

Clustering Dokumen Fatwa MUI berdasarkan Bidang Permasalahan Menggunakan *K-Means*

Fajar Rohman Hariri¹, Ala Syauqi², Ainatul Mardiyah³

^{1,2,3}123Teknik Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim Malang
E-mail: ¹fajar@ti.uin-malang.ac.id, ²syauqi@ti.uin-malang.ac.id, ³ainadimana@gmail.com

Abstrak – Sejak Majelis Ulama Indonesia (MUI) berdiri pada tahun 1975 sampai sekarang, lembaga ini telah menghasilkan fatwa sebanyak 201 buah yang mencakup berbagai bidang. Seperti masalah ibadah, ahwal al-syakhshiyah, keluarga berencana, masalah makanan dan minuman, kebudayaan, hubungan antar agama, dan lain-lain. Text mining merupakan salah satu teknik yang digunakan untuk menggali data yang tersembunyi dari data yang berbentuk text. Salah satu metode dalam text mining adalah clustering. Penelitian kali ini mengimplementasikan metode *K-Means Clustering* pada dokumen – dokumen fatwa MUI untuk mengelompokkan fatwa yang ada berdasarkan kemiripan masalah yang dibahas. *Silhouette Coefficient* digunakan untuk analisa terhadap cluster yang dihasilkan, dengan nilai terbaik 0.0994699359772012 dengan pengelompokan 10 cluster.

Kata Kunci — Clustering, fatwa, *K-means*

Abstract – Since the Indonesian Ulema Council (MUI) was established in 1975 until now, this institution has produced 201 edicts covering various fields. Such as matters of worship, ahwal al-syakhshiyah, family planning, food and beverage issues, culture, interfaith relations, and others. Text mining is one of the techniques used to dig up hidden data from data in the form of text. One method in text mining is clustering. The present study implements the *K-Means Clustering* method in MUI fatwa documents to classify existing fatwas based on the similarity of the issues discussed. *Silhouette Coefficient* is used for the analysis of the resulting clusters, with the best value of 0.0994699359772012 with 10 clusters grouping

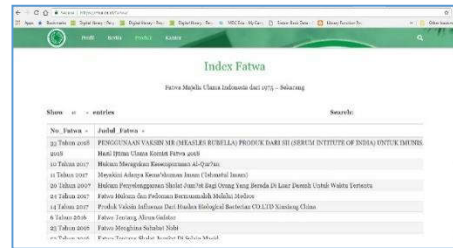
Keywords — Clustering, fatwa, *K-means*.

1. PENDAHULUAN

Fatwa merupakan produk hukum Islam yang sudah ada semenjak masa Nabi SAW, yang kemudian menjadi produk hukum Islam yang berkembang hingga sekarang. Fatwa-fatwa ulama Islam yang terhimpun dalam kitab-kitab fiqh dan keputusan-keputusan lembaga fatwa adalah sebagian dari ijtihad yang bersifat kasuistik karena merupakan respon atau jawaban terhadap pertanyaan yang diajukan oleh peminta fatwa. Jadi, fatwa merupakan salah satu solusi dalam memecahkan persoalan yang terjadi di zaman modern ini.

Majelis Ulama Indonesia (MUI), yang merupakan wadah musyawarah para ulama, zu'ama, dan cendekiawan muslim serta menjadi pengayom bagi seluruh muslim Indonesia adalah lembaga paling berkompeten dalam menjawab dan memecahkan setiap masalah sosial keagamaan yang senantiasa timbul dan dihadapi masyarakat.

Sejak Majelis Ulama Indonesia (MUI) berdiri pada tahun 1975 sampai sekarang, lembaga ini telah menghasilkan fatwa sebanyak 201 buah yang mencakup berbagai bidang. Seperti masalah ibadah, ahwal al-syakhshiyah, keluarga berencana, masalah makanan dan minuman, kebudayaan, hubungan antar agama, dan lain-lain. Informasi mengenai fatwa MUI dapat diakses pada laman <https://mui.or.id/fatwa/>. Namun seperti terlihat pada gambar 1, belum ada pengelompokan fatwa pada laman tersebut.

Gambar 1. Web <https://mui.or.id/fatwa>

Text mining merupakan salah satu teknik yang digunakan untuk menggali data yang tersembunyi dari data yang berbentuk text. Salah satu metode dalam text mining adalah clustering. Teknik Clustering sendiri merupakan teknik pengelompokan yang banyak dipakai dalam data mining. Menurut Santosa (2007) tujuan utama dari metode cluster adalah pengelompokan sejumlah data / objek ke dalam cluster (group) sehingga dalam setiap cluster akan berisi data yang semirip mungkin[1].

Text mining telah banyak digunakan untuk pengelompokan dokumen. Cepi Slamet (2016) mengelompokkan surat-surat dalam Al-Qur'an berdasarkan ayat dengan menggunakan algoritma clustering k-means[2]. Deka (2014) menggunakan teknik clustering k-means untuk mengelompokkan judul buku sesuai dengan kategorinya sehingga memudahkan pustakawan dalam pengelompokan peletakan buku dan merancang strategi dalam meningkatkan minat baca mahasiswa pada Universitas Islam Indonesia[3]. Eko Yulian (2018) mengimplementasikan Text Mining dengan K-Means Clustering pada Tema LGBT dalam Arsip Tweet Masyarakat Kota Bandung.[4]

Penelitian kali ini mengimplementasikan metode K-Means Clustering pada dokumen – dokumen fatwa MUI untuk mengelompokkan fatwa yang ada berdasarkan kemiripan masalah yang dibahas. Dan untuk stemming dokumennya menggunakan algoritma nazief & andriani karena algoritma ini memiliki akurasi yang lebih baik dibandingkan algoritma porter.[5]

2. METODE PENELITIAN

2.1 Jenis Penelitian

Penelitian ini merupakan penelitian eksperimental, karena untuk mendapatkan performa terbaik algoritma K-Means dalam mengcluster, dilakukan beberapa kali percobaan dengan parameter yang berbeda-beda. Data didapatkan dari dokumen fatwa MUI yang didapatkan dari laman <https://mui.or.id/fatwa/> dengan ekstensi pdf.

2.2 Metode Analisis Data

Untuk analisis data, dari dokumen fatwa yang diperoleh, akan dilakukan percobaan dengan beberapa scenario ujicoba yang dijelaskan pada tabel 1 dibawah

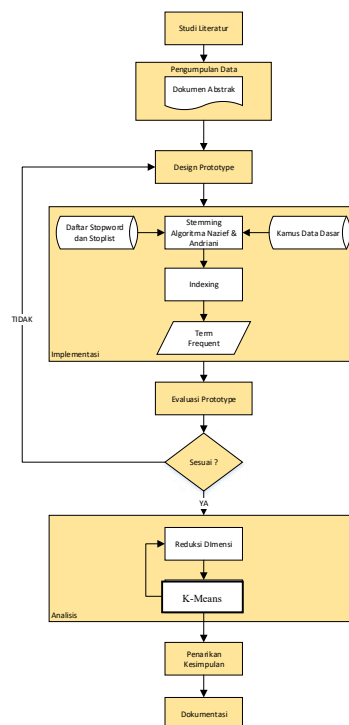
Table 1. Skenario Uji Coba

Reduksi Dimensi (%)	Jumlah Cluster
0 (Semua)	2 , 3 , 4 , 5 , 6 , 7 , 8 , 9 , 10
10	2 , 3 , 4 , 5 , 6 , 7 , 8 , 9 , 10
20	2 , 3 , 4 , 5 , 6 , 7 , 8 , 9 , 10
30	2 , 3 , 4 , 5 , 6 , 7 , 8 , 9 , 10
40	2 , 3 , 4 , 5 , 6 , 7 , 8 , 9 , 10
50	2 , 3 , 4 , 5 , 6 , 7 , 8 , 9 , 10
60	2 , 3 , 4 , 5 , 6 , 7 , 8 , 9 , 10
70	2 , 3 , 4 , 5 , 6 , 7 , 8 , 9 , 10
80	2 , 3 , 4 , 5 , 6 , 7 , 8 , 9 , 10
90	2 , 3 , 4 , 5 , 6 , 7 , 8 , 9 , 10

Untuk kemudian akan dibandingkan cluster yang dihasilkan menggunakan Silhouette Coefficient

2.3 Metode Analisis Data

Secara umum, penelitian dilakukan dengan beberapa tahapan seperti pada gambar 2:



Gambar 2. Alur Penelitian

Alur penelitian dari penelitian ini adalah:

2.3.1 Studi Literatur

Pada tahap ini dilakukan pengumpulan informasi yang diperlukan untuk penelitian. Informasi yang diperlukan diperoleh dengan mempelajari literatur- literatur mengenai permasalahan clustering dokumen, metode K Means dan algoritma stemming untuk bahasa Indonesia.

2.3.2 Pengumpulan Data

Mencari data-data yang terkait dengan sistem, antara lain mengumpulkan data dokumen fatwa MUI

2.3.3 Design Prototype

Membuat design prototype untuk proses stemming, pembentukan term frequent.

2.3.4 Implementasi

2.3.4.1 Stemming Algoritma Nazief & Andriani.

Proses pemecahan kata dan mengambil kata dasar dari dokumen abstrak menggunakan Algoritma Nazief & Andriani.

2.3.4.2 Indexing

Dari data kata dasar masing-masing dokumen akan dilist dan dilihat frekuansi masing-masing kata untuk selanjutnya dibentuk sebuah matrix term frequent.

2.3.4.3 Evaluasi Prototype

Dilakukan analisa terhadap prototype yang ada, jika sudah sesuai lanjut ke proses selanjutnya, jika masih belum, dilakukan perbaikan pada prototype sampai sesuai.

2.3.4.4 Analisis

a.Reduksi Dimensi

Dokumen – dokumen abstrak yang ada akan menghasilkan banyak kata dasar, dan ada beberapa kata yang hanya akan muncul di satu atau sedikit dokumen, sehingga untuk selanjutnya akan dilakukan proses reduksi dimensi dengan membuang kata-kata itu.

b. Perhitungan Metode K Means Clustering

Dari data setelah di reduksi dimensi, akan diolah menggunakan metode K Means untuk menghasilkan cluster data.

2.3.4.5 Penarikan Kesimpulan

Setelah dilakukan pengamatan dari beberapa skenario percobaan, dilihat bagaimana performa metode K Means, dan dilihat bagaimana kondisi terbaik yang menghasilkan akurasi terbaik.

2.3.4.6 Dokumentasi

Pada tahap terakhir ini dilakukan penyusunan laporan dan dokumentasi dari pelaksanaan secara keseluruhan.

2.4 K Means Clustering

Clustering merupakan proses mengelompokkan atau penggolongan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota satu kelas atau cluster dan meminimumkan kesamaan antar cluster [6].

Sementara clustering akan membagi data ke dalam grup-grup yang mempunyai objek yang karakteristiknya sama. Menurut Han & Kamber (2006), algoritme k-means bekerja dengan cara membagi data ke dalam k buah cluster yang telah ditentukan[7].

Perhitungan jarak yang digunakan dalam penelitian ini adalah cosine similarity. Tahap - tahap Algoritme dasar k-means seperti berikut:

1. Tentukan jumlah k sebagai cluster yang ingin dibentuk.
2. Menentukan pusat cluster secara acak sebanyak k.
3. Menentukan jarak setiap data terhadap pusat cluster(centroid)
4. Mengelompokkan setiap data yang bersangkutan berdasarkan kedekatannya dengan centroid (jarak terkecil).
5. Menentukan pusat cluster baru. Memperbaharui nilai centrid dari rata-rata cluster yang bersangkutan dengan menggunakan persamaan (1)

$$y_j(t + 1) = \frac{1}{N_{sj}} \sum X_j \dots \dots (1)$$

6. Ulangi langkah 3 hingga 5 sampai anggota yang ada pada tiap cluster tidak berubah.
7. Jika langkah 6 sudah terpenuhi, maka nilai pusat cluster pada perulangan terakhir akan digunakan sebagai parameter untuk kelompok dokumen fatwa.

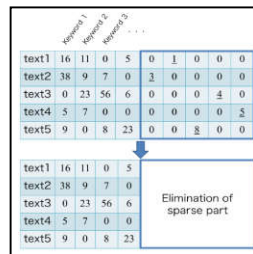
2.5 Reduksi Dimensi Berdasarkan Karakteristik Data Artikel

Proses reduksi dimensi dilakukan untuk mengecilkan dimensi data sehingga waktu komputasi dibutuhkan lebih sedikit. Namun proses reduksi dimensi harus memperhatikan karakteristik data, karena dimensi yang hilang bisa jadi juga menghilangkan karakteristik data[8]. Oleh karena itu dalam penelitian ini dipilih reduksi dimensi dengan menghilangkan sparse part. Term yang memiliki frekuensi yang kebanyakan bernilai nol akan dihilangkan, karena term tersebut merepresentasikan term yang sangat jarang dijumpai dalam data atau dengan kata lain term boleh diabaikan.

Besarnya pengurangan dimensi term sesuai dengan parameter yang ditentukan. Karena ditakutkan proses ini akan menghilangkan karakteristik data, maka dibuat parameter nilai yang berfungsi untuk membatasi reduksi dimensi. Sehingga meskipun dalam suatu term kebanyakan nilainya nol, term tersebut tidak akan dihilangkan jika ada salah satu dari nilai frekuensinya melebihi parameter yang telah ditentukan.[9]

Data TF yang sudah masuk dalam database memiliki nilai yang berbeda-beda. Term yang memilik TF yang kebanyakan bernilai nol dapat diabaikan, dengan catatan tidak ada nilai TF yang

melebihi parameter yang sudah ditentukan. Dengan kata lain proses reduksi dimensi adalah dengan cara menghilangkan fitur atau term yang yang frekuensinya semua/kebanyakan bernilai 0. Jika ada beberapa yang tidak bernilai 0 maka TF tersebut nilainya harus melebihi parameter tertentu. Hal ini dimaksudkan agar data tidak kehilangan karekteristiknya. Sehingga proses dapat berjalan dengan baik.[9]



Gambar 3. Ilustrasi Reduksi Dimensi

Dalam gambar 3, ditunjukkan misal parameter adalah 40%. Maka reduksi dimensi dilakukan dengan menghilangkan kolom yang nilainya bukan 0 kurang dari 40% data yang ada. Kolom ke 5-9 akan dihilangkan karena kolom tersebut tidak memenuhi syarat.

2.6 Algoritma Nazief & Adriani

Algoritma stemming untuk bahasa yang satu berbeda dengan algoritma stemming untuk bahasa lainnya. Sebagai contoh bahasa Inggris memiliki morfologi yang berbeda dengan bahasa Indonesia sehingga algoritma stemming untuk kedua bahasa tersebut juga berbeda. Proses stemming pada teks berbahasa Indonesia lebih rumit/kompleks karena terdapat variasi imbuhan yang harus dibuang untuk mendapatkan root word (kata dasar) dari sebuah kata. Pada umumnya kata dasar pada bahasa Indonesia terdiri dari kombinasi [10] :

$$\text{Prefiks 1} + \text{Prefiks 2} + \text{Kata dasar} + \text{Sufiks 3} + \text{Sufiks 2} + \text{Sufiks 1}$$

Algoritma Nazief & Adriani yang dibuat oleh Bobby Nazief dan Mirna Adriani ini memiliki tahap-tahap sebagai berikut: (Nazief, 1996)

1. Pertama cari kata yang akan diistem dalam kamus kata dasar. Jika ditemukan maka diasumsikan kata adalah root word, maka algoritma berhenti.
2. Inflection Suffixes (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa particles (“-lah”, “-kah”, “-tah” atau “-pun”) maka langkah ini diulangi lagi untuk menghapus Possesive Pronouns (“-ku”, “-mu”, atau “-nya”) jika ada.
3. Hapus Derivation Suffixes (“-i”, “-an” atau “-kan”). Jika kata ditemukan di kamus, maka algoritma berhenti. Jika tidak maka ke langkah 3a
 - a. Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
 - b. Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4
4. Hilangkan derivation prefixes DP {“di-”, “ke-”, “se-”, “me-”, “be-”, “pe”, “te-”} dengan iterasi maksimum adalah 3 kali:
5. Apabila setelah langkah 4 kata dasar masih belum ditemukan, maka proses recoding dilakukan dengan mengacu pada aturan pada Tabel 3. Recoding dilakukan dengan menambahkan karakter recoding di awal kata yang dipenggal. Pada Tabel 3, karakter recoding adalah huruf kecil setelah tanda hubung (‘-’) dan terkadang berada sebelum tanda kurung. Sebagai contoh, kata “menangkap” (aturan 15), setelah dipenggal menjadi “nangkap”. Karena tidak valid, maka recoding dilakukan dan menghasilkan kata “tangkap”.
6. Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai root word dan proses selesai.

2.7 Silhouette Coefficient

Silhouette Coefficient merupakan salah satu metode yang digunakan untuk menguji kualitas dan kekuatan dari sebuah cluster. Metode silhouette coefficient merupakan gabungan dari metode cohesion dan metode separation.

Metode cohesion sendiri merupakan suatu metode yang digunakan untuk mengukur seberapa dekat relasi antar objek dalam satu cluster yang sama. Sedangkan metode separation digunakan untuk mengukur seberapa jauh sebuah cluster terpisah dengan cluster yang lain.

Silhouette memiliki tiga tahap dalam perhitungannya, Berikut tahap perhitungan silhouette coefficient [11]:

- a. Menghitung rata-rata jarak objek dengan semua dokumen yang berada dalam satu cluster dengan menggunakan persamaan (2)

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \dots\dots\dots (2)$$

- b. Kemudian menghitung jarak objek dengan semua dokumen antar cluster dengan menggunakan persamaan (3)

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \dots\dots\dots (3)$$

- c. Kemudian menghitung nilai silhouette dengan menggunakan persamaan (4)

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \dots\dots\dots (4).$$

3. HASIL DAN PEMBAHASAN

2.1. Preprocessing

Tahapan preprocessing :

1. Tokenisasi

Untuk proses tokenisasi, dalam membaca isi file pdf pada penelitian ini menggunakan library itextsharp , library ini cukup bagus dalam merubah isi file pdf menjadi text.

2. Stopword and Stoplist Removal

Term yang dihasilkan pada proses sebelumnya bermacam-macam, diantara term yang ada, ada beberapa term yang dianggap tidak berguna atau tidak bernilai, kata-kata tersebut contohnya seperti kata sambung, kata depan, dll. Ada total 694 kata yang termasuk dalam daftar kata tersebut. Yang selanjutnya dapat dihilangkan untuk dilakukan proses selanjutnya.

3. Stemming

Metode stemming yang dipakai adalah Algoritma Nazief & Adriani. Dari data dokumen fatwa yang ada, setelah dilakukan proses tokenisasi, stopword removal, selanjutnya untuk setiap kata, diambil kata dasarnya dengan algoritma tersebut dengan menerapkan regular expression.

2.1. Indexing

Proses selanjutnya setelah preprosesing adalah indexing. Dari dokumen 114 fatwa MUI yang ada, dilakukan preprocessing, dan merubahnya menjadi matrix term frequent, dihasilkan 1531 kata (term).

3.3 Reduksi Dimensi

Dari 114 data dokumen Fatwa setelah diproses menghasilkan 1531 term (kata). Dari 1531 kata yang ada, ada beberapa kata yang hanya muncul di satu atau dua dokumen saja, sehingga kata tersebut dapat dihiraukan. Oleh karena itu, untuk menghilangkan kata-kata tersebut, dan juga untuk mempercepat proses perhitungan, dilakukan proses reduksi dimensi.

Hasil jumlah term yang dihasilkan setelah dilakukan proses reduksi dimensi ada pada tabel 2 berikut:

Table 2. Reduksi Dimensi

Nilai Reduksi Dimensi	Jumlah Term
Semua	1531
10 %	142
20 %	85
30 %	53
40 %	36
50 %	26
60 %	17
70 %	13
80 %	12
90 %	5

Nilai % diatas berarti kata-kata yang diambil adalah kata yang muncul di lebih dari berapa % dokumen. Semakin sedikit parameter, maka akan semakin sedikit pula term yang dihasilkan.

Hasil Ujicoba

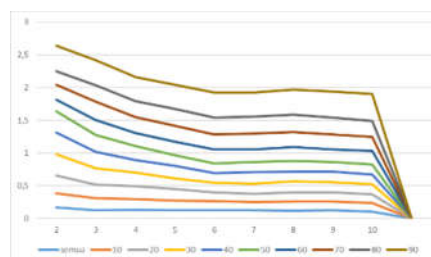
Setelah dilakukan proses reduksi dimensi, selanjutnya dilakukan proses clustering menggunakan K-Means, dengan skenario reduksi dimensi yang berbeda dan jumlah cluster yang berbeda pula, menghasilkan nilai *Silhouette Coefficient* seperti pada tabel 3 berikut:

Tabel 3. Nilai *Silhouette Coefficient*

Nilai Reduksi Dimensi	Jumlah Cluster	Silhouette Coefficient
Semua	1	0.1649371946060189
	2	0.12279165341720683
	3	0.13094693386662373
	4	0.12124174878015036
	5	0.12495083075064418
	6	0.1244709201902999
	7	0.1150903333536572
	8	0.12064776006622427
	9	0.0994699359772012
	10	0.1649371946060189
10 %	1	0.2144288464388275
	2	0.1812965580770138
	3	0.16285280984387443
	4	0.14925202054926154
	5	0.13818131815563744
	6	0.119962854205404
	7	0.14069639326717095
	8	0.13698631701949565
	9	0.13098312673626197
	10	0.2144288464388275
20 %	1	0.2729950242160394
	2	0.20968726410975805
	3	0.19563359392854573
	4	0.17485867458215276
	5	0.1342581939189252
	6	0.13161799627307585
	7	0.14154789594362577
	8	0.13922407596279435
	9	0.13515472390073188
	10	0.2729950242160394
30 %	1	0.32600398146007575
	2	0.2532506901403773
	3	0.21108398948167365
	4	0.16325211960996985
	5	0.14916841347306525
	6	0.15526875340543358
	7	0.16642560322077452
	8	0.1515127335717922
	9	0.15466482901713619
	10	0.32600398146007575
40 %	1	0.329624933383928
	2	0.2514621548901702
	3	0.190130194700792
	4	0.19287410674513245
	5	0.1411363560021326

	6	0.17502081325938906
	7	0.14810789662027868
	8	0.16198442139821792
	9	0.15205629032324183
	10	0.329624933383928
	1	0.32600398146007575
	2	0.2532506901403773
	3	0.21108398948167365
	4	0.16325211960996985
50 %	5	0.14916841347306525
	6	0.15526875340543358
	7	0.16642560322077452
	8	0.1515127335717922
	9	0.15466482901713619
	10	0.32600398146007575
	1	0.18115956477308773
	2	0.23059548215881986
	3	0.20314281425548864
	4	0.20650599762433872
	5	0.21472454507956212
60 %	6	0.19076784096144808
	7	0.2111980464354233
	8	0.19216209743025658
	9	0.20420568461477492
	10	0.18115956477308773
	1	0.23010869480249674
	2	0.2781602299478894
	3	0.2464817557624663
	4	0.2421251904589407
	5	0.23333114404313102
70 %	6	0.24454471880262768
	7	0.22809354933963893
	8	0.23043826883271196
	9	0.21122942009780532
	10	0.23010869480249674
	1	0.20501599480013052
	2	0.2549665654959047
	3	0.24234693598971377
	4	0.2577037803707858
	5	0.2585567138631424
80 %	6	0.25846169331829627
	7	0.26418921915536747
	8	0.2555517486921774
	9	0.24499837764482585
	10	0.20501599480013052
	1	0.38923906481625614
	2	0.3816575830450635
	3	0.36863147657601486
	4	0.36829006115330776
	5	0.3808156497345422
90 %	6	0.36592893453646164
	7	0.3887861244090915
	8	0.3997905614211763
	9	0.4169874444020241
	10	0.38923906481625614

Hasil dari reduksi dimensi dengan Silhouette Coefficient ditunjukkan oleh gambar 4 di bawah ini:



Gambar 4. Hasil Skenario Ujicoba

Dari hasil di atas diketahui bahwa reduksi dimensi dan jumlah cluster mempengaruhi kualitas cluster yang dihasilkan.

4. SIMPULAN

Dari hasil penelitian didapatkan bahwa:

1. Reduksi dimensi dan jumlah cluster mempengaruhi kualitas cluster yang dihasilkan.
2. Berdasarkan hasil Silhouette Coefficient yang dihasilkan, semakin kecil reduksi dimensi, ternyata nilainya semakin besar, menunjukkan kualitas cluster yang kurang baik.
3. Jumlah cluster 10 memiliki nilai Silhouette Coefficient terkecil di setiap scenario reduksi dimensi.

Metode K-Means cukup baik dalam mengelompokkan dokumen Fatwa MUI dengan nilai Silhouette Coefficient terkecil 0.0994699359772012 pada saat tanpa reduksi dimensi dan jumlah cluster 9.

5. SARAN

Saran-saran untuk untuk penelitian lebih lanjut untuk menutup kekurangan penelitian ini:

1. Perlu dilakukan pengecekan clustering dokumen fatwa MUI menggunakan metode yang lain
2. Perlu adanya penelitian yang membandingkan metode – metode yang ada dalam mengcluster document fatwa sehingga dapat diketahui mana metode yang paling optimal.

DAFTAR PUSTAKA

- [1] [1] Santoso, Budi. 2007. Data Mining: Teknik Pemanfaatan data untuk keperluan bisnis, Yogyakarta : Graha Ilmu
- [2] Slamet, Cepy, et.al. 2016. Clustering the Verses of the Holy Qur'an using K-Means Algorithm . Asian Journal of Information Technology 5159-5162 ISSN: 1682-3925. <URL: <http://www.medwelljournals.com/abstract/?doi=ajit.2016.5159.5162>> .
- [3] Nugraha , Deka Dwinavinta Candra,, et.all. 2014. Klasterisasi Judul Buku dengan Menggunakan Metode K – Means. Seminar Nasional Aplikasi Teknologi Informasi (SNATI).<URL: <http://journal.uui.ac.id/Snati/article/view/3282>>.
- [4] Yulian, Eko. 2018. Text Mining dengan K-Means Clustering pada Tema LGBT dalam Arsip Tweet Masyarakat Kota Bandung. JURNAL MATEMATIKA “MANTIK”. Vol. 04 No. 01. <URL:<http://journal.uui.ac.id/Snati/article/view/3282>>.
- [5] Agusta, Ledi. 2009 . Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia. Konferensi Nasional Sistem dan Informatika. 106-39
- [6] Tan, P.N., Steinbach, M. & Kumar, V., 2006. Introduction to Data Mining. Boston : Pearson Education
- [7] Han, J & Kamber, M., 2006. Data Mining Concept and Techniques Second Edition. Han, J & Kamber, M., 2006. Data Mining Concept and Techniques Second Edition.
- [8] Yang, Yaming dan Pedersen, 1997, J.O, A Comparative Study on Feature Selection in Text Categorization, School of Computer Science, Carnegie Mellon University, US
- [9] Toyota, T., Nobuhara, H. Visualization of the Internet News Based on Efficient Self-Organizing Map Using Restricted Region Search and Dimensionality Reduction. JACIII. 16: 222. 2012.
- [10] Nazief, Bobby dan Mirna Adriani. 1996. Confix-Stripping: Approach to Stemming Algorithm for Bahasa Indonesia, Faculty of Computer Science University of Indonesia.
- [11] Handoyo, R., Rumani, R.M. & Michrandi, S.N. 2014. Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K-Means pada Pengelompokan Dokumen. JSM STMIK Mikroskil. Vol. 15, No. 2.

