasil pengelompokan usia jamaah umroh dapat membantu Auliya Tour & Travel dalam melihat kelompok jamaah berdasarkan usia.

SIMPULAN

Berdasarkan penelitian yang dilakukan dalam mengelompokan data jamaah umroh pada Auliya Tour & Travel Kisaran dengan *K-Means Clustering*, maka dapat diambil kesimpulan sebagai berikut:

1. Variabel yang mempengaruhi dalam pengelompokan data terdiri jenis kelamin, usia dan paket yang

- telah mengalami proses transformasi data.
2. Data yang diperoleh dari Auliya Tour & Travel masih berbentuk teks sehingga diperlukan penyeleksian dan transformasi data ke dalam bentuk numberik agar dapat diproses pada algoritma K-Means.
 3. Perancangan sistem metode K-Means dimulai dari mengumpulkan data, seleksi dan transformasi data setelah itu sistem akan menghitung dan menampilkan.
 4. Pengujian data sebanyak 170 record diolah dengan *Software RapidMiner* untuk mendapatkan hasil pengelompokan dalam waktu singkat.

UCAPAN TERIMA KASIH

Terima kasih kepada Pimpinan dan Staf Auliya Tour & Travel Kisaran yang telah banyak memberikan dukungan dalam penelitian ini.

DAFTAR PUSTAKA

- [1] Kurniawansyah, A. S. (2018) 'Implementasi Metode Artificial Neural Network dalam Memprediksi Hasil Ujian Kompetensi Kebidanan (Studi Kasus: Akademi Kebidanan Dehasen Bengkulu)', *Jurnal Pseudocode*, 5(1), Februari 2018, pp. 37-44.
- [2] Musharyadi, F. (2017) 'Tingkat Pemahaman Mahasiswa Terhadap Norma Agama Islam Menggunakan Algoritma K-Means Clustering', *Menara Ilmu* 11(78), November 2017, pp. 48-54.
- [3] Nuryaman, Y., Asistiyasari, A. dan Yudha, A. (2018) 'Komparasi Algoritma K-Mean dan AHC untuk klasifikasi curah hujan di Indonesia', *Ikraith-Informatika*, 2(2), Juli 2018, pp. 70-75.
- [4] Aranda, J. dan Natasya, W. A. G (2016) 'Penerapan Metode K-Means Cluster Analysis Pada Sistem Pendukung Keputusan Pemilihan Konsentrasi Untuk Mahasiswa International Class STMIK AMIKOM Yogyakarta', *Seminar Nasional Teknologi Informasi dan Multimedia 2016*, (January 2018), pp. 6-7.
- [5] Mardalius, Mardalius. "Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means." *JURTEKSI* 4.2 (2018): 123-132.
- [6] Sibuea, Mustika Larasati, and Andy Safta. "Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering ." *JURTEKSI* 4.1 (2017): 85-92