

PENGELOMPOKAN SURAT DALAM AL QUR'AN MENGGUNAKAN ALGORITMA K-MEANS

Tutik Khotimah

Dosen Fakultas Teknik, Program Studi Teknik Informatika

Universitas Muria Kudus

Email: tutik.khotimah@gmail.com

ABSTRAK

Kitab suci Al Qur'an terdiri dari 114 surat. Masing-masing surat memiliki jumlah ayat yang berbeda. Ayat-ayat tersebut ada yang panjang dan ada yang pendek. Banyak atau sedikit jumlah ayat dalam suatu surat, dan panjang atau pendek ayat memberi pengaruh terhadap lama waktu membaca surat-surat dalam Al Qur'an. Penelitian ini bertujuan untuk melakukan pengelompokan (*clustering*) surat-surat dalam Al Qur'an berdasarkan jumlah ayat, jumlah ruku' dan lama waktu membaca surat. Algoritma *clustering* yang digunakan adalah *K-means*. Dari percobaan yang dilakukan, nilai optimal diperoleh pada *iterasi* ke-14 dengan nilai *k* sebanyak 4 *cluster*. *Cluster_0* terdiri dari 18 anggota dengan nilai *centroid* pada variabel ayat sebesar 106,389, variabel ruku' sebesar 9,778, dan variabel lama sebesar 1333,5. *Cluster_1* terdiri dari 26 anggota dengan nilai *centroid* pada variabel ayat sebesar 62,038, variabel ruku' sebesar 4,885, dan variabel lama sebesar 662,5. *Cluster_2* memiliki 7 anggota dengan nilai *centroid* pada variabel ayat sebesar 183,143, variabel ruku' sebesar 22,857 dan variabel lama sebesar 3310,571. Sedangkan *cluster_3* memiliki 63 anggota dengan nilai *centroid* pada variabel ayat sebesar 22,635, variabel ruku' sebesar 1,508, dan variabel lama sebesar 145,635.

Kata kunci: *Clustering, K-means, Al Qur'an*

ABSTRACT

Holy Qur'an consists of 114 surah . Each letter has a different number of ayat . The ayat have a long and some short . A lot or a little amount of ayat in a surah , as well as long and short ayat giving effect to a long time to read the surah in the Qur'an . This study aims to clustering surah in the Qur'an based on the ayat , the ruku' and a long time to read the letter . Clustering algorithm used is the K - means clustering . From the experiments conducted , the optimal value is obtained at iteration 14th with k= 4 . Cluster_0 consists of 18 members with the centroid of the variable ayat=106.389 , the variable ruku'= 9.778 , and the variable duration=1333.5 . Cluster_1 consists of 26 members with the centroid of the variable ayat= 62.038 , the variable ruku'=4.885 , and the variable duration=662.5 . Cluster_2 has 7 members with the centroid of the variable ayat=183.143 , variable ruku'=22.857 and the variable duration=3310.571 . While cluster_3 has 63 members with the centroid of the variable ayat=22.635 , the variable ruku'=1,508 , and the variable duration=145.635 .

Keywords: *Clustering, K-means, Al Qur'an*

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Al Qur'an adalah kitab suci yang digunakan sebagai pedoman hidup manusia, khususnya umat Islam. Maka, membaca dan mempelajari Al Qur'an adalah sesuatu yang wajib dan bernilai sebagai amal ibadah. Sejak zaman sahabat, Al Qur'an telah dibagi-bagi menjadi beberapa bagian, di antaranya adalah dua bagian, tiga bagian, lima bagian, tujuh bagian dan sebagainya. Pembagian tersebut dilakukan untuk mempermudah dalam penghafalan atau pembacaan Al Qur'an pada amalan-amalan harian, atau pembacaan Al Qur'an dalam sholat.

Al Qur'an terdiri dari 114 surat, masing-masing surat terdiri dari beberapa ayat. Jumlah ayat dalam Al Qur'an mencapai 6236 ayat. Al Qur'an dibagi lagi menjadi 30 bagian yang sama banyak yang disebut dengan juz. Setiap juz dibagi menjadi beberapa ruku'. Surat-surat dalam Al Qur'an memiliki jumlah ruku' yang berbeda-beda tergantung banyaknya ayat dalam surat dan panjang pendeknya masing-masing ayat.

Berdasarkan panjang pendeknya ayat, surat-surat dalam Al Qur'an dikelompokkan menjadi empat macam [1] yaitu: *ath-thiwal*, *al-mi'uun*, *al-matsaani*, dan *al-mufashshol*. *Ath-thiwal* adalah surat-surat yang memiliki ayat yang panjang. *Al-mi'uun* adalah surat-surat yang memiliki ayat lebih dari atau mendekati 100. *Al-matsaani* adalah surat-

surat pendek yang sering dibaca. Sedangkan Al-mufashshol adalah surat-surat selain ath-thiwal, al-mi'uun, dan al-matsaani. Pengelompokan tersebut bertujuan untuk memudahkan manusia dalam menghafal, mempelajari, dan mengkaji Al Qur'an. Namun, para ulama' masih berbeda pendapat mengenai pengelompokan tersebut.

Pengelompokan surat-surat dalam Al Qur'an adalah hal yang penting. Untuk itu, diperlukan suatu teknik yang dapat digunakan untuk mengelompokkan surat-surat Al Qur'an agar dapat mempermudah manusia dalam menghafal, mempelajari, dan mengkaji Al Qur'an. Pada penelitian ini, pengelompokan tersebut didasarkan pada jumlah ayat dan jumlah ruku' dari masing-masing surat, mengingat bahwa banyak sedikitnya jumlah ayat dalam suatu surat dan panjang pendeknya masing-masing ayat memberi pengaruh terhadap lama waktu yang dibutuhkan untuk membaca surat tersebut. Lama waktu yang dibutuhkan itu juga digunakan sebagai variabel pengelompokan. Namun, lama waktu yang dibutuhkan masing-masing orang adalah berbeda-beda, tergantung tingkat kefasihan dalam membaca Al Qur'an. Semakin fasih seseorang, ia akan semakin lancar dan cepat dalam membaca Al Qur'an. Sehingga pada penelitian ini lama waktu yang digunakan adalah lama waktu yang diperlukan oleh imam Masjidil Haram, Syaikh Abdurrahman As-Sudais.

1.2 Tujuan

Tujuan dari penelitian ini adalah untuk melakukan pengelompokan surat-surat dalam Al Qur'an berdasarkan jumlah ayat, jumlah ruku' dan lama waktu membaca surat dengan menggunakan algoritma *K-means*.

1.3 Manfaat

Hasil penelitian diharapkan dapat memberikan kontribusi terhadap pembagian Al Qur'an agar dapat membantu memudahkan manusia dalam menghafal, mempelajari, dan mengkaji Al Qur'an.

2. TINJAUAN PUSTAKA

2.1 Clustering

Pengelompokan (*clustering*) merupakan bagian dari ilmu *data mining* yang bersifat tanpa arahan (*unsupervised*) [2]. *Clustering* adalah proses pembagian data ke dalam kelas atau *cluster* berdasarkan tingkat kesamaannya [3]. Dalam *clustering*, data yang memiliki kesamaan dimasukkan ke dalam *cluster* yang sama, sedangkan data yang tidak memiliki kesamaan dimasukkan dalam *cluster* yang berbeda [4].

2.2 K-means

K-means adalah salah satu algoritma *clustering* yang menggunakan metode partisi [3]. *K-means* adalah algoritma *clustering* yang membagi masing-masing item data ke dalam satu *cluster* [4][5]. Berikut ini adalah langkah-langkah dalam algoritma *K-means* [5]:

1. Tentukan jumlah *cluster* (*k*) pada *data set*
2. 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:

$$\bar{V}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} \quad (1)$$

Keterangan:

- \bar{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* dalam *cluster* tersebut

3. Pada masing-masing record, hitung jarak terdekat dengan *centroid*.

Jarak *centroid* yang digunakan adalah *Euclidean Distance*, dengan rumus seperti pada persamaan (2):

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

Keterangan:

D_e = *Euclidean Distance*

i = banyaknya objek

(x, y) = koordinat objek

(s, t) = koordinat *centroid*

4. Kelompokkan objek berdasarkan jarak ke *centroid* terdekat
5. Ulangi langkah ke-2, lakukan *iterasi* hingga *centroid* bernilai optimal

3. METODE PENELITIAN

Penelitian ini adalah penelitian eksperimen dengan langkah-langkah penelitian sebagai berikut:

1. Pengumpulan data

Data yang digunakan adalah data jumlah ayat dan jumlah ruku' dalam masing-masing surat dalam Al Qur'an. Data ini diambil dari mushof Al Qur'an. Data lain yang dibutuhkan adalah data durasi pembacaan surat Al Qur'an oleh imam Masjidil Haram, Syaikh Abdurrahman bin Abdul Aziz As-Sudais An-Najdi berupa *file* murotal¹.

2. *Pre-processing*

Pada tahap *pre-processing* dilakukan *data integration* dan *data transformation*. *Data integration* yaitu penggabungan data dari data mushof Al Qur'an dan data dari *file* murotal sehingga menjadi 1 tabel. *Data transformation* dilakukan dengan mengubah *value* dari durasi pembacaan yang bertipe *time* menjadi tipe *integer*.

3. *Clustering*

Pada tahap ini, algoritma yang digunakan adalah algoritma *K-means*, dengan nilai k sebanyak 4 *cluster*. Percobaan dilakukan berulang-ulang dengan menggunakan *software* RapidMiner hingga diperoleh nilai optimal, yaitu nilai dimana nilai *centroid* tidak mengalami perubahan. Nilai optimal ini diperoleh pada *iterasi* ke-14 ke atas.

4. HASIL DAN PEMBAHASAN

4.1 Hasil *Clustering*

Cluster Model yang dihasilkan dari pengelompokan surat dalam Al Qur'an dengan algoritma *K-means* ditampilkan dalam Gambar 1. Pada Gambar 1 tampak bahwa keanggotaan *Cluster 0* terdiri dari 18 *item*, keanggotaan *Cluster 1* terdiri dari 26 *item*, keanggotaan *Cluster 2* terdiri dari 7 *item*, dan keanggotaan *Cluster 3* terdiri dari 63 *item*. Jumlah keseluruhan *item* sebanyak 114.

Gambar 2 menampilkan nilai *centroid* pada masing-masing *cluster* dengan atribut yang digunakan adalah Ayat, Ruku', dan Lama. Pada *cluster_0*, *centroid* atribut Ayat berada pada *value* 106,389, atribut Ruku' berada pada *value* 9,778 dan atribut Lama berada pada *value* 1333,500. Pada *cluster_1*, *centroid* atribut Ayat berada pada *value* 62,038, atribut Ruku' berada pada *value* 4,885 dan atribut Lama berada pada *value* 662,500. Pada *cluster_2*, *centroid* atribut Ayat berada pada *value* 183,143, atribut Ruku' berada pada *value* 22,857 dan atribut Lama berada pada *value* 3310,571. Sedangkan pada *cluster_3*, *centroid* atribut Ayat berada pada *value* 22,635, atribut Ruku' berada pada *value* 1,508 dan atribut Lama berada pada *value* 145,635.

¹ <http://quranicaudio.com>

Cluster Model	
Cluster 0:	18 items
Cluster 1:	26 items
Cluster 2:	7 items
Cluster 3:	63 items
Total number of items:	114

Gambar 1. Cluster model pengelompokan surat Al Qur'an

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
Ayat	106.389	62.038	183.143	22.635
Ruku'	9.778	4.885	22.857	1.508
Lama	1333.500	662.500	3310.571	145.635

Gambar 2. Nilai centroid

Ada pun daftar keanggotaan *cluster* dapat dilihat pada tabel 1 berikut ini:

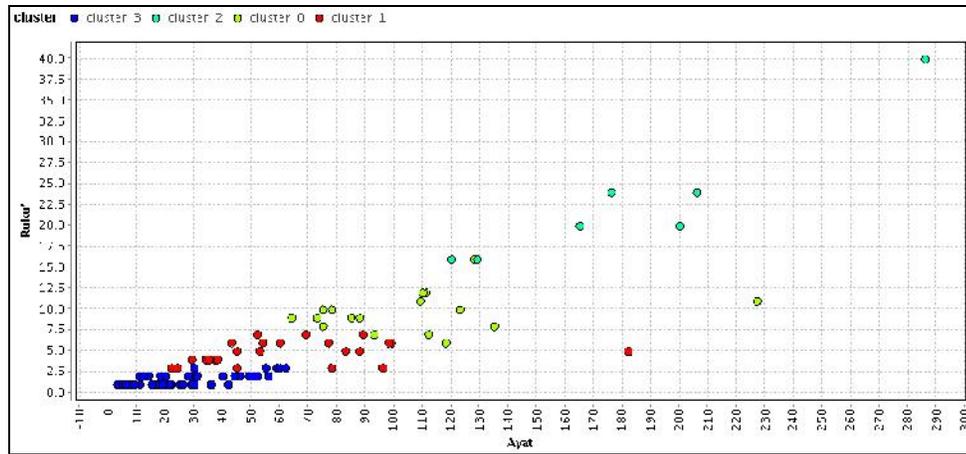
Tabel 1. Daftar anggota *cluster*

No	Cluster	Anggota	Jml
1	Cluster_0	Al Anfaal, Yunus, Huud, Yusuf, An Nahl, Al Israa', Al Kahfi, Thaahaa, Al Anbiyaa', Al Hajj, Al Mu'minuun, An Nuur, Asy Syu'araa', An Naml, Al Qashash, Al Ahzab, Az Zumar, Al Mu'min	18
2	Cluster_1	Ar Ra'd, Ibrahim, Al Hijr, Maryam, Al Furqaan, Al 'Ankabuut, Ar Ruum, Luqman, Saba', Faathir, Yaasiin, Ash Shaaffat, Shaad, Fushshilat, Asy Syuura, Az Zukhruf, Al Jaatsiyah, Al Ahqaaf, Muhammad, Al Fath, Qaaf, Ar Rahmaan, Al Waaqi'ah, Al Hadiid, Al Mujaadilah, Al Hasyr	26
3	Cluster_2	Al Baqarah, Ali 'Imran, An Nisaa', Al Maa-idah, Al An'am, Al A'raaf, At Taubah	7
4	Cluster_3	Al Faatihah, As Sajdah, Ad Dukhaan, Al Hujuraat, Adz Dzariyaat, Ath Thuur, An Najm, Al Qamar, Al Mumtahanah, Ash Shaff, Al Jumu'ah, Al Munaafiquun, At Taghaabun, Ath Thalaaq, At Tahriim, Al Mulk, Al Qalam, Al Haaqqah, Al Ma'aarij, Nuh, Al Jin, Al Muzzammil, Al Muddatstsir, Al Qiyaamah, Al Insaan, Al Mursalaat, An Naba', An Nazi'at, 'Abasa, At Takwiir, Al Infithaar, Al Muthaffiifiin, Al Insiyiqaaq, Al Buruuj, Ath Thaarig, Al A'laa, Al Ghaasyiyah, Al Fajr, Al Balad, Asy Syams, Al Lail, Adh Dhuhaa, Alam Nasyrah, At Tiin, Al 'Alaq, Al Qadr, Al Bayyinah, Al Zalzalah, Al 'Aadiyaat, Al Qaari'ah, At Takaatsur, Al 'Ashr, Al Humazah, Al Fiil, Quraisy, Al Maa'uun, Al Kautsar, Al Kaafiruun, An Nashr, Al Lahab, Al Ikhlash, Al Falaq, An Naas	63
Total			144

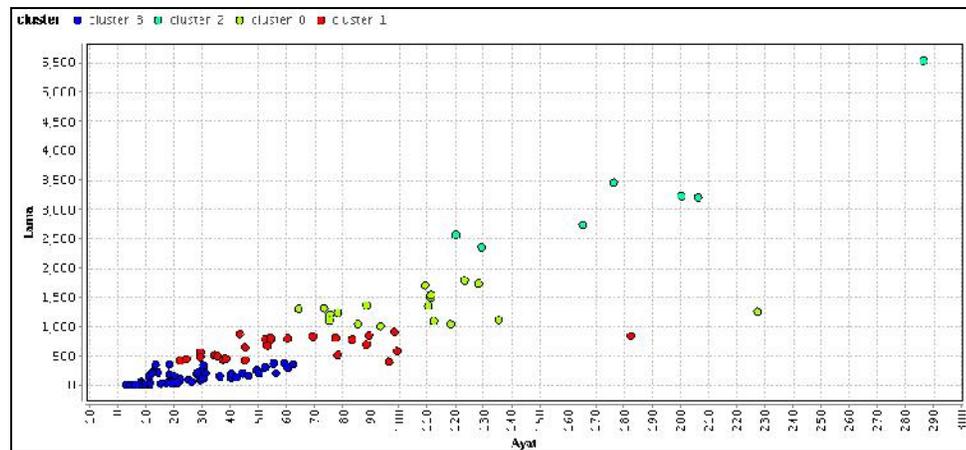
4.2 Analisis Cluster

Keanggotaan *cluster* dikelompokkan berdasarkan kemiripannya. Dari pengamatan hasil *clustering* diperoleh bahwa surat-surat yang menempati *cluster_2* memiliki jumlah ayat yang banyak (120-190an), demikian juga jumlah ruku'nya (15-40an), dan waktu yang dibutuhkan untuk membaca adalah lama yaitu berkisar antara 2.000 - 5.500an detik. *Cluster_2* merupakan kelompok surat yang panjang, yang terdiri dari 7 surat yaitu: Al Baqarah, Ali 'Imran, An Nisaa', Al Maa-idah, Al An'am, Al A'raaf, At Taubah. *Cluster_3* terdiri dari surat-surat pendek yang memiliki jumlah ayat antara 3-62 dan jumlah ruku' 1-3. Waktu yang dibutuhkan untuk membaca surat-surat pada *cluster_3* adalah sebentar, yaitu kurang dari 400 detik. *Cluster_0* merupakan kumpulan dari surat-surat yang membutuhkan waktu 1.000-2.000 detik untuk membacanya. Sedangkan *cluster_1* terdiri dari surat-surat yang bacaannya membutuhkan waktu

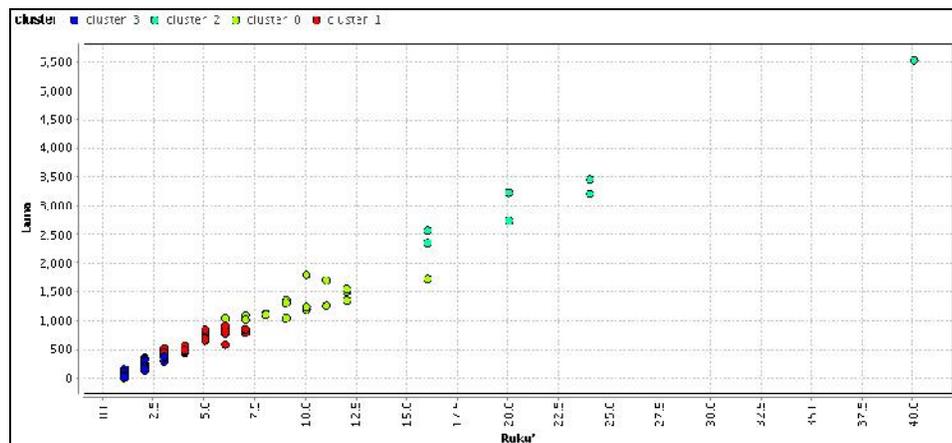
sekitar 400-900an detik. Komposisi keanggotaan *cluster* berdasarkan jumlah ayat dan jumlah ruku' ditampilkan pada Gambar 3. Komposisi keanggotaan *cluster* berdasarkan jumlah ayat dan lama membaca ditampilkan pada Gambar 4. Sedangkan komposisi keanggotaan *cluster* berdasarkan jumlah ruku' dan lama membaca ditampilkan pada Gambar 5.



Gambar 3. Analisis *cluster* berdasarkan jumlah ayat dan jumlah ruku'



Gambar 4. Analisis *cluster* berdasarkan jumlah ayat dan lama membaca



Gambar 5. Analisis *cluster* berdasarkan jumlah ruku' dan lama membaca

5. KESIMPULAN

Dari hasil penelitian ini dapat disimpulkan sebagai berikut:

1. Telah dilakukan pengelompokan surat-surat dalam Al Qur'an dengan menggunakan algoritma *K-means*.
2. Pengelompokan menghasilkan 4 *cluster* dengan anggota *cluster_0* sebanyak 18 surat, *cluster_1* sebanyak 26 surat, *cluster_2* sebanyak 7 surat, *cluster_3* sebanyak 63 surat. Pengelompokan ini telah mencapai nilai optimal pada *iterasi* ke-14.

DAFTAR PUSTAKA

- [1] Al-Qaththan, Syaikh Manna. (2007). *Pengantar Studi Ilmu Al Qur'an*, Pustaka Al-Kautsar.
- [2] Agusta, Yudi. 2007. *K-means – Penerapan, Permasalahan dan Metode Terkait*. *Jurnal Sistem dan Informatika*, Vol 3 bulan Pebruari 2007. hal 47-60.
- [3] Han, Jiawei, Micheline Kamber. (2006). *Data Mining: Concepts and Techniques, Second Edition*. Elsevier.
- [4] Bramer, Max. (2007). *Principles of Data Mining*. Springer.
- [5] Larose, Daniel T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Wiley.