

**Pengelompokan Tematik Al-Qur'an Menggunakan Metode *LSA*
dengan Pembobotan *TF* dan *Probabilistic-IDF***

SKRIPSI

Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Teknik Pada Fakultas Teknik
Universitas Islam Riau Pekanbaru



JHODY DAMARA
153510502

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS ISLAM RIAU
PEKANBARU
2020**

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Dengan menyebut nama Allah SWT yang Maha Pengasih lagi Maha Penyayang, Penulis ucapkan puji syukur atas kehadiran-Nya, yang telah melimpahkan rahmat, hidayah, dan inayah-Nya kepada penulis, sehingga penulis dapat menyelesaikan proposal skripsi yang berjudul “Pengelompokan Tematik Al-Qur'an Menggunakan Metode *LSA* dengan Pembobotan *TF* dan *Probabilistic-IDF*” ini tepat pada waktunya.

Dalam penyusunan skripsi ini, Penulis telah banyak mendapat hambatan dan halangan yang berarti. Untuk itu, Penulis menyampaikan ucapan terima kasih kepada semua pihak yang telah banyak membantu dalam pembuatan skripsi ini.

1. Prof. Dr. H. Syafrinaldi SH.,MCL selaku rektor Universitas Islam Riau.
2. Bapak Dr. Eng. Muslim, MT selaku Dekan Fakultas Teknik.
3. Ibu Ause Labellapansa, ST., M.Kom selaku ketua Program Studi Teknik Informatika.
4. Nesi Syafitri, S.Kom., M.Kom selaku penasehat akademik yang senantiasa memberikan masukan dan motivasi kepada penulis.
5. Bapak Dr. Arbi Haza Nasution, M.IT., B.IT selaku pembimbing yang telah ikhlas dan sabar memberikan bimbingan, motivasi dan arahan di tengah kesibukan beliau.
6. Seluruh Dosen Prodi Teknik Informatika yang mendidik serta memberi arahan.

7. Kepada Seluruh Staff Tata Usaha yang telah membantu dalam kelancaran proses penyelesaian skripsi ini.
8. Kedua orang tua dan keluarga, yang selalu mendo'akan, serta memberikan dukungannya.

Terlepas dari semua itu, penulis menyadari sepenuhnya bahwa masih ada kekurangan, baik dari segi susunan kalimat maupun tata bahasanya. Oleh karena itu dengan tangan terbuka penulis menerima segala saran dan kritik agar penulis dapat menyempurnakan laporan ini.

Akhir kata penulis berharap semoga laporan ini dapat memberikan manfaat, inspirasi, dan dapat dipergunakan dengan baik.

Wassalamu'alaikum Wr. Wb.

Pekanbaru, April 2020

Jhody Damara

AL-QUR'AN THEMATIC CLUSTERING USING LSA METHODS WITH TF WEIGHTING AND PROBABILISTIC-IDF

Jhody Damara
Faculty of Engineering
Informatics Engineering Program
Islamic University of Riau
Email: jhodydmr@student.uir.ac.id

ABSTRACT

Al-Qur'an is the Muslim holy book which contains the words of Allah SWT, which were revealed in Arabic to the Prophet Muhammad and read them as worthy of worship. The Qur'an functions as a guide / guide for humanity in achieving happiness in living in this world and the hereafter. In studying the science of the Qur'an, a person is not only required to read it at least must know about the grouping of verses based on themes contained in the Qur'an. Where the grouping can function to further study the theme. The problem is that there is no corpus of the Qur'anic thematic grouping that explains the theme of hell which is fast and accurate. This research aims to improve the efficiency of scientific search and reduce the difficulty in grouping surahs and verses. In this study the dataset used is, the Al-Qur'an translation dataset on the Indonesian version of qurandatabase.org. This grouping is done using the Latent Semantic Analysis (LSA) method by weighting Term Frequency and Probabilistic-IDF which is an algebraic statistical method that extracts the hidden semantic structure from words and sentences. This research resulted in a grouping with the sub theme "The Character of Hell Experts and Their Crimes" which obtained a score of f-score score of the Al-Qur'an Cordoba evaluation of 71.85% from the U matrix assessment, while from the Qur'an thematic experts obtained additional contributions 28 surahs and verses into the sub theme "The Character of Experts of Hell and Their Crimes" with an increase in f1-score of 9.32%.

Keywords: *Al-Qur'an, Cordoba, Latent Semantic Analysis (LSA), Term Frequency, Probabilistic-Inverse Document Frequency (IDF)*

**PENGELOMPOKAN TEMATIK AL-QUR'AN MENGGUNAKAN
METODE *LSA* DENGAN PEMBOBOTAN *TF*
DAN *PROBABILISTIC-IDF***

Jhody Damara
Fakultas Teknik
Program Studi Teknik Informatika
Universitas Islam Riau
Email : jhodydmr@student.uir.ac.id

ABSTRAK

Al-Qur'an adalah kitab suci umat islam yang berisi firman-firman Allah SWT, yang diwahyukan dalam bahasa Arab kepada Nabi Muhammad SAW dan membacanya bernilai ibadah. Al-Qur'an berfungsi sebagai petunjuk/pedoman bagi umat manusia dalam mencapai kebahagiaan hidup di dunia dan akhirat. Dalam mempelajari ilmu Al-Qur'an, seseorang tidak hanya diwajibkan membacanya minimal harus mengetahui tentang pengelompokan ayat-ayat berdasarkan tema yang terdapat dalam Al-Qur'an. Dimana pengelompokan tersebut dapat berfungsi untuk mengkaji lebih lanjut mengenai tema tersebut. Maka yang menjadi permasalahan adalah belum adanya korpus pengelompokan tematik Al-Qur'an yang menjelaskan tentang tema neraka yang cepat dan akurat. Penelitian ini bertujuan untuk meningkatkan efisiensi pencarian ilmiah dan mengurangi kesulitan dalam pengelompokan surah dan ayat. Dalam penelitian ini dataset yang digunakan yaitu, dataset terjemahan Al-Qur'an pada *qurandatabase.org* versi indonesia. Pengelompokan ini dilakukan menggunakan metode *Latent Semantic Analysis (LSA)* dengan pembobotan *Term Frequency* dan *Probabilistic-IDF* yang merupakan metode statistik aljabar yang mengekstrak struktur semantik yang tersembunyi dari kata dan kalimat. Penelitian ini menghasilkan pengelompokan dengan sub tema "Sifat Ahli Neraka dan Kejahatan Mereka" yang memperoleh nilai *f1-score* evaluasi Al-Qur'an Cordoba sebesar 71,85% dari penilaian matriks U , sedangkan dari pakar tematik Al-Qur'an diperoleh kontribusi penambahan 28 surah dan ayat ke dalam sub tema "Sifat Ahli Neraka dan Kejahatan Mereka" dengan peningkatan *f1-score* 9,32%.

Kata kunci : Al-Qur'an, Cordoba, *Latent Semantic Analysis(LSA)*, *Term Frequency*, *Probabilistic-Invers Document Frequency(IDF)*

DAFTAR ISI

	Halaman
KATA PENGANTAR	i
ABSTRACT	iii
ABSTRAK	iv
DAFTAR ISI	v
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	x
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah.....	4
1.3 Rumusan Masalah.....	4
1.4 Batasan Masalah.....	5
1.5 Tujuan Penelitian.....	6
1.6 Manfaat Penelitian.....	6
BAB II LANDASAN TEORI	7
2.1 Studi Kepustakaan.....	7
2.2 Dasar Teori.....	8
2.2.1 Tematik.....	9
2.2.2 Al-Qur'an.....	9
2.2.3 Neraka.....	10
2.2.4 Sub-sub Tema Neraka.....	11
2.2.5 Phyton.....	13
2.2.6 <i>Text Preprocessing</i>	14
2.2.7 <i>Bag Of Words</i>	15

2.2.8	<i>Latent Semantic Analysis (LSA)</i>	16
2.2.9	<i>Term Frequency dan Inverse Document Frequency</i>	17
2.2.10	<i>Singular Value Decomposition (SVD)</i>	19
2.2.11	Evaluasi.....	20
2.2.12	<i>Flowchart</i>	21
BAB III METODOLOGI PENELITIAN		23
3.1	Alat dan Bahan Penelitian.....	23
3.1.1	Alat Penelitian.....	23
3.1.2	Bahan Penelitian	23
3.2	Analisis Sistem.....	24
3.2.1	Gambaran Umum Sistem.....	25
3.3	Rancangan Sistem.....	27
3.3.1	Text Preprocessing.....	27
3.3.2	<i>Clean Data</i>	30
3.3.3	<i>Bag of words</i>	31
3.3.4	<i>Latent Semantic Analysis (LSA)</i>	31
3.3.5	Pembobotan <i>TF</i> dan <i>Probabilistic-IDF</i>	32
3.3.6	<i>Singular Value Decomposition (SVD)</i>	36
3.3.7	Evaluasi.....	38
BAB IV HASIL DAN PEMBAHASAN		40
4.1	Data Tematik Al-Qur'an Tema Neraka	40
4.2	<i>Text Preprocessing</i>	41
4.3	<i>Bag of words</i>	44
4.4	Pembobotan kata terhadap <i>Bag of words</i>	44
4.4.1	<i>Term Frequency</i>	46
4.4.2	<i>Probabilistic Inverse Document Frequency</i>	47
4.4.3	<i>Term Frequency</i> dan <i>Probabilistic Inverse Document Frequency</i>	47
4.5	Visualisasi Kemunculan <i>Bag of words</i> terhadap Dokumen.....	48

4.6	Matriks dari Hasil Perkalian <i>TF</i> dan <i>Probabilistic IDF</i>	50
4.7	<i>Singular value decomposition (SVD)</i>	51
4.8	Pengelompokan Surah dan Ayat Perkonsep Berdasarkan Matriks U	52
4.9	Evaluasi Sistem.....	55
4.9.1	Data Evaluasi	55
4.9.2	Evaluasi Pengelompokan Berdasarkan Al-Qur'an Cordoba ...	58
4.9.3	Evaluasi Pakar Berdasarkan Hasil Data <i>False Positif</i> (FP)	62
4.9.4	Perbedaan Hasil Evaluasi Al-Qur'an Cordoba dengan Hasil Evaluasi Pakar.....	66
BAB V KESIMPULAN DAN SARAN		68
5.1	Kesimpulan	68
5.2	Saran	68
DAFTAR PUSTAKA		70

DAFTAR TABEL

	Halaman
Tabel 2. 1 <i>Variants of term frequency (TF) weight</i>	17
Tabel 2. 2 <i>Variants of inverse document frequency (IDF) weight</i>	18
Tabel 2. 3 Simbol dan Fungsi <i>Flowchart</i>	21
Tabel 3. 1 Dokumen Tema Neraka	34
Tabel 3. 2 <i>Bag of Words</i> Tema Neraka.....	34
Tabel 3. 3 Proses Perhitungan <i>TF</i>	35
Tabel 3. 4 Proses Perhitungan <i>Probabilistic-IDF</i>	35
Tabel 3. 5 Proses Perhitungan <i>TF</i> dan <i>Probabilistic-IDF</i>	36
Tabel 3. 6 Contoh Data Awal <i>SVD (U)</i>	37
Tabel 3. 7 Contoh Pengelompokan <i>SVD (U)</i>	37
Tabel 3. 8 Contoh Pembentukan Konsep.....	38
Tabel 4. 1 Data tematik Al-Qur'an tema neraka	40
Tabel 4. 2 Hasil <i>text preprocessing</i>	42
Tabel 4. 3 Daftar <i>Stopword Update</i>	43
Tabel 4. 4 <i>Bag of words</i> tema neraka.....	44
Tabel 4. 5 Set nilai nol terhadap <i>Bag of words</i>	45
Tabel 4. 6 Menghitung kemunculan <i>bag of words</i>	45
Tabel 4. 7 Hasil perhitungan <i>Term Frequency</i>	46
Tabel 4. 8 Hasil perhitungan <i>Probabilistic Inverse Document Frequency</i>	47
Tabel 4. 9 Hasil Perkalian <i>TF</i> dan <i>Probabilistic IDF</i>	48
Tabel 4. 10 Hasil pengelompokan <i>SVD (U)</i>	52
Tabel 4. 11 Inputan Data <i>Gold Standard</i>	55
Tabel 4. 12 Hasil Data Evaluasi Tematik Al-Qur'an Cordoba	59
Tabel 4. 13 Data <i>False Positif (FP)</i>	62
Tabel 4. 14 Hasil Evaluasi Pakar Tematik	63
Tabel 4. 15 Perbedaan Hasil Evaluasi Kedua Pakar	66
Tabel 4. 16 Perbedaan Hasil Evaluasi Al-Qur'an Cordoba dengan Pakar.....	67

DAFTAR GAMBAR

	Halaman
Gambar 3. 1 Diagram Alir Sistem.....	25
Gambar 3. 2 <i>Flowchart Text Preprocessing</i>	28
Gambar 3. 3 Proses <i>Case folding</i>	29
Gambar 3. 4 Proses <i>Tokenization</i>	29
Gambar 3. 5 Proses <i>Stemming</i>	29
Gambar 3. 6 Proses <i>filtering</i> menggunakan <i>stopwords</i>	30
Gambar 3. 7 <i>Bag Of Words Tema Neraka</i>	31
Gambar 3. 8 Flowchart <i>Latent semantic analysis</i>	32
Gambar 4. 1 Visualisasi Hasil Perkalian <i>TF dan Probabilistic IDF</i>	48
Gambar 4. 2 Grafik kemunculan <i>bag of words</i>	49
Gambar 4. 3 Inputan untuk Proses <i>SVD</i>	50
Gambar 4. 4 Nilai matriks <i>U</i>	51
Gambar 4. 7 Hasil Evaluasi <i>F1-Score</i> Terbaik	61

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Data Tema Neraka	
Lampiran 2 Hasil Evaluasi Pakar	
Lampiran 3 SK Pembimbing Skripsi	
Lampiran 4 Kartu Konsultasi Bimbingan Skripsi	
Lampiran 5 SK Komprehensif Skripsi	
Lampiran 6 Berita Acara Ujian Skripsi	
Lampiran 7 Surat Keterangan Bebas Plagiarisme	



BAB I PENDAHULUAN

1.1 Latar Belakang

Al-Qur'an adalah kitab suci umat Islam yang berisi firman-firman Allah SWT, yang diwahyukan dalam bahasa Arab kepada Nabi Muhammad SAW dan membacanya bernilai ibadah. Al-Qur'an berfungsi sebagai petunjuk/pedoman bagi umat manusia dalam mencapai kebahagiaan hidup di dunia dan akhirat. Dalam mempelajari ilmu Al-Qur'an, seseorang tidak hanya diwajibkan membacanya minimal harus mengetahui tentang pengelompokan ayat-ayat berdasarkan tema yang terdapat dalam Al-Qur'an. Dimana pengelompokan tersebut dapat berfungsi untuk mengkaji lebih lanjut mengenai tema tersebut.

Tematik Al-Qur'an adalah suatu metode untuk memahami makna tema yang terdapat dalam Al-Qur'an dengan cara menganalisis seluruh ayat Al-Qur'an tentang tema yang ingin dipahami. Contohnya pada saat penulis ingin memahami dan membahas tentang tema neraka. Pertama-tama yang penulis lakukan ialah mengumpulkan terlebih dahulu semua ayat Al-Qur'an yang membahas tentang tema neraka. Kemudian menganalisa ayat Al-Qur'an tersebut satu per-satu, bagaimana neraka menurut ayat per-ayat dalam Al-Qur'an, tujuannya untuk kemudahan seseorang dalam memahami isi Al-Qur'an tersebut berdasarkan tema neraka secara cepat dan akurat.

Didalam Al-Qur'an kata neraka dalam bahasa Arab disebut dengan istilah *Nar*. Kata itu mengandung pengertian api dan neraka. *Nar* yang mengandung pengertian api di dalam Al-Qur'an adalah sebagai berikut :

بَلَىٰ مَنْ كَسَبَ سَيِّئَةً وَأَحَاطَتْ بِهِ خَطِيئَتُهُ فَأُولَٰئِكَ أَصْحَابُ النَّارِ هُمْ فِيهَا خَالِدُونَ

“(Bukan demikian), yang benar: barangsiapa berbuat dosa dan ia telah diliputi oleh dosanya, mereka itulah penghuni neraka, mereka kekal di dalamnya.” (Q.S Al-Baqarah [2]: 81)

إِنَّ الْمُنَافِقِينَ فِي الدَّرَكِ الْأَسْفَلِ مِنَ النَّارِ وَلَنْ تَجِدَ لَهُمْ نَصِيرًا

“Sesungguhnya orang-orang munafik itu (ditempatkan) pada tingkatan yang paling bawah dari neraka. dan kamu sekali-kali tidak akan mendapat seorang penolongpun bagi mereka”. (QS. An-Nisa' [4] : 145)

Dengan demikian, dapat dikatakan bahwa neraka ialah suatu tempat penyiksaan yang penuh dengan kobaran api dan aneka macam siksaan yang sangat pedih, yang dipersiapkan oleh Allah di akhirat. Suatu tempat yang disediakan untuk memenjarakan, menghukum dan mengurung para musuh-musuh Allah dan rasul-rasul-Nya dan mereka yang selalu bergelimang dosa.

Seiring dengan kemajuan ilmu pengetahuan dan teknologi, beberapa penerbit buku yang menerbitkan Al- Qur'an dan Terjemahan melampirkan jenis kategori ayat-ayat Al-Qur'an ke dalam beberapa kelas. Dalam Al-Qur'an dan Terjemahan yang diterbitkan oleh penerbit Cordoba, melampirkan bahwa dalam Al-Qur'an terdapat 15 klasifikasi ayat-ayat Al- Qur'an. Namun tidak semua ayat memiliki

label kelas, ada beberapa ayat yang belum memiliki kelas. Sehingga ayat-ayat yang belum memiliki label kelas tersebut dapat diprediksi dengan menggali informasi dari ayat yang sudah memiliki label kelasnya (Hilwah, Kudus, & Sunendiari, 2017)

Solusi yang ditawarkan oleh penulis adalah mencoba untuk membuat pengelompokan tematik Al-Qur'an tentang tema neraka berdasarkan subtema, surat, dan terjemahannya yang merujuk pada Al-Qur'an Cordoba rujukan terlengkap atau disebut juga dengan Al-Qur'an amazing, sehingga menghasilkan korpus yang nantinya bisa digunakan untuk penelitian lebih lanjut. Menurut Baker (2010:93) korpus merupakan kumpulan teks baik tulisan lisan maupun lisan yang tersimpan dalam komputer. Baker mendefinisikan korpus terdapat pada media elektronik saja. Sedangkan Menurut Setiawan (2017) korpus merupakan kumpulan tulisan yang ditulis oleh seseorang baik berupa *hard copy* dan *soft copy*. Korpus dalam bentuk *hard copy* dapat dicontohkan seperti buku, majalah, kamus, dan koran. Contoh *soft copy* dapat berupa aplikasi, *website*, kamus online, dan lain sebagainya. Jadi kesimpulannya korpus merupakan kumpulan teks baik secara lisan maupun tulisan yang ada di media cetak maupun elektronik dan dapat dijadikan sumber data.

Berdasarkan uraian di atas, maka yang menjadi permasalahan adalah belum adanya korpus pengelompokan tematik Al-Qur'an yang menjelaskan tentang tema neraka yang cepat dan akurat. Maka dari itu dibutuhkan teknologi untuk mempermudah seseorang belajar dan memahami kandungan ayat Al-Qur'an. Metode tematik dipilih sebagai salah satu solusinya karena menghasilkan

korpus-korpus yang nantinya bisa digunakan untuk penelitian lebih lanjut. Contohnya “visualisasi tematik Al-Qur’an berbasis knowledge graph” yang dibuat oleh peneliti sebelumnya sehingga tampilan pengelompokan ayat lebih menarik dan mempermudah seseorang untuk memahaminya. Atas dasar latar belakang tersebut penulis tertarik untuk melakukan penelitian dengan Judul : **“Pengelompokan Tematik Al-Qur’an Menggunakan Metode LSA dengan Pembobotan TF dan Probabilistic-IDF”**.

1.2 Identifikasi Masalah

Adapun identifikasi masalah yang dapat diambil dari latar belakang tersebut adalah sebagai berikut:

1. Proses pencarian ayat Al-Qur’an tentang tema neraka masih menggunakan cara manual sehingga pengelompokan surah dan ayat dilakukan dengan cara mengumpulkan para hadist, tafsir, dan para ulama untuk membuat pengelompokan tematik Al-Qur’an.
2. Belum adanya korpus pengelompokan tematik Al-Qur’an tentang tema neraka.

1.3 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah disebutkan, maka permasalahan pada penelitian ini diambil sebuah rumusan masalah sebagai berikut:

1. Bagaimana cara membuat sebuah pengelompokan tematik Al-Qur'an berdasarkan terjemahan Al-Qur'an versi indonesia sesuai dengan tema neraka menggunakan metode *Latent Semantic Analysis*?
2. Bagaimana cara menyajikan data Al-Qur'an yang telah dikelompokan sesuai tema neraka dengan metode *Latent Semantic Analysis* ?

1.4 Batasan Masalah

Agar permasalahan tidak menyimpang dari maksud dan tujuan penyusunan proposal, penelitian dan juga mengingat luasnya permasalahan, penulis mempunyai pembatasan masalah yaitu :

1. Pengelompokan tematik Al-Qur'an ini hanya menyajikan ayat Al-Qur'an yang berkaitan dengan tema neraka.
2. Sumber data pengelompokan tematik Al-Qur'an ini diperoleh dari <http://qurandatabase.org/>
3. Pengelompokan tematik Al-Qur'an ini merujuk pada Al-Qur'an Cordoba rujukan terlengkap atau disebut juga dengan Al-Qur'an amazing.
4. Pengelompokan tematik Al-Qur'an ini menggunakan pembobotan *term frequency* dan *Probabilistic inverse document frequency*.
5. Tahap evaluasi dilakukan dengan dua cara yaitu, pertama dengan menggunakan Al-Qur'an Cordoba dan yang kedua dengan menggunakan pakar tematik Al-Qur'an.

1.5 Tujuan Penelitian

Adapun tujuan dari penelitian ini untuk mempermudah dalam membuat pengelompokan tematik Al-Qur'an secara otomatis sesuai dengan tema yang diinginkan secara efisien dan akurat sehingga dapat dengan mudah memahaminya, selain itu penelitian ini juga menghasilkan korpus yang nantinya bisa dimanfaatkan oleh penelitian lebih lanjut.

1.6 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut :

1. Tersedianya korpus pengelompokan tematik Al-Qur'an yang bertema neraka sehingga meningkatkan efisiensi dan efektifitas dalam mempelajari ayat Al-Qur'an yang menjelaskan tema neraka.
2. Mempermudah peneliti berikutnya dalam memperoleh data tentang pengelompokan Al-Qur'an tema Neraka.
3. Membantu para ahli tafsir Al-Qur'an dalam membuat pengelompokan untuk surah dan ayat yang bertemakan Neraka.

BAB II

LANDASAN TEORI

2.1 Studi Kepustakaan

Untuk menyusun skripsi ini, penulis juga menggunakan bahan acuan kepustakaan yang bersumber pada penelitian-penelitian sebelumnya. Hal ini berguna sebagai pembanding serta bahan referensi bagi penulis. Penelitian pertama yang menjadi rujukan yaitu penelitian yang dilakukan oleh Husni Teja Sukmana, (2018) mengenai “*Development of Semantic Web of Indonesian Qur'an*”. Penelitian ini bertujuan untuk menyediakan sebuah korpus Al-Qur'an berbahasa Indonesia dan melakukan analisis semantik terhadap korpus yang sudah dibentuk. Korpus ini diharapkan nantinya dapat dijadikan sebagai basis pengetahuan dalam membangun berbagai aplikasi di bidang penelusuran informasi, pengolahan bahasa manusia, dan komputasi linguistik pada domain Al-Qur'an berbahasa Indonesia.

Penelitian selanjutnya dilakukan oleh Mochamad Irfan Dary, (2015) membahas tentang “Analisis dan Implementasi *Short Text Similarity* dengan Metode *Latent Semantic Analysis* untuk Mengetahui Kesamaan Ayat Al-Qur'an”. Tujuan dari penelitian ini untuk membantu memahami sebuah konsep dengan cara mencari hubungan antara konsep tersebut dengan konsep lainnya menggunakan metode *Latent Semantic Analysis*. Dari hasil pengujian didapat bahwa dari berbagai dimensi yang digunakan, akurasi maksimum adalah 71% dan F-

Measure 40%. Lebih baik dibandingkan *term document matrix* biasa (tanpa *LSA*) dengan akurasi 85% tetapi F-Measure 0%.

Penelitian terakhir dilakukan oleh Luthfiarta, Zeniarja, & Salam, (2013) membahas tentang “Algoritma *Latent Semantic Analysis (LSA)* pada Peringkasan Dokumen Otomatis untuk Proses *Clustering* Dokumen”. Dalam penelitian ini algoritma *LSA* dapat melakukan proses reduksi kalimat dengan lebih baik dibandingkan algoritma *feature based* sehingga mendapatkan hasil akurasi proses *clustering* dokumen yang lebih akurat. Hasil penelitian menunjukkan tingkat akurasi menggunakan peringkasan dokumen otomatis dengan *LSA* dalam proses *clustering* dokumen mencapai 71,04% yang diperoleh pada tingkat peringkasan dokumen otomatis dengan *LSA* 40% dibandingkan dengan hasil *clustering* tanpa peringkasan dokumen otomatis yang hanya mencapai tingkat akurasi 65,97%.

Pada kasus ini penulis membahas tentang bagaimana cara metode *latent semantic analysis* mengelompokkan surah, ayat, dan terjemahannya sesuai dengan sub-sub tema yang ada pada Al-Qur’an Cordoba khusus tema neraka berdasarkan *bag of words* yang sudah terbentuk. Kemudian pengelompokan ini nantinya akan dievaluasi dengan menggunakan perhitungan *precision*, *recall*, dan *f1-score*.

2.2 Dasar Teori

Penelitian ini dilakukan tidak terlepas dari teori-teori yang sudah ada, dasar teori diperlukan untuk mengetahui sumber dari teori yang dikemukakan pada penelitian ini.

2.2.1 Tematik

Tafsir Maudhu'i (tematik) adalah mengumpulkan ayat-ayat Al-Qur'an yang mempunyai tujuan yang satu, bersama-sama membahas topic, judul, tema tertentu dengan menertibkannya sedapat mungkin sesuai dengan masa turunnya selaras dengan sebab-sebab turunnya. Kemudian memperhatikan ayat-ayat tersebut dengan penjelasan-penjelasan, keterangan-keterangan dan hubungannya dengan ayat lain kemudian mengistimbatkan hukum-hukum (Yamani, 2015).

Dari pengertian tersebut dapat dipahami bahwa tematik adalah metode yang menjelaskan beberapa ayat Al-Qur'an mengenai suatu judul ataupun tema tertentu, dengan memperhatikan urutan dari turunan masing-masing ayat, sesuai dengan turunannya yang dijelaskan dengan berbagai macam keterangan dari segala segi dan diperbandingkan dengan keterangan berbagai ilmu pengetahuan yang benar dan membahas topik ataupun tema yang sama, sehingga lebih mempermudah dan memperjelas masalah, karena Al-Qur'an banyak mengandung berbagai macam tema pembahasan yang perlu dibahas secara tematik, agar pembahasan dapat dilakukan dengan tuntas dan sempurna.

2.2.2 Al-Qur'an

Al-Qur'an adalah salah satu kitab suci yang terpelihara keasliannya, terdiri dari 6236 ayat, 114 surat dan 30 juz. Dalam Al-Qur'an terdapat 2 jenis surah yang dibedakan berdasarkan waktu dan tempat turunnya ayat yaitu Makkiyah dan Madaniyah (Jadhira et al., 2018).

Al-Qur'an merupakan wahyu yang diturunkan oleh Allah SWT kepada Nabi Muhammad SAW melalui perantara Malaikat Jibril dengan bahasa Arab,

sebagai mukjizat yang diturunkan secara mutawatir untuk dijadikan petunjuk dan pedoman hidup bagi setiap umat Islam yang ada di muka bumi. Al-Qur'an bukan sekedar mengatur hubungan manusia dengan Rabbnya, tetapi juga mengatur hubungan manusia dengan manusia, dan manusia dengan alam sekitarnya.

Pengelompokan ayat Al-Quran perlu kita ketahui, selain memudahkan kita untuk memahami semua ayat-ayat Al-Quran, juga agar kita lebih cepat untuk menemukan arti dari masing-masing ayat Al-Quran tersebut. Al-Quran dapat dikelompokkan kedalam berbagai pengelompokan seperti pengelompokan Al-Quran berdasarkan tematik, berdasarkan turunnya atau berdasarkan pembahasannya dan lain sebagainya.

2.2.3 Neraka

Neraka adalah tempat di alam akhirat yang diciptakan oleh Allah SWT untuk manusia dan jin yang tidak taat kepada Allah Swt dan bagi mereka yang tidak percaya bahwa Allah adalah tuhan semesta alam ketika manusia hidup di dalam dunia. Didalam neraka penuh dengan balasan kesedihan dan penderitaan bahkan lebih lagi penghuni menjadi makhluk hina sehina- hinanya. Allah berfirman dalam Q.S. Hud: 106-107 yang berbunyi :

فَأَمَّا الَّذِينَ شَفَعُوا فِي النَّارِ لَهُمْ فِيهَا زَفِيرٌ وَشَهِيقٌ

خَالِدِينَ فِيهَا مَا دَامَتِ السَّمَاوَاتُ وَالْأَرْضُ إِلَّا مَا شَاءَ رَبُّكَ إِنَّ رَبَّكَ فَعَّالٌ لِّمَا يُرِيدُ

Artinya: “Maka adapun orang-orang yang sengsara, maka (tempatnya) didalam neraka, disana mereka mengeluarkan dan menarik napas dengan merintih, :106* - Mereka kekal didalamnya selama ada langit dan bumi, kecuali jika Tuhanmu

menghendaki (yang lain). Sungguh Tuhanmu Maha Pelaksana terhadap apa yang Dia kehendaki “.:107

2.2.4 Sub-sub Tema Neraka

Berdasarkan rujukan dari Al-Qur'an Cordoba amzing yang menjadi subtema pada penelitian ini adalah sebagai berikut :

1. Nama-Nama Neraka, contohnya pada surah 20 ayat 74 :

إِنَّهُ مَنْ يَأْتِ رَبَّهُ مُجْرِمًا فَإِنَّ لَهُ جَهَنَّمَ لَا يَمُوتُ فِيهَا وَلَا يَحْيَىٰ

“Sesungguhnya barangsiapa datang kepada Tuhannya dalam keadaan berdosa, maka sesungguhnya baginya neraka Jahannam. Ia tidak mati di dalamnya dan tidak (pula) hidup”.

2. Keabadian Neraka, contohnya pada surah 2 ayat 39 :

وَالَّذِينَ كَفَرُوا وَكَذَّبُوا بِآيَاتِنَا أُولَٰئِكَ أَصْحَابُ النَّارِ هُمْ فِيهَا خَالِدُونَ

“Adapun orang-orang yang kafir dan mendustakan ayat-ayat Kami, mereka itu penghuni neraka; mereka kekal di dalamnya”.

3. Memasuki Neraka, contohnya pada surah 3 ayat 192 :

رَبَّنَا إِنَّكَ مَنْ تُدْخِلِ النَّارَ فَقَدْ أَخْرَيْتَهُ وَمَا لِلظَّالِمِينَ مِنْ أَنْصَارٍ

“Ya Tuhan kami, sesungguhnya barangsiapa yang Engkau masukkan ke dalam neraka, maka sungguh telah Engkau hinakan ia, dan tidak ada bagi orang-orang yang zalim seorang penolongpun”.

4. Penjaga Neraka, contohnya pada surah 43 ayat 77 :

وَنَادُوا يَا مَالِكُ لِيَقْضِ عَلَيْنَا رَبُّكَ قَالَ إِنَّكُمْ مَأْكُونُونَ

“Mereka berseru: Hai Malik biarlah Tuhanmu membunuh kami saja. Dia menjawab: Kamu akan tetap tinggal (di neraka ini)”.

5. Sifat Neraka, contohnya pada surah 25 ayat 66 :

إِنَّهَا سَاءَتْ مُسْتَقَرًّا وَمُقَامًا

“Sesungguhnya jahannam itu seburuk-buruk tempat menetap dan tempat kediaman”.

6. Sifat Ahli Neraka dan Kejahatan Mereka, contohnya pada surah 2 ayat 162 :

خَالِدِينَ فِيهَا لَا يُخَفَّفُ عَنْهُمْ الْعَذَابُ وَلَا هُمْ يُنظَرُونَ

“Mereka kekal di dalam laknat itu; tidak akan diringankan siksa dari mereka dan tidak (pula) mereka diberi tangguh”.

7. Siksaan Ahli Neraka Dilipatgandakan, contohnya pada surah 25 ayat 69 :

يُضَاعَفْ لَهُ الْعَذَابُ يَوْمَ الْقِيَامَةِ وَيَخْلُدْ فِيهِ مُهَانًا

“(yakni) akan dilipat gandakan azab untuknya pada hari kiamat dan dia akan kekal dalam azab itu, dalam keadaan terhina”.

8. Mereka yang Kekal Dalam Neraka, contohnya pada surah 98 ayat 6 :

إِنَّ الَّذِينَ كَفَرُوا مِنْ أَهْلِ الْكِتَابِ وَالْمُشْرِكِينَ فِي نَارِ جَهَنَّمَ خَالِدِينَ فِيهَا أُولَئِكَ هُمْ شَرُّ الْبَرِيَّةِ

“Sesungguhnya orang-orang yang kafir yakni ahli Kitab dan orang-orang yang musyrik (akan masuk) ke neraka Jahannam; mereka kekal di dalamnya. Mereka itu adalah seburuk-buruk makhluk”.

9. Makanan dan Minuman Ahli Neraka, contohnya pada surah 14 ayat 16 :

مِنْ وَرَائِهِ جَهَنَّمُ وَيُسْقَىٰ مِنْ مَاءٍ صَدِيدٍ

“di hadapannya ada Jahannam dan dia akan diberi minuman dengan air nanah”.

10. Pohon Zaquum, contohnya pada surah 37 ayat 63 :

إِنَّا جَعَلْنَاهَا فِتْنَةً لِلظَّالِمِينَ

“Sesungguhnya Kami menjadikan pohon zaqqum itu sebagai siksaan bagi orang-orang yang zalim”.

11. Percakapan Ahli Neraka, contohnya pada surah 33 ayat 67 :

وَقَالُوا رَبَّنَا إِنَّا أَطَعْنَا سَادَتَنَا وَكُبَرَاءَنَا فَأَضَلُّونَا السَّبِيلًا

“Dan mereka berkata: "Ya Tuhan kami, sesungguhnya kami telah mentaati pemimpin-pemimpin dan pembesar-pembesar kami, lalu mereka menyesatkan kami dari jalan (yang benar)”.

2.2.5 Phyton

Python adalah salah satu bahasa pemrograman tingkat tinggi yang bersifat interpreter, interactive, object-oriented, dan dapat beroperasi hampir di semua platform: Mac, Linux, dan Windows. Python termasuk bahasa pemrograman yang mudah dipelajari karena sintaks yang jelas, dapat dikombinasikan dengan penggunaan modul-modul siap pakai, dan struktur data tingkat tinggi yang efisien (Prasetya & Nurviyanto, 2012).

Bahasa ini muncul pertama kali pada tahun 1991, dirancang oleh seorang bernama Guido van Rossum. Sampai saat ini Python masih dikembangkan oleh *Python Software Foundation*. Seperti halnya bahasa pemrograman dinamis lainnya, python umumnya digunakan sebagai bahasa skrip meski pada praktiknya penggunaan bahasa ini lebih luas mencakup konteks pemanfaatan yang umumnya tidak dilakukan dengan menggunakan bahasa skrip.

2.2.6 *Text Preprocessing*

Text preprocessing merupakan suatu proses menormalisasikan teks pada dokumen sehingga informasi yang dimuat merupakan bagian yang padat dan ringkas namun tetap merepresentasikan informasi yang termuat didalamnya. Pada umumnya, proses ini dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah diproses oleh sistem (Mujilahwati, 2016).

Dalam tahap ini, terdapat beberapa proses diantaranya sebagai berikut :

a. *Case folding*

Proses *case folding* dilakukan untuk mengubah kapitalisasi karakter (huruf) menjadi kecil untuk semua kata atau huruf mulai dari ‘a’ sampai ‘z’. Contohnya “Saya Suka Belajar Bahasa Python” maka akan berubah menjadi “saya suka belajar bahasa python”.

b. *Tokenization*

Proses *tokenization* dilakukan untuk pemotongan string input berdasarkan tiap kata yang menyusunnya dengan kata lain memisahkan kalimat menjadi tiap-

tiap kata. Contohnya “saya adalah mahasiswa uir” maka akan berubah menjadi “saya”, “adalah”, “mahasiswa”, “uir”.

c. Stemming

Tahap *stemming* merupakan proses untuk mengubah semua kata yang telah dipilih pada proses *tokenization* menjadi kata yang berupa kata dasar. Contohnya “mempunyai” maka akan berubah menjadi “punya”.

d. Filtering

Tahap *filtering* adalah tahap mengambil kata-kata penting dari hasil token dan kemudian membuang kata-kata yang tidak penting dengan menggunakan *stopwords* sehingga nantinya menghasilkan *bag of words*.

e. Stopwords

Stopwords merupakan kumpulan daftar kata-kata yang kemungkinan besar tidak akan memberikan pengaruh prediksi, seperti imbuhan dan kata ganti. Kata yang termasuk *stopwords* akan dibuang, karena kata tersebut tidak merepresentasikan isi dokumen walaupun sering muncul. Contohnya “yang”, “dan”, “di”, “dari”, “untuk”, dan lain-lain.

2.2.7 Bag Of Words

Semua dokumen dapat direpresentasikan secara sederhana menggunakan *Bag of Words* (BoW). BoW adalah sebuah model yang merepresentasikan objek secara global misalnya kalimat teks atau dokumen sebagai bag (multiset) kata tanpa memperdulikan tata bahasa bahkan urutan kata untuk menjaga keanekaragamannya. Dengan kata lain, BoW merupakan kumpulan kata-kata unik atau kata kunci dalam dokumen (Mardiana & Nyoto, 2015).

BoW digunakan dalam pengolahan bahasa alami dan pencarian informasi, dan dikenal sebagai model ruang vektor. Pada model ini, tiap kalimat dalam dokumen digambarkan sebagai *token*, mengabaikan tata bahasa dan bahkan urutan kata namun menghitung frekuensi kejadian atau kemunculan kata dari dokumen. BoW merupakan sebuah konsep yang diambil dari analisis teks, yaitu merepresentasikan dokumen sebagai sebuah kantung informasi-informasi penting tanpa mengurutkan setiap katanya.

2.2.8 *Latent Semantic Analysis (LSA)*

Latent Semantic Analysis (LSA) adalah suatu metode untuk mengekstrak sebuah tulisan dalam suatu dokumen dan kemudian mengaplikasikannya dalam perhitungan matematis. Penilaian dengan metode ini lebih kepada kata-kata yang ada dalam tulisan tanpa memperhatikan urutan kata dan tata bahasa dalam tulisan tersebut, sehingga suatu kalimat yang dinilai adalah berdasarkan kata-kata kunci yang ada pada kalimat tersebut (Winata, F., & Rainarli, E., 2016).

LSA merupakan salah satu algoritma pengembangan dari bidang ilmu *Information Retrieval*, yang mampu menghimpun sejumlah besar dokumen dalam basis data dan menghubungkan relasi antar dokumen dengan mencocokkan kueri (*query*) yang diberikan. Lebih spesifik lagi, algoritma *LSA* merupakan suatu metode pembuatan representasi istilah (*term*) berbasis vektor yang dianggap mampu menangkap inti sari (*semantic*) dari suatu dokumen maupun kalimat.

Menurut Landauer dan Dumais (1997) cara kerja *LSA* ialah dengan menghasilkan sebuah model yang didapat dengan mencatat kemunculan-kemunculan kata dari tiap-tiap dokumen yang direpresentasikan dalam sebuah

matriks yang dinamakan *term-document matrix*, maka setelah itu dilakukan proses *Singular Value Decomposition (SVD)* yang akan digunakan untuk mendapatkan *Cosine Similarity* (nilai kemiripan) antara satu dokumen dengan dokumen lainnya.

2.2.9 Term Frequency dan Inverse Document Frequency

Metode *Term Frequency* dan *Invers Document Frequency (TF-IDF)* merupakan metode yang digunakan menentukan seberapa jauh keterhubungan kata (*term*) terhadap dokumen dengan memberikan bobot setiap kata. Metode *TF-IDF* ini menggabungkan dua konsep yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut (Fitri, 2013).

Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai *TF* dan *IDF* pada setiap *token* (kata) di setiap dokumen dalam korpus. *Term frequency (TF)* adalah frekuensi dari kemunculan sebuah *term* dalam dokumen yang bersangkutan.

Berikut ini jenis-jenis formula pembobotan *Term Frequency* :

Tabel 2. 1 Variants of term frequency (TF) weight

Skema Pembobotan	TF Weigth
<i>Binary</i>	0,1
<i>Raw count</i>	<i>ft, d</i>
<i>Term frequency</i>	$ft, d / \sum_{t' \in d} ft', d$
<i>Log normalization</i>	log (1+ ft, d)
<i>Double normalization 0.5</i>	$0.5 + 0.5 \cdot \frac{ft, d}{\max\{t' \in d\}ft', d}$

(Sumber : Poletini, 2004)

Keterangan :

$f(t, d)$ pilihan paling sederhana adalah dengan menggunakan hitungan mentah dari suatu istilah dalam dokumen yaitu, berapa kali istilah t muncul dalam dokumen d . Dimana t adalah kata kunci dan d adalah dokumen. Kemudian $\max\{t' \in d\}$ adalah Jumlah kata kunci yang sering muncul pada satu dokumen.

Sedangkan *Inverse document frequency (IDF)* adalah suatu statistik yang mengkarakteristikan sebuah *term* dalam keseluruhan koleksi dokumen.

Berikut ini jenis-jenis formula pembobotan *Inverse Document Frequency* :

Tabel 2. 2 Variants of inverse document frequency (IDF) weight

Skema Pembobotan	IDF Weigth
Unary	1
<i>inverse document frequency</i>	$\log \frac{N}{nt} = -\log \frac{nt}{N}$
<i>inverse document frequency smooth</i>	$\log\left(\frac{N}{1+nt}\right)$
<i>inverse document frequency max</i>	$\log\left(\frac{\max\{t' \in d\}nt'}{1+nt}\right)$
<i>Probabilistic inverse document frequency</i>	$\log \frac{N-nt}{nt}$

(Sumber : Poletini, 2004)

Keterangan :

N adalah total jumlah dokumen dalam korpus, nt adalah jumlah kata kunci yang muncul pada banyaknya dokumen, dan $\max\{t' \in d\}nt'$ adalah jumlah kata kunci yang sering muncul pada banyaknya dokumen.

Setelah mendapatkan nilai dokumen dengan rumus *TF* dan *IDF* yang ditentukan langkah selanjutnya untuk mendapatkan hasil akhir dari pembobotan

yaitu dengan cara mengalikan hasil dari nilai *TF* dan *IDF* dokumen tersebut sehingga nantinya akan menghasilkan nilai yang lebih baik.

Berikut ini formula perkalian *TF* dengan *IDF* :

$$TFIDF = TF * IDF \dots\dots\dots (2.1)$$

TF adalah frekuensi dari kemunculan sebuah *term* dalam dokumen yang bersangkutan, Sedangkan *IDF* adalah suatu statistik yang mengkararakteristikan sebuah *term* dalam keseluruhan koleksi dokumen.

2.2.10 Singular Value Decomposition (SVD)

Singular Value Decomposition merupakan sebuah teknik yang biasa digunakan untuk mendekomposisi sebuah matriks dengan 3 komponen yaitu 2 matriks orthogonal serta 1 matriks yang berisi nilai-nilai singular (Arining , 2017).

SVD adalah komponen pemrosesan yang mengkompresi informasi yang berkaitan dalam jumlah besar ke dalam ruang yang lebih kecil. *SVD* merupakan proses awal dari *LSA* yang merepresentasikan isi kata dalam matriks dua dimensi yang berasal dari terjemahan Al-Qur'an tema nerakan, dimana kolom merepresentasikan surah dan ayat, dan baris mewakili *bag of words*. Berikut ini formula dari *SVD* adalah sebagai berikut :

$$A \approx USV^T \dots\dots\dots (2.2)$$

Dimana :

A : matrix asal

U : *orthonormal eigenvector* dari AAT

S : matriks diagonal

V^T : *transpose* dari *orthogonal* matriks V

Dekomposisi ini memungkinkan dimensi matriks asal untuk dilakukan reduksi dimensi. Dengan proses reduksi dimensi terhadap perkalian matriks *SVD*, maka akan diperoleh penyederhanaan dan pembobotan dari matriks asal dengan mengambil sebagian besar dari struktur penting antara kata kunci dengan kalimatnya.

2.2.11 Evaluasi

Evaluasi adalah suatu proses yang sistematis untuk menentukan atau membuat keputusan, sampai sejauh mana tujuan program telah tercapai (Norman E. Grounoud, 1985).

Evaluasi merupakan kegiatan atau upaya yang meliputi pengukuran dan penilaian yang direncanakan untuk mendukung tercapainya tujuan (program, produksi, prosedur). Untuk selanjutnya hasil dari kegiatan atau upaya tersebut digunakan sebagai bahan pengambilan keputusan atas objek yang dievaluasi.

Pada penelitian ini proses evaluasi yang akan digunakan yaitu metode *Precision*, *Recall*, dan *F1-Score*. Tujuannya untuk mengetahui seberapa besar tingkat keakuratan pengelompokan yang dihasilkan dari metode *Latent Semantic Analysis*. *Precision* merupakan rasio prediksi benar *positif* dibandingkan dengan keseluruhan hasil yang diprediksi *positif*, sedangkan *recall* merupakan rasio prediksi benar *positif* dibandingkan dengan keseluruhan data yang benar *positif*, dan *F1-Score* merupakan perbandingan rata-rata *precision* dan *recall* yang dibobotkan.

Berikut formula untuk perhitungan *precision*, *recall*, dan *f1-score*:

$$\text{Precision} = \frac{TP}{TP+FP} \dots\dots\dots (2.3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \dots\dots\dots (2.4)$$

$$\text{F1-Score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \dots\dots\dots (2.5)$$

Keterangan :

True Positive (TP) = jumlah data *positif* yang terdeteksi benar.

False Positive (FP) = jumlah data *negatif* namun terdeteksi sebagai data *positif*




False Negative (FN) = jumlah data *positif* namun terdeteksi sebagai data *negatif*.

2.2.12 Flowchart

Flowchart adalah bagan-bagan yang mempunyai arus yang menggambarkan langkah-langkah penyelesaian suatu masalah. *Flowchart* merupakan cara penyajian dari suatu algoritma yang digunakan untuk menunjukkan dimana alur kerja dan proses yang sedang dikerjakan dalam suatu sistem secara keseluruhan (Ladjamudin, 2006:265).

Berikut ini simbol *flowchart* dan fungsinya dapat dilihat pada tabel 2.3 sebagai berikut :

Tabel 2. 3 Simbol dan Fungsi *Flowchart*

No	Simbol	Fungsi
1.		Simbol <i>Start</i> atau <i>End</i> yang berfungsi sebagai awal atau akhir dari sebuah <i>flowchart</i> .
2.		Simbol proses untuk pemerosesan yang terjadi pada sebuah alur kerja.
3.		Simbol yang menyatakan bagian dari program (sub program).

No	Simbol	Fungsi
4.		Simbol persiapan yang digunakan untuk memberi nilai awal suatu besaran.
5.		Simbol <i>input</i> atau <i>output</i> yang berfungsi sebagai masukan atau keluaran proses.
6.		Simbol yang menyatakan penyambung ke simbol lain dalam satu halaman.
7.		Simbol yang menyatakan penyambung ke halaman lainnya.
8.		Simbol yang menyatakan sebagai dokumen pada alur kerja.
9.		Simbol <i>decision</i> (keputusan) yang digunakan untuk penyeleksian kondisi dalam sebuah program.
10.		Simbol yang menyatakan sebagai media penyimpanan drum magnetic.
11.		Simbol yang menyatakan input atau output menggunakan disket.
12.		Simbol yang menyatakan operasi yang dilakukan secara manual.
13.		Simbol yang menyatakan arah aliran pekerjaan (proses).

(Sumber: Siallagan, 2009:6)

BAB III

METODOLOGI PENELITIAN

3.1 Alat dan Bahan Penelitian

Alat dan bahan yang digunakan dalam penelitian ini adalah sebagai berikut:

3.1.1 Alat Penelitian

Pada penelitian ini penulis menggunakan alat penelitian berupa perangkat keras dan perangkat lunak, yaitu :

1. Perangkat keras yang terdiri dari :
 - a. HP Pavilion g4 Notebook PC
 - b. 3GB RAM
 - c. Intel Core i3-2310M
2. Perangkat lunak yang terdiri dari :
 - a. Microsoft Windows 7 Professional 64-bit
 - b. Python 3.7 (64-bit)
 - c. Jupyter notebook
 - d. Drawing tool yaitu Edraw Max 9.1

3.1.2 Bahan Penelitian

Bahan penelitian yang digunakan oleh penulis untuk mencakup hasil yang diinginkan yaitu berupa Al-Qur'an Cordoba rujukan terlengkap atau disebut juga dengan Al-Qur'an amazing yang bertema neraka.

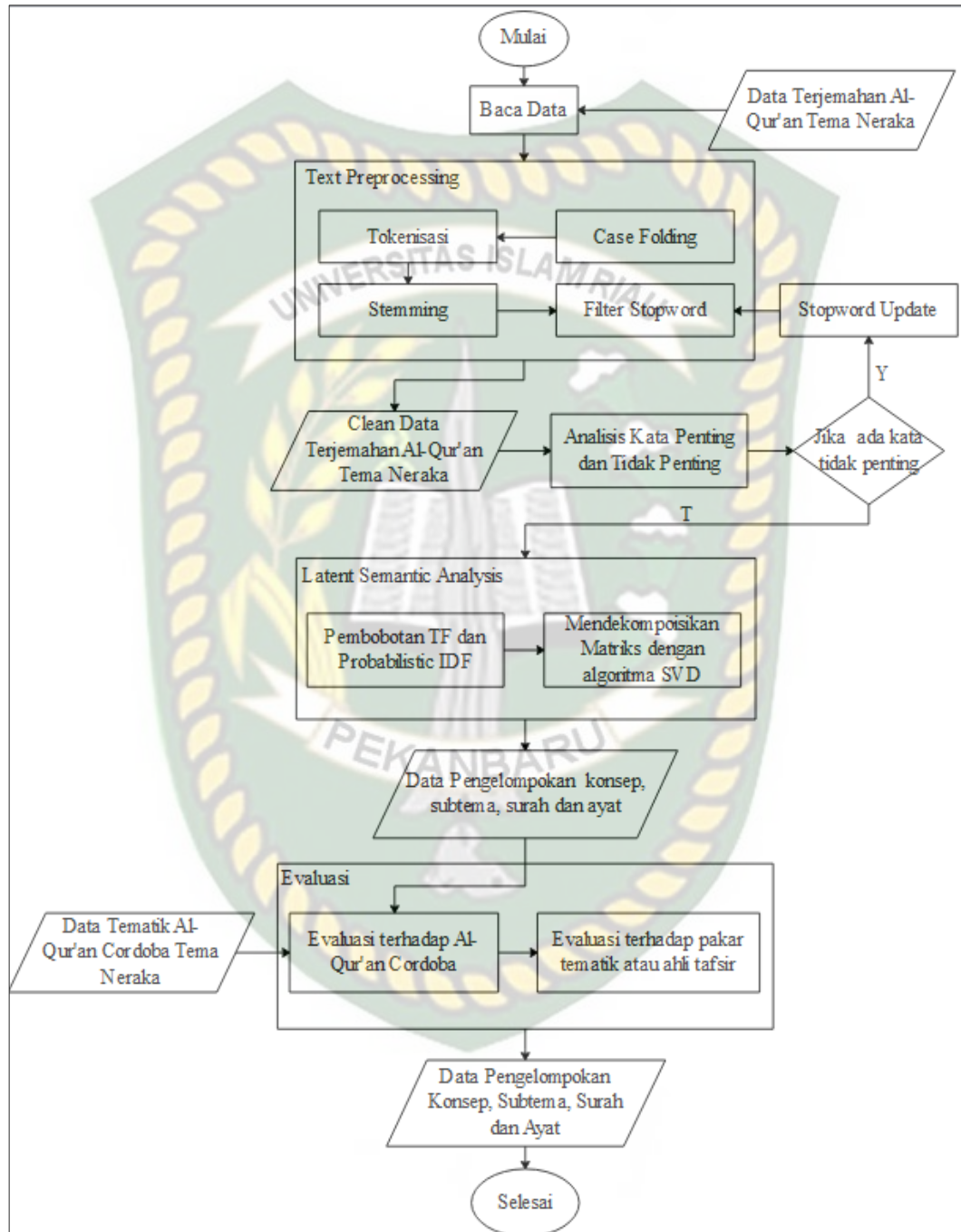
3.2 Analisis Sistem

Penelitian ini menggunakan korpus atau sekumpulan dokumen teks terjemah ayat Al-Qur'an versi Indonesia yang diambil dari qurandatabase.org dan merujuk pada Al-Qur'an Cordoba yang bertema neraka. Dalam penelitian ini dokumen yang digunakan yaitu sebanyak 372 surah dan ayat, dimana setiap surah dan ayat pada Al-Qur'an tersebut dianggap sebagai satu dokumen. Dokumen yang digunakan adalah khusus terjemahan yang bertema neraka saja sesuai dengan batasan masalah penulis.

Sebelum melakukan pengelompokan tematik Al-Qur'an proses yang pertama kali dilakukan yaitu *text preprocessing* untuk mendapatkan *bag of words* yang kemudian akan dihitung menggunakan pembobotan *TF* dan *Probabilistic IDF*. Setelah itu hasil dari pembobotan tersebut akan menjadi data inputan untuk proses selanjutnya dengan menggunakan metode *latent semantic analysis*. Pada metode ini hasil dari pembobotan tersebut akan diolah kembali dengan menggunakan algoritma *singular value decomposition* sehingga menghasilkan pengelompokan surah dan ayat sesuai dengan subtema berdasarkan nilai matriks *U* yang didapat dari proses *SVD* tersebut.

Setelah itu hasil dari pengelompokan tersebut dievaluasi dengan menggunakan perhitungan *precision*, *recall*, dan *f1-score* untuk melihat akurasi kecocokan dengan Al-Qur'an Cordoba. Selain itu hasil dari evaluasi tersebut nantinya akan melibatkan ahli tafsir untuk mengevaluasi kembali data yang tidak valid dengan sistem, sehingga dapat menambahkan kontribusi yang lebih baik terhadap pengelompokan tematik Al-Qur'an pada sub-sub tema neraka.

3.2.1 Gambaran Umum Sistem



Gambar 3. 1 Diagram Alir Sistem

Berikut penjelasan dari diagram alir sistem yang akan dibangun pada Gambar 3.1 :

1. Baca data terjemahan Al-Qur'an tema neraka yang menjadi data inputan terlebih dahulu.
2. Setelah itu melakukan tahap *preprocessing text* yaitu *case folding* untuk merubah semua karakter huruf menjadi huruf kecil, *tokenization* untuk merubah kalimat menjadi tiap-tiap kata, *stemming* untuk merubah semua kata menjadi kata dasar, dan *filtering* untuk menghapus semua kata yang tidak penting dengan menggunakan library *stopwords* indonesian.
3. Setelah tahap *preprocessing* maka akan dilakukan analisa terhadap clean data atau *bag of words*, apabila masih ada kata yang kemungkinan besar tidak akan memberikan pengaruh prediksi terhadap tema neraka maka akan dimasukkan kedalam *stopwords update*.
4. Setelah itu tahapan proses dengan menggunakan metode *latent semantic analysis* terhadap *bag of words* .
5. Pada proses metode *latent semantic analysis* akan dilakukan pembobotan terhadap *bag of words* dengan menggunakan pembobotan *TF* dan *Probabilistic-IDF* sehingga menghasilkan matriks yang berukuran besar.
6. Setelah itu mendekomposisikan matriks yang berukuran besar tersebut dengan menggunakan algoritma *singular value decomposition (SVD)* sekaligus mengelompokkan surah dan ayat menjadi perkonsep berdasarkan hasil dari pembobotan *bag of words* yang telah diproses menggunakan *SVD*.

7. Terakhir yaitu tahap evaluasi digunakan untuk mengamati hasil pengelompokan dari pengujian sistem dengan data tematik Al-Qur'an Cordoba atau disebut juga dengan Al-Qur'an amazing yang bertema neraka dengan melihat akurasi kecocokan menggunakan perhitungan *precision*, *recall*, dan *f1-score* sehingga menghasilkan data pengelompokan tematik Al-Qur'an berdasarkan konsep, subtema, surah dan ayat untuk tema neraka dengan mengambil persentase *f1-score* yang terbaik. Selain itu data yang tidak tepat dari hasil evaluasi sistem akan dievaluasi kembali oleh pakar tematik untuk memperoleh tambahan atau kontribusi yang lebih baik terhadap pengelompokan tematik Al-Qur'an.

3.3 Rancangan Sistem

Rancangan sistem merupakan tahap selanjutnya setelah analisis sistem, mendapatkan gambaran dengan jelas tentang apa yang dikerjakan pada analisis sistem, maka dilanjutkan dengan memikirkan bagaimana membentuk sistem tersebut.

Perancangan sistem adalah suatu fase dimana diperlukan suatu keahlian perancangan untuk elemen-elemen komputer yang akan menggunakan sistem yaitu pemilihan peralatan dan program komputer untuk sistem yang baru. (Kristanto, 2008 : 61)

3.3.1 Text Preprocessing

Text preprocessing merupakan tahapan dari proses awal terhadap teks untuk mempersiapkan teks menjadi data yang akan diolah lebih lanjut. Dalam

penelitian ini penulis menggunakan 4 tahapan *text preprocessing* yang akan digunakan yaitu *case folding*, *tokenization*, *stemming*, dan *filtering*.

Berikut ini tahapan *text preprocessing* dapat dilihat pada gambar 3.2 :



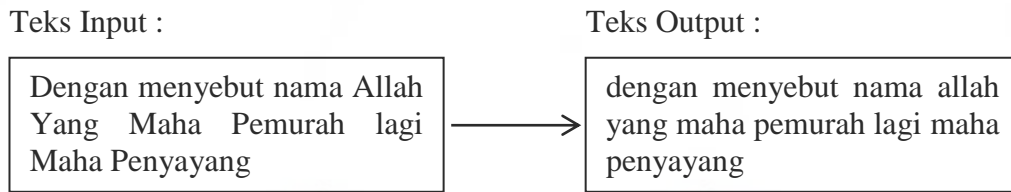
Gambar 3. 2 *Flowchart Text Preprocessing*

Berikut ini penjelasan dari *flowchart text preprocessing* di atas :

3.3.1.1 *Case folding*

Case folding adalah mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter.

Berikut ini contoh dari proses *case folding* :

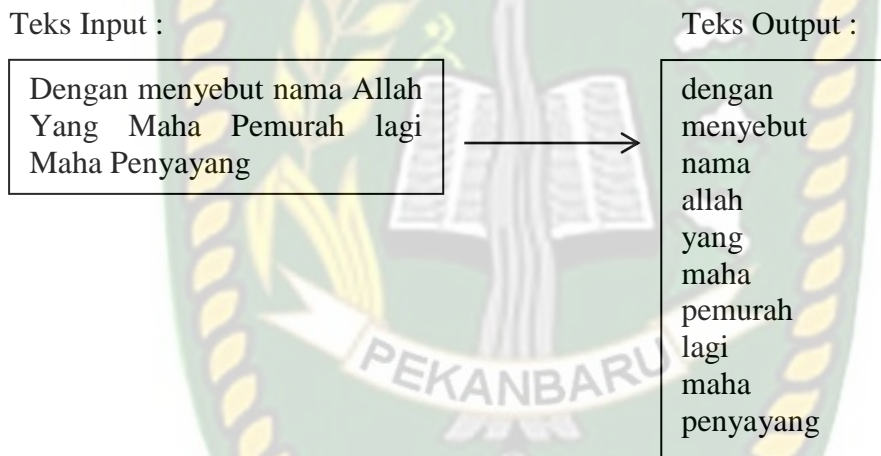


Gambar 3. 3 Proses *Case folding*

3.3.1.2 *Tokenization*

Tahap *tokenization* adalah tahap pemotongan string input berdasarkan pada tiap kata yang menyusunnya.

Berikut ini contoh dari proses *tokenization* :

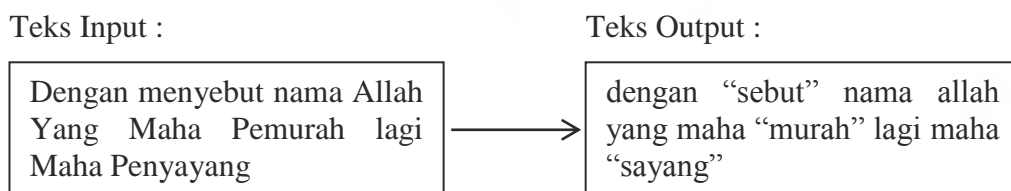


Gambar 3. 4 Proses *Tokenization*

3.3.1.3 *Stemming*

Tahap *stemming* merupakan proses untuk mengubah semua kata yang telah dipilih pada proses *tokenization* menjadi kata yang berupa katan dasar.

Berikut ini contoh dari proses *stemming* :

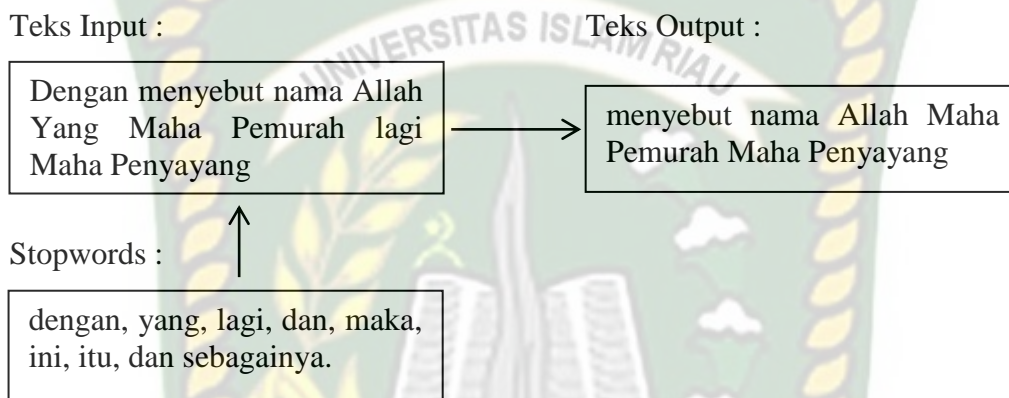


Gambar 3. 5 Proses *Stemming*

3.3.1.4 Filtering

Tahap *filtering* adalah tahap mengambil kata - kata penting dari hasil *tokenization* dan membuang kata yang kurang penting dengan menggunakan *stopwords*.

Berikut ini contoh dari proses *filtering* menggunakan *stopwords* :



Gambar 3. 6 Proses *filtering* menggunakan *stopwords*

Stopwords merupakan kumpulan daftar kata-kata yang kemungkinan besar tidak akan memberikan pengaruh prediksi, seperti imbuhan dan kata ganti. Kata yang termasuk *stopwords* akan dibuang, karena kata tersebut tidak merepresentasikan isi dokumen walaupun sering muncul.

3.3.2 Clean Data

Clean data merupakan data terjemahan Al-Qur'an tema neraka yang sudah diproses melalui tahap *text preprocessing* dan selanjutnya akan dilakukan pembobotan terhadap *term* yang dihasilkan, kemudian akan diproses lagi dengan menggunakan metode *latent semantic analysis* dan algoritma *singular value decomposition* dengan menggunakan *bag of words* tema neraka yang dihasilkan dari tahapan *text preprocessing*.

3.3.3 *Bag of words*

Bag of words adalah sebuah gambaran sederhana yang digunakan dalam pengolahan bahasa alami dan pencarian informasi. Dikenal sebagai model ruang vektor. *Bag of words* juga merupakan tas atau tempat penyimpanan kata-kata yang nantinya dipakai untuk pembobotan terhadap metode *latent semantic analysis*. Kata-kata yang digunakan untuk *bag of words* ini telah dianalisis dan kemudian diambil melalui bahan terjemahan ayat Al-Qur'an Cordoba rujukan terlengkap atau disebut juga dengan Al-Qur'an amazing sesuai dengan batasan masalah penulis yaitu "Tema Neraka".

Berikut ini contoh dari *bag of words* dari tema neraka :

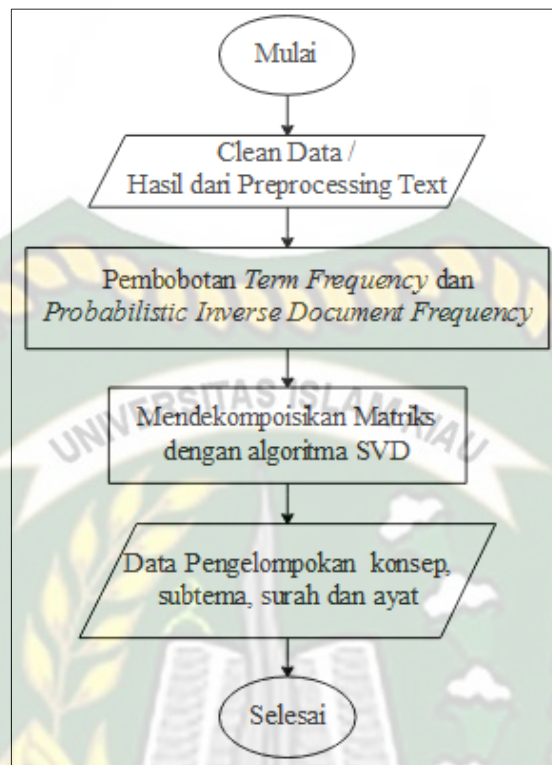
Bag of words = [neraka, hawiyah, jahim, saqar, lazza, huthamah, sair, wail, jahannam, kafir, azab, iblis, api, air, mendidih, sesat, zabaniyah]

Gambar 3. 7 *Bag Of Words Tema Neraka*

3.3.4 *Latent Semantic Analysis (LSA)*

Latent semantic analysis merupakan suatu metode untuk menemukan keterkaitan, kemiripan dan hubungan antara dokumen-dokumen, penggalan dari dokumen, dan kata-kata yang muncul pada dokumen-dokumen dengan memanfaatkan komputasi statistik untuk menggali dan merepresentasikan konteks yang digunakan sebagai sebuah arti kata untuk sejumlah korpus yang besar.

Berikut alur dari tahapan *latent semantic analysis* dapat dilihat pada gambar 3.8 dibawah ini :



Gambar 3. 8 Flowchart *Latent semantic analysis*

Pada gambar 3.8 dapat dilihat bahwa metode ini digunakan setelah data terjemahan ayat Al-Qur'an melakukan tahapan *text preprocessing*, kemudian data tersebut akan diproses dengan melakukan pembobotan menggunakan algoritma *term frequency (TF)* dan *Probabilistic-inverse document frequency (IDF)* terhadap *bag of words* yang dihasilkan. Setelah itu dilakukan proses *Singular Value Decomposition (SVD)* yang digunakan untuk mengkompresi informasi yang berkaitan dalam jumlah besar kedalam ruang yang lebih kecil sekaligus menghasilkan pengelompokan subtema, surah dan ayat menjadi perkonsep.

3.3.5 Pembobotan *TF* dan *Probabilistic-IDF*

Data yang telah melalui tahap *preprocessing* harus berbentuk numerik. Untuk mengubah data tersebut menjadi numerik yaitu menggunakan metode pembobotan *TF* dan *Probabilistic-IDF*. Metode *TF* dan *Probabilistic-IDF* merupakan

metode pembobotan *term* yang banyak digunakan sebagai metode pembandingan terhadap metode pembobotan baru. Pada metode ini, hasil akhir perhitungan bobot *term* dalam sebuah dokumen dilakukan dengan mengalikan nilai *Term Frequency* dengan *Probabilistic Inverse Document Frequency*.

Dalam perhitungan bobot menggunakan *TF* dan *IDF*, langkah yang pertama kali dihitung yaitu nilai *TF* perkata dengan bobot masing-masing kata menggunakan formula berikut :

$$ft, d / \sum_{t' \in d} ft, d \dots\dots\dots (3.1)$$

Keterangan :

ft, d = banyaknya kata kunci yang muncul pada tiap-tiap dokumen

$\sum_{t' \in d} ft, d$ = banyaknya kata pata tiap-tiap dokumen.

Kemudian menghitung nilai *Probabilistic-IDF* dengan menggunakan formula sebagai berikut :

$$\log \frac{N-nt}{nt} \dots\dots\dots (3.2)$$

Keterangan :

N = jumlah keseluruhan data yang terdapat pada dokumen.

nt = jumlah kemunculan kata kunci pada banyaknya dokumen.

Berikut ini alur perhitungan *Term Frequency (TF)* dan *Probalistic Inverse Document Frequency (IDF)* dapat dilihat pada tabel berikut :

Tabel 3. 1 Dokumen Tema Neraka

Document	Isi Document	Jumlah Kata
d1	kemudian sesungguhnya tempat kembali mereka benar-benar ke neraka jahim.	2 kata
d2	“Apakah yang memasukkan kamu ke dalam Saqar (neraka)?”	2 kata
d3	sesungguhnya neraka jahannam itu (padanya) ada tempat pengintai,	2 kata
d4	“dan diperlihatkan dengan jelas neraka jahim kepada orang-orang yang sesat.”	3 kata

Pada tabel 3.1 merupakan data terjemahan ayat Al-Qur'an tema neraka yang terdiri dari 4 dokumen dan masing – masing dokumen terdiri dari 2 buah kata kunci kecuali dokumen pada d4 yang memiliki 3 buah kata kunci.

Berikut ini tabel kata kunci atau *Bag of words* untuk tema neraka :

Tabel 3. 2 *Bag of Words* Tema Neraka

No	Bag Of Words
1	Neraka
2	Jahannam
3	Jahim
4	Kafir

Pada gambar 3.2 di atas merupakan kata kunci yang sesuai dengan tema neraka untuk dijadikan sebagai *bag of words* yang nantinya akan melakukan pembobotan *TFIDF*.

Berikut ini tabel proses perhitungan *Term Frequency (TF)*:

Tabel 3. 3 Proses Perhitungan *TF*

Doc	<i>Bag Of Words</i>				
	Neraka	Jahim	Jahannam	Saqar	Sesat
d1	0.5	0.5	0	0	0
d2	0.5	0	0	0.5	0
d3	0.5	0	0.5	0	0
d4	0.333	0.333	0	0	0.333

Pada tabel 3.3 merupakan hasil proses perhitungan *Term Frequency (TF)* dengan cara membagikan jumlah kata kunci (t) dengan jumlah kata yang ada pada dokumen (d). Contohnya pada d1 kata “neraka” muncul 1 kali kemudian jumlah kata kunci pada d1 ada 2 kata maka perhitungan dilakukan dengan formula (3.1), maka hasilnya adalah $1 / 2 = 0.5$.

Berikut ini tabel proses perhitungan *Probabilistic-IDF* :

Tabel 3. 4 Proses Perhitungan *Probabilistic-IDF*

<i>Bag Of Words</i>				
Neraka	Jahim	Jahannam	Saqar	Sesat
0.4771	0.4771	0.4771	0.4771	0.4771

Pada tabel 4.3 merupakan hasil proses perhitungan *Probabilistic-IDF (IDF)* dengan formula (3.2), dimana N adalah jumlah keseluruhan data yang terdapat pada dokumen dan nt adalah jumlah kemunculan kata kunci pada banyaknya dokumen. Contohnya pada kata “neraka” muncul sebanyak 4 kali pada keseluruhan dokumen. Maka perhitungan dilakukan dengan formula (3.2), maka hasilnya adalah $\log (4-4 / 4) = 0.4771$

Langkah terakhir yaitu mengalikan masing – masing hasil dari perhitungan *TF* dengan hasil perhitungan *Probabilistic-IDF* untuk mendapatkan hasil akhir yang lebih baik.

Berikut ini proses perhitungan *TF* dan *Probabilistic-IDF* :

Tabel 3. 5 Proses Perhitungan *TF* dan *Probabilistic-IDF*

Doc	<i>Bag Of Words</i>				
	Neraka	Jahim	Jahannam	Saqar	Sesat
d1	0.2385	0.2385	0	0	0
d2	0.2385	0	0	0.2385	0
d3	0.2385	0	0.2385	0	0
d4	0.1588	0.1588	0	0	0.1588

Pada gambar 3.5 merupakan proses terakhir yaitu mencari nilai *TFIDF* dengan cara mengalikan hasil dari *TF* dengan hasil dari *Probabilistic-IDF*. Contohnya pada d1 kata “neraka”. Nilai *TF* dari kata “neraka” adalah 0.5 sedangkan nilai *Probabilistic-IDF* dari kata “neraka” adalah 0.4771 Maka perhitungan dilakukan dengan cara $0.5 * 0.4771 = 0.2385$.

3.3.6 Singular Value Decomposition (SVD)

Metode *singular value decomposition* (SVD) disini sangat berguna untuk mempermudah pengolahan data serta menyederhanakan matriks berukuran apa saja yang didapat setelah proses menghitung pembobotan *TF* dan *Probabilistic-IDF*. SVD juga memiliki kemampuan pemodelan hubungan antara kata - kata dan kalimat yang dapat membantu untuk meningkatkan akurasi dan menghilangkan fitur-fitur yang dirasa kurang penting dalam dokumen sehingga dapat memproduksi model data yang lebih baik. Selain itu SVD juga memiliki kelebihan dalam efisiensi waktu proses untuk dataset berskala besar.

Pada tahap ini hasil dari perhitungan *TFIDF* akan diproses secara otomatis menggunakan *library* python yang ada pada sistem, sehingga tahap ini secara otomatis menghasilkan sebuah pengelompokan yang sesuai dengan jumlah

subtema yang terbentuk. Selain itu tahap ini juga akan menghasilkan surah dan ayat yang secara otomatis dikelompokkan oleh sistem berdasarkan hasil dari proses *SVD* nilai matriks *U*.

Berikut ini contoh alur pengelompokan berdasarkan hasil dari proses *SVD* nilai matriks *U* dapat dilihat pada tabel berikut :

Tabel 3. 6 Contoh Data Awal *SVD* (*U*)

Bagofwords	d1	d2	d3	d4
Neraka				
Jahim				
Jahannam				
Saqar				
Sesat				

Pada tabel 3.6 merupakan contoh data awal untuk proses pengelompokan yang dilakukan oleh sistem. Pada tabel tersebut dapat dilihat bahwa kolom yang ditandai dengan warna abu-abu merupakan kemunculan dari *Bagofwords*. Langkah selanjutnya yaitu proses pengelompokan dokumen berdasarkan kemunculan *Bagofwords*.

Berikut ini contoh pengelompokan dokumen berdasarkan kemunculan *Bagofwords* dapat dilihat pada tabel 3.7 :

Tabel 3. 7 Contoh Pengelompokan *SVD* (*U*)

Bagofwords	d1	d4	d2	d3
Neraka				
Jahim				
Jahannam				
Saqar				
Sesat				

Pada tabel 3.7 menunjukkan proses terjadinya pengelompokan dokumen berdasarkan kemunculan *Bagofwords*, dimana kemunculan *Bagofwords* pada dokumen-dokumen tersebut akan terkelompok secara otomatis. Langkah terakhir yaitu pembentukan konsep sesuai dengan dokumen yang sudah terkelompok.

Berikut ini contoh pembentukan konsep dari hasil pengelompokan dokumen dapat dilihat pada tabel 3.8 :

Tabel 3. 8 Contoh Pembentukan Konsep

Bagofwords	d1	d4	d2	d3	
Neraka					Konsep-1
Jahim					Konsep-2
Jahannam					Konsep-3
Saqar					Konsep-4
Sesat					Konsep-5

Pada tabel 3.8 menunjukkan terdapat 5 konsep yang terbentuk dari hasil pengelompokan dokumen berdasarkan kemunculan *Bagofwords*.

3.3.7 Evaluasi

Dalam penerapan metode *latent semantic analysis* sebuah kata perlu diberi bobot dengan menggunakan perhitungan *F1-Score* berdasarkan pada nilai *Precision* dan *Recall*, karena semakin sering suatu kata muncul pada suatu dokumen, maka diduga kata tersebut semakin penting dalam dokumen. Sehingga dokumen tersebut dapat dikelompokkan sesuai dengan konsep dan tema yang telah ditentukan.

Pada tahap ini evaluasi dilakukan untuk melihat akurasi kecocokan terhadap pengelompokan masing-masing konsep, surah dan ayat yang didapat dari hasil algoritma *singular value decomposition* dengan data tematik Al-Qur'an

Cordoba tema neraka. Dari hasil evaluasi tersebut akan didapat hasil data yang tepat dan tidak tepat. Untuk hasil data yang tidak tepat akan di evaluasi kembali oleh pakar tematik Al-Qur'an sehingga nantinya dapat menambah kontribusi terhadap pengelompokan berdasarkan subtema, surah dan ayat pada Al-Qur'an tema neraka.



Dokumen ini adalah Arsip Miik :

Perpustakaan Universitas Islam Riau

BAB IV HASIL DAN PEMBAHASAN

Berikut ini merupakan tampilan hasil dan pembahasan dari pengelompokan tematik Al-Qur'an menggunakan metode *latent semantic analysis* dengan pembobotan *term frequency* dan *Probabilistic inverse document frequency*. Tampilan ini dibuat sedemikian rupa untuk mendukung segala sesuatu yang dibutuhkan agar terciptanya sebuah sistem yang valid.

4.1 Data Tematik Al-Qur'an Tema Neraka

Data tematik Al-Qur'an tema neraka ini merupakan data inputan awal yang bersumber dari qurandatabase.org yang merujuk pada Al-Quran Cordoba khusus tema neraka. Data yang di ambil untuk menjadi inputan awal adalah surah, ayat, dan terjemahan tema neraka sebanyak 372 data kemudian di proses menggunakan bahasa pemrograman python.

Berikut tampilan data tematik Al-Qur'an tema neraka dapat dilihat pada tabel 4.1 dibawah ini :

Tabel 4. 1 Data tematik Al-Qur'an tema neraka

No	Surah	Ayat	Terjemahan
0	2	119	Sesungguhnya Kami telah mengutusmu (Muhammad) ...
1	2	206	Dan apabila dikatakan kepadanya: "Bertakwalah ...
2	3	112	Mereka diliputi kehinaan di mana saja mereka b...
3	3	162	Apakah orang yang mengikuti keridhaan Allah sa...
4	3	197	Itu hanyalah kesenangan sementara, kemudian te...
...

No	Surah	Ayat	Terjemahan
367	38	63	Apakah kami dahulu menjadikan mereka olok-olok...
368	40	47	Dan (ingatlah), ketika mereka berbantah-bantah...
369	40	48	Orang-orang yang menyombongkan diri menjawab: ...
370	40	50	Penjaga Jahannam berkata: "Dan apakah belum da...
371	67	9	Mereka menjawab: "Benar ada", sesungguhnya tel...

Pada tabel 4.1 merupakan data terjemahan tema neraka sebanyak 372 surah dan ayat yang menjadi data inputan awal untuk proses pengelompokan tematik Al-Qur'an.

4.2 Text Preprocessing

Proses ini merupakan tahap awal untuk mempersiapkan teks menjadi data yang akan diolah lebih lanjut. Pada proses ini menggunakan 4 tahapan *text preprocessing* yang akan di gunakan yaitu: *case folding* untuk merubah semua karakter huruf menjadi huruf kecil, *tokenization* untuk merubah kalimat menjadi tiap-tiap kata, *stemming* untuk merubah semua kata menjadi kata dasar, dan *filtering* untuk menghapus semua kata yang tidak penting dengan menggunakan *library stopwords* indonesian.

Berikut hasil dari tahapan *text preprocessing* dapat dilihat pada tabel 4.2 dibawah ini :

Tabel 4. 2 Hasil *text preprocessing*

Surah	Ayat	Terjemahan	Lower_Tokenizing	Stemming	Stopwords	Bagofwords	
0	2	119	Sesungguhnya Kami telah mengutusmu (Muhammad) ...	[sesungguhnya, kami, telah, mengutusmu, (, muh...	[sungguh, kami, telah, utus, , muhammad, , den...	[, , , , , , huni, neraka,]	huni neraka
1	2	206	Dan apabila dikatakan kepadanya: "Bertakwalah ...	[dan, apabila, dikatakan, kepadanya, ; , " , be...	[dan, apabila, kata, kepada, , , takwa, kepada...	[, , , dosa, , , neraka, jahannam, , nerak...	dosa neraka jahannam neraka jahannam b...
2	3	112	Mereka diliputi kehinaan di mana saja mereka b...	[mereka, diliputi, kehinaan, di, mana, saja, m...	[mereka, liput, hina, di, mana, saja, mereka, ...	[hina, , , , , , kafir, , durhaka,]	hina kafir durhaka
3	3	162	Apakah orang yang mengikuti keridhaan Allah sa...	[apakah, orang, yang, mengikuti, keridhaan, al...	[apakah, orang, yang, ikut, keridhaan, allah, ...	[, jahannam, , buruk,]	jahannam buruk
4	3	197	Itu hanyalah kesenangan sementara, kemudian te...	[itu, hanyalah, kesenangan, sementara, , kemu...	[itu, hanya, senang, sementara, , kemudian, te...	[, jahannam, , jahannam, , buruk,]	jahannam jahannam buruk
...
367	38	63	Apakah kami dahulu menjadikan mereka olok-olok...	[apakah, kami, dahulu, menjadikan, mereka, olo...	[apakah, kami, dahulu, jadi, mereka, olok, , a...	[olok, ,]	olok
368	40	47	Dan (ingatlah), ketika mereka berbantah-bantah...	[dan, (, ingatlah,) , , ketika, mereka, berba...	[dan, , ingat, , , ketika, mereka, berbantah-b...	[, , berbantah-bantah, neraka, , , , azab, ...	berbantah-bantah neraka azab api neraka
369	40	48	Orang-orang yang menyombongkan diri menjawab: ...	[orang-orang, yang, menyombongkan, diri, menja...	[orang, yang, sombong, diri, jawab, , , sunggu...	[, , neraka, hamba-hamba, , ,]	neraka hamba-hamba-
370	40	50	Penjaga Jahannam berkata: "Dan apakah belum da...	[penjaga, jahannam, berkata, ; , " , dan, apaka...	[jaga, jahannam, kata, , , dan, apakah, belum...	[jahannam, , , , , , jahannam, , , ,]	jahannam jahannam kafir
371	67	9	Mereka menjawab: "Benar ada", sesungguhnya tel...	[mereka, menjawab, ; , " , benar, ada, , , s...	[mereka, jawab, , , benar, ada, , , sungguh, t...	[, , , , , dusta, , , , , sesat, ,]	dusta sesat

Pada tabel 4.2 merupakan tahap *text preprocessing* yang dilakukan menggunakan data terjemahan Al-Qur'an tema neraka yang berjumlah 372 data terjemahan. Proses yang pertama kali dilakukan yaitu *case folding*, fungsinya untuk merubah semua kalimat data terjemahan menjadi huruf kecil. Setelah itu proses *tokenizing* yaitu untuk memisahkan kalimat menjadi tiap-tiap kata. Selanjutnya proses *stemming* merupakan proses merubah semua kata menjadi kata dasar. Kemudian yang terakhir yaitu proses *filtering stopwords* menggunakan *library stopwords* indonesian, fungsinya untuk menghapus kata-kata yang kemungkinan besar tidak akan memberikan pengaruh prediksi, seperti imbuhan dan kata ganti karena kata tersebut tidak merepresentasikan isi dokumen walaupun sering muncul. Selain itu penulis juga menambahkan *stopwords update* sesuai dengan kebutuhan dan analisis penelitian terhadap pengelompokan tematik Al-Qur'an.

Berikut ini penambahan *stopwords update* sesuai dengan kebutuhan dapat dilihat pada tabel 4.3 :

Tabel 4. 3 Daftar *Stopword Update*

Stopwords Update
(['air','arsy','baiknya','rusa','kaki','arab','serah','pedoman','istana','segenap','kasih','dahi','daki','diminumnya','dinampakkan','dinding','dur','emas','gelegak','gemuk','gurau','halal','hancur','situ','gila','susu','berkali','kali','kurangnya','mata','olah','sekurang','setidak','tama','bani','nafkah','ahli','alam','akrab','alih','amal','angin','arah','asai','asap','awas','badwi','bahaya','bala','bantu','batang','bathil','bau','beban','belah','belas','berita','bit','bunga','cabang','campur','senda','dalamnya','sungguh','allah','masuk','tuhan','balas','hapus','tidaknya','cari','puluh','ikrar','tenggelam','ayat','hadap','balas','hasta','haus','hias','hidang','hijrah','hubung','hujan','iring','isteri','buah','iman','sedia','kerja','baca','laku','hak','dengar','muka','hadap','jaga','lipat','ringan','rahasia','rasul','hidup','ya','selimut','barangsiapa','tinggal','jalan','lihat','diam','manusia','firman','tolong','niscaya','dunia','kecuali','laki','perempuan','sembah','seret','nyala','barang','rugi','mohon','bawa','mudah','temu','intai','mahhluk','risalah','liput','rombong','penuh','kitab','ganda','lempar','buka','lindung','golong','turun','bangkit','tiada','hai','senang','jatuh','sempit','kalah','larang','suka','lawan','sengaja','harap','kulit','usaha','kelak','rezeki','negeri','tatkala','suara','jungkir','sisi','jelek','timpal','mengada','adakan','janji','dasar','prasnangka','lupa','kaum','kutu','keri','gelagak','maksud','keri','pekak','saleh','ganti','perkasa','bisu','padam','karunia','saksi','berangkat','rasulullah','terik','fitnah','bangun','ikan','nyata','habis','gejolak','rapat','asa','shiddiqien','anggap','heran','terima','surga','benda','takut','wahyu','neraka','belaka','jurang','runtuh','tanda','tebus','tepi','agama','manfaat','binatang','patut','hikmah','hilang','luas','mekah','tindas','berat','lekat','keliling','telinga','tangguh','andaikata','kepong','istirahat','lapis','ragu','desak','pasu','gabung','mundur','belok','lampau','geram','perhati','umpan','salam','ampun','besi','rantai','pandang','ilah','adil','ulang','bantah','berbantah','hindar','tingkat','pagi','nama','berfikir','langit','ternak','batas','kabar','tanah','samping','gagal','akibat','ibadat','alir','rintih','tunduk','izin','golak','muhammad','masjidilharam','henti','umpan','beli','sentosa','aman','sungkur','alas','seraya','sakit','taat','sumpah','nama','rang','lesu','sumbat','sanggup','gelagak','keluarga','kiri','bilang','beli','ngeri','cabut','alangkah','ibrahim','perintah','benci','indah','ku','nikmat','orng','dingin','tujuh','pakai','kena','umpama','sabar','teriak','umur','berani','belulang','bencana','berani','bohong','bukti','bunuh','darah','derita','hangus','hapus','ikat','ilmu','israil','jahat','jerumus','jihad','jual','jumpa','kalang','kasar','kaya','kebaji','lontar','luluh','luth','maaf','madinah','madu','mahfuzh','makhluk','sungai','marah','maryam','masuklah','maut','mayang','merta','mimpi','minyak','miskin','muram','murka','murtad','orang','musa','nafas','naung','nilai','nuh','nyala','pahala','paham','paksa','panggil','pasang','payah','penjuru','perak','perut','petang','pikul','pincang','potong','puji','puncak','punggung','putera','rabb','rendah','rusak','sabut','salah','saring','sayang','sejuk','sembilan','sembunyi','semesta','sengsara','serang','sesal','shalat','simpan','siram','sisat','sombong','suami','sukar','sumber','suruh','tali','tangan','tanggung','tarik','tasbih','teguh','tekan','tel','teman','tempuh','perang','tentara','tuang','tukar','tulang','tulis','tunggu','ubah','uji','unta','urus','usus','utus','makmur','alat','berhala','bintang','hukum','keburu','pimpin','tiang','tubi','kuning','lapar','penjara'])

4.3 *Bag of words*

Bag of words merupakan kumpulan kata-kata yang berasal dari tahapan *text preprocessing* sebelumnya dan kemudian akan diproses kembali dengan menggunakan pembobotan kata dengan formula *Term Frequency* dan *Probabilistic-Inverse Document Frequency*.

Berikut hasil *bag of words* yang diperoleh dari tahapan *text preprocessing* dapat dilihat pada tabel 4.4 dibawah ini :

Tabel 4. 4 *Bag of words* tema neraka

Bag of words
'ingkar', 'huni', 'hina', 'didih', 'munafik', 'jahannam', 'durhaka', 'hawiyah', 'azab', 'haram', 'makan', 'kikir', 'neraka', 'khianat', 'pintu', 'zabaniyah', 'pohon', 'cela', 'bakar', 'jahim', 'hisab', 'batil', 'kufur', 'zalim', 'akhirat', 'huthamah', 'kekal', 'api', 'cambuk', 'dengki', 'musyrik', 'panas', 'iblis', 'nanah', 'zaqqum', 'dosa', 'malik', 'buruk', 'aniaya', 'belunggu', 'mati', 'syaitan', 'dusta', 'jin', 'riba', 'kafir', 'olok', 'lalai', 'siksa', 'najis', 'ria', 'jannah', 'quraisy', 'binasa', 'celaka', 'saqar', 'minum', 'sesat'

Pada tabel 4.4 merupakan kumpulan kata kunci atau *Bag of words* yang didapat dari proses *text preprocessing*. Pada penelitian ini terdapat 58 kata yang menjadi *bag of words* untuk tema neraka.

4.4 Pembobotan kata terhadap *Bag of words*

Sebelum melakukan pembobotan terhadap kata yang terdapat pada *Bag of words* langkah yang pertama kali dilakukan yaitu mengeset semua kata menjadi nol, agar dapat menghitung jumlah kemunculan kata terhadap dokumen yang nantinya akan diproses lagi menggunakan formula *TF* dan *Probabilistic-IDF*.

Berikut ini hasil *bag of words* yang sudah diset dengan nilai nol dapat dilihat pada tabel 4.5 :

Tabel 4. 5 Set nilai nol terhadap *Bag of words*

	ingkar	huni	hina	didih	munafik	jahannam	durhaka	hawiyah	azab	haram	...	siksa	najis	ria	jannah	quraisy	binasa	celaka
Bagofwords																		
huni neraka	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
dosa neraka jahannam neraka jahannam buruk	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
hina kafir durhaka	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
jahannam buruk	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
jahannam jahannam buruk	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
...
olok	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
neraka azab api neraka	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
neraka	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
jahannam jahannam kafir	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
dusta sesat	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0

Pada gambar 4.5 menunjukkan bahwa semua *Bag of words* diset menjadi 0. Setelah mengeset nilai *Bag of words* menjadi nol langkah selanjutnya yaitu menghitung jumlah kemunculan *Bag of words* terhadap dokumen.

Berikut ini hasil kemunculan *Bag of words* terhadap dokumen dapat dilihat pada tabel 4.6 dibawah ini :

Tabel 4. 6 Menghitung kemunculan *bag of words*

	huni	neraka	dosa	jahannam	buruk	hina	kafir	durhaka	dengki	api	...	jannah	iblis	zabaniyah	cambuk	kikir	batil	quraisy
Bagofwords																		
huni neraka	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
dosa neraka jahannam neraka jahannam buruk	0.0	2.0	1.0	2.0	1.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
hina kafir durhaka	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
jahannam buruk	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
jahannam jahannam buruk	0.0	0.0	0.0	2.0	1.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...
olok	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
neraka azab api neraka	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
neraka	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
jahannam jahannam kafir	0.0	0.0	0.0	2.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
dusta sesat	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Pada tabel 4.5 menunjukkan setiap kata yang muncul 1 kali pada tiap dokumen maka nilai kata tersebut akan bertambah 1. Dan jika kata tidak ada pada setiap dokumen maka kata tersebut akan tetap bernilai 0.

4.4.1 Term Frequency

Setelah mendapatkan nilai *Bag of words* pada tiap-tiap dokumen maka tahap selanjutnya yaitu menghitung nilai *Term Frequency* dengan persamaan berikut :

$$ft, d / \sum_{t' \in d} ft, d \dots\dots\dots (4.1)$$

Keterangan :

ft, d = banyaknya kata kunci yang muncul pada tiap-tiap dokumen

$\sum_{t' \in d} ft, d$ = banyaknya kata pata tiap-tiap dokumen.

Berikut ini hasil perhitungan *Term Frequency* terhadap *Bag of words* dapat dilihat pada tabel 4.7 :

Tabel 4. 7 Hasil perhitungan *Term Frequency*

	huni	neraka	dosa	jahannam	buruk	hina	kafir	durhaka	dengki	api	...	jannah	iblis	zabaniyah	cambuk	kikir
Bagofwords																
huni neraka	0.5	0.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	...	0.0	0.0	0.0	0.0	0.0
dosa neraka jahannam neraka jahannam buruk	0.0	0.333333	0.166667	0.333333	0.166667	0.000000	0.000000	0.000000	0.0	0.00	...	0.0	0.0	0.0	0.0	0.0
hina kafir durhaka	0.0	0.000000	0.000000	0.000000	0.000000	0.333333	0.333333	0.333333	0.0	0.00	...	0.0	0.0	0.0	0.0	0.0
jahannam buruk	0.0	0.000000	0.000000	0.500000	0.500000	0.000000	0.000000	0.000000	0.0	0.00	...	0.0	0.0	0.0	0.0	0.0
jahannam jahannam buruk	0.0	0.000000	0.000000	0.666667	0.333333	0.000000	0.000000	0.000000	0.0	0.00	...	0.0	0.0	0.0	0.0	0.0
...
olok	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	...	0.0	0.0	0.0	0.0	0.0
neraka azab api neraka	0.0	0.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.25	...	0.0	0.0	0.0	0.0	0.0
neraka	0.0	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	...	0.0	0.0	0.0	0.0	0.0
jahannam jahannam kafir	0.0	0.000000	0.000000	0.666667	0.000000	0.000000	0.333333	0.000000	0.0	0.00	...	0.0	0.0	0.0	0.0	0.0
dusta sesat	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	...	0.0	0.0	0.0	0.0	0.0

Pada tabel 4.7 menunjukkan hasil perhitungan *Term Frequency* terhadap *Bag of words* dengan menggunakan formula pada persamaan (4.1).

4.4.2 *Probabilistic Inverse Document Frequency*

Selanjutnya melakukan proses perhitungan *Probabilistic Inverse Document Frequency* dengan persamaan berikut :

$$\log \frac{N-nt}{nt} \dots\dots\dots (4.2)$$

Keterangan :

N = jumlah keseluruhan data yang terdapat pada dokumen.

nt = jumlah kemunculan kata kunci pada banyaknya dokumen.

Berikut hasil perhitungan *Probabilistic Inverse Document Frequency* terhadap *Bag of words* dapat dilihat pada tabel 4.8 dibawah ini :

Tabel 4. 8 Hasil perhitungan *Probabilistic Inverse Document Frequency*

	ingkar	huni	hina	didih	munafik	jahannam	durhaka	hawiyah	azab	haram	...
0	2.569374	2.569374	2.569374	2.569374	2.569374	2.569374	2.569374	2.569374	2.569374	2.569374	...

Pada tabel 4.8 menunjukkan hasil perhitungan *Probabilistic Inverse Document Frequency* terhadap *Bag of words* dengan menggunakan formula pada persamaan (4.2).

4.4.3 *Term Frequency dan Probabilistic Inverse Document Frequency*

Setelah mendapatkan nilai *Term Frequency* dan *Probabilistic Inverse Document Frequency* terhadap *Bag of words* maka langkah selanjutnya yaitu menghitung nilai akhir dengan cara mengalikan hasil dari *Term Frequency* dengan *Probabilistic Inverse Document Frequency*.

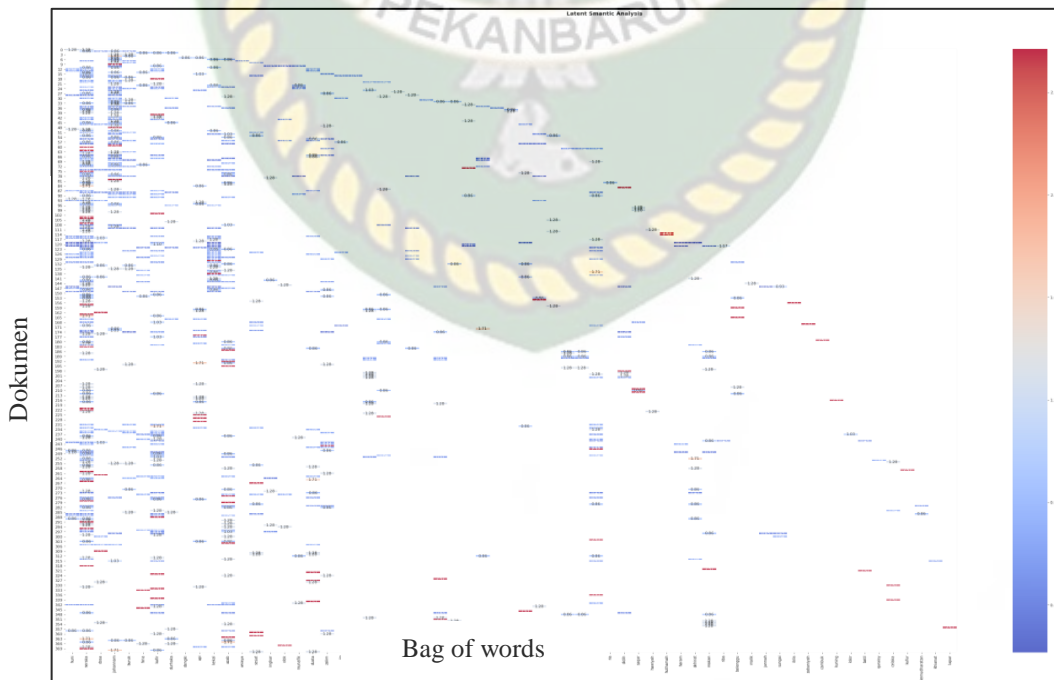
Berikut hasil perkalian *term frequency* dan *Probabilistic Inverse Document Frequency* terhadap *Bag of words* :

Tabel 4. 9 Hasil Perkalian *TF* dan *Probabilisic IDF*

	huni	neraka	dosa	jahannam	buruk	hina	kafir	durhaka	dengki	api	...	jannah	iblis	zabaniyah	cambuk
0	1.284687	1.284687	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	...	0.0	0.0	0.0	0.0
1	0.000000	0.856458	0.428229	0.856458	0.428229	0.000000	0.000000	0.000000	0.0	0.000000	...	0.0	0.0	0.0	0.0
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.856458	0.856458	0.856458	0.0	0.000000	...	0.0	0.0	0.0	0.0
3	0.000000	0.000000	0.000000	1.284687	1.284687	0.000000	0.000000	0.000000	0.0	0.000000	...	0.0	0.0	0.0	0.0
4	0.000000	0.000000	0.000000	1.712916	0.856458	0.000000	0.000000	0.000000	0.0	0.000000	...	0.0	0.0	0.0	0.0
...
367	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	...	0.0	0.0	0.0	0.0
368	0.000000	1.284687	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.642343	...	0.0	0.0	0.0	0.0
369	0.000000	2.569374	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	...	0.0	0.0	0.0	0.0
370	0.000000	0.000000	0.000000	1.712916	0.000000	0.000000	0.856458	0.000000	0.0	0.000000	...	0.0	0.0	0.0	0.0
371	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	...	0.0	0.0	0.0	0.0

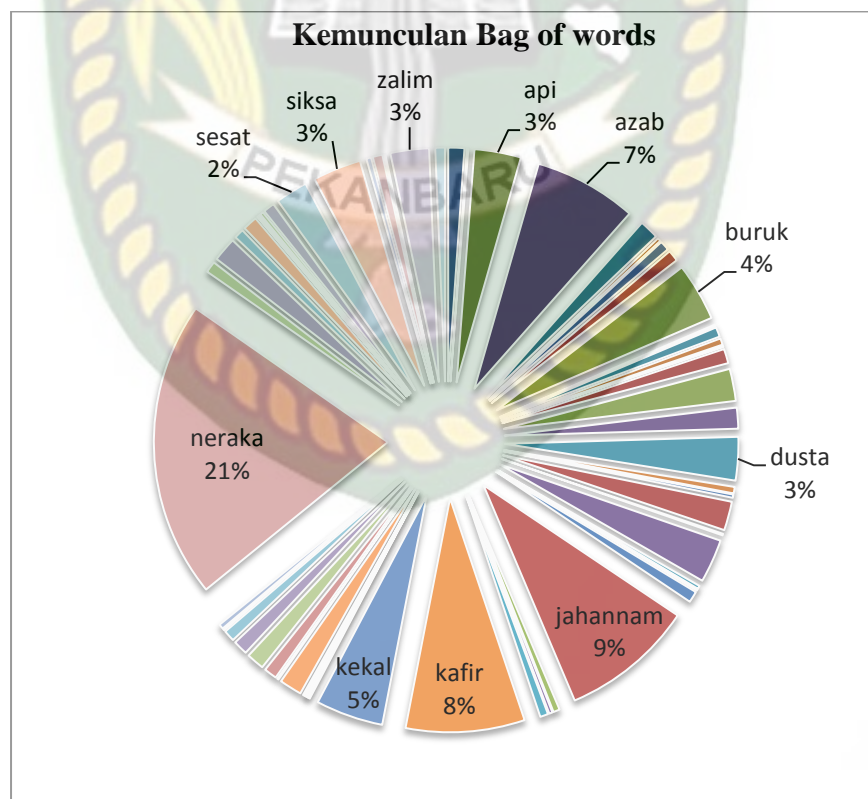
Pada tabel 4.9 menunjukkan hasil akhir dari perhitungan *Term Frequency* dengan *Probabilistic Inverse Document Frequency* terhadap *Bag of words* menggunakan formula $TFIDF = TF * Probabilistic-IDF$.

4.5 Visualisasi Kemunculan *Bag of words* terhadap Dokumen



Gambar 4. 1 Visualisasi Hasil Perkalian *TF* dan *Probabilisic IDF*

Pada gambar 4.1 merupakan visualisasi *Bag of words* yang sudah melalui tahap perkalian *TF* dan *Probabilistic-IDF*. Dari gambar tersebut dapat di lihat bahwa kata yang muncul pada dokumen mempunyai nilai. Kemunculan tersebut dapat ditandai dengan warna yang di mulai dari warna biru hingga warna merah. Semakin banyak jumlah kemunculan dari kata tersebut maka warna yang dihasilkan berwarna merah, begitu juga sebaliknya apabila *Bag of words* muncul lebih sedikit maka warna yang dihasilkan yaitu warna biru. Pada gambar ini dinilai bukan dari seberapa banyak kata itu muncul pada dokumen itu sendiri, tetapi seberapa sering kata itu muncul pada keseluruhan dokumen. Untuk lebih jelasnya dapat dilihat pada grafik dibawah ini.



Gambar 4. 2 Grafik kemunculan *bag of words*

Dari gambar 4.2 menunjukkan bahwa persentasi kata yang paling sering muncul pada tema neraka yaitu kata neraka itu sendiri sebesar 21%, kemudian disusul dengan kata jahannam 9%, kafir 8%, azab 7%, kekal 5%, dan seterusnya.

4.6 Matriks dari Hasil Perkalian *TF* dan *Probabilistic IDF*

(0, 0)	1.284686954807523
(0, 1)	1.284686954807523
(1, 1)	0.856457969871682
(1, 2)	0.428228984935841
(1, 3)	0.856457969871682
(1, 4)	0.428228984935841
(2, 5)	0.856457969871682
(2, 6)	0.856457969871682
(2, 7)	0.856457969871682
(3, 3)	1.284686954807523
(3, 4)	1.284686954807523
(4, 3)	1.712915939743364
(4, 4)	0.856457969871682
(5, 3)	0.856457969871682
(5, 8)	0.856457969871682
(5, 9)	0.856457969871682
(6, 3)	0.856457969871682
(6, 10)	0.856457969871682
(6, 11)	0.856457969871682
(7, 1)	0.5138747819230093
(7, 3)	1.0277495638460186
(7, 4)	0.5138747819230093
(7, 12)	0.5138747819230093
(8, 3)	1.284686954807523
(8, 4)	0.6423434774037615
:	:
(358, 11)	0.428228984935841
(358, 18)	0.428228984935841
(359, 13)	2.569373909615046
(360, 6)	1.284686954807523
(360, 18)	1.284686954807523
(361, 13)	2.569373909615046
(362, 11)	2.569373909615046
(363, 1)	1.712915939743364

Gambar 4. 3 Inputan untuk Proses *SVD*

Pada gambar 4.3 merupakan hasil dari pembobotan pada tiap-tiap kata yang telah melalui tahapan perkalian *TF* dan *Probabilistic-IDF*. Hasil ini sudah dalam bentuk matriks sehingga nantinya dapat di proses oleh *SVD*. Selain itu untuk pembobotan yang bernilai 0 (nol) tidak akan di munculkan karena *SVD* hanya memproses pembobotan yang ada nilainya saja.

4.7 Singular value decomposition (SVD)

Pada tahapan ini algoritma *SVD* diperlukan untuk mempermudah pengolahan data serta menyederhanakan matriks yang berukuran apa saja yang didapat setelah proses perhitungan *TF* dan *Probabilistic-IDF* dengan mendekomposisikan menjadi 3 komponen matriks yaitu menghasilkan nilai U , S , dan V^T . Tujuannya untuk menghilangkan fitur-fitur yang dirasa kurang penting dalam dokumen sehingga dapat memproduksi model data yang lebih baik.

Pada penelitian ini penulis menggunakan nilai dari matriks U dalam melakukan pengelompokan, karena nilai U merupakan proses dari terjemahan dokumen itu sendiri yang dikelompokkan berdasarkan konsep.

Berikut hasil dari nilai matriks U yang dikelompokkan oleh sistem secara otomatis dapat dilihat pada gambar 4.4 dibawah ini :

	0	1	2	3	4	5	6	\		
0	0.076878	-0.034896	-0.004698	0.008105	-0.020371	-0.001553	-0.002740			
1	0.059712	0.044907	-0.052069	-0.042989	-0.014281	0.003551	-0.002410			
2	0.009453	0.051258	0.018095	0.073979	0.003213	-0.001851	0.012971			
3	0.017273	0.112143	-0.076822	-0.073111	-0.006018	-0.000977	0.002503			
4	0.020750	0.134799	-0.094312	-0.089216	-0.004394	-0.007424	-0.000909			
..			
367	0.001461	0.003047	0.006529	-0.002394	-0.001971	0.001521	-0.005561			
368	0.081225	-0.018671	0.050842	-0.027596	0.072036	-0.009663	-0.006892			
369	0.146075	-0.068711	-0.014928	0.006689	-0.028352	-0.022210	0.001174			
370	0.025802	0.167297	-0.070297	-0.007281	0.003094	-0.024670	-0.008093			
371	0.005062	0.006844	0.000399	0.004635	-0.014440	0.190069	-0.174024			
	7	8	9	10						
0	0.012481	0.015862	-0.005131	0.006994						
1	0.060658	-0.028106	0.012144	-0.008866						
2	-0.000585	0.023575	0.008186	-0.029275						
3	0.002740	0.000379	-0.003027	-0.017774						
4	-0.000891	-0.005422	-0.005656	-0.010401						
..						
367	-0.005175	-0.003688	0.003520	-0.004067						
368	-0.001885	-0.005962	-0.003474	-0.015316						
369	-0.008676	-0.014243	-0.007415	0.001045						
370	-0.003572	-0.012283	-0.004814	0.001368						
371	-0.084638	0.003690	0.093493	0.011401						
	[372 rows x 11 columns]									

Gambar 4. 4 Nilai matriks U

Pada gambar 4.4 menunjukkan bahwa nilai U merupakan nilai matriks dari dokumen terjemahan itu sendiri yang telah dikelompokkan berdasarkan

konsep. Selanjutnya hasil pengelompokan tersebut diproses oleh sistem untuk menampilkan surah dan ayat yang telah dikelompokkan secara otomatis berdasarkan konsep.

4.8 Pengelompokan Surah dan Ayat Perkonsep Berdasarkan Matriks U

Pada tahap ini hasil dari *singular value decomposition* nilai matriks U digunakan untuk melakukan pengelompokan surah dan ayat dari keseluruhan dokumen yang terdiri dari 11 konsep sesuai dengan subtema neraka yaitu 11 subtema. Berikut hasil pengelompokan surah dan ayat berdasarkan nilai matriks U dapat dilihat pada tabel 4.10 di bawah ini :

Tabel 4. 10 Hasil pengelompokan SVD (U)

No	Pengelompokan	Surah dan Ayat	Jumlah
1	Konsep 0	(79, 39), (90, 20), (92, 14), (26, 96), (44, 47), (27, 90), (71, 25), (67, 7), (25, 12), (79, 36), (45, 34), (83, 16), (41, 19), (37, 163), (19, 70), (33, 66), (40, 48), (37, 23), (67, 8), (54, 48), (38, 59), (17, 63), (38, 85), (52, 13), (78, 21), (89, 23), (40, 49), (48, 13), (57, 15), (41, 24), (52, 27), (40, 46), (40, 47), (84, 12), (2, 167), (77, 32), (77, 31), (69, 31), (70, 15), (21, 99), (23, 104), (2, 119), (67, 10), (22, 51), (16, 62), (19, 72), (26, 94), (40, 7), (18, 53), (82, 14), (50, 24), (41, 40), (29, 25), (74, 27), (74, 26), (74, 42), (102, 6), (81, 12), (37, 68), (73, 12), (101, 10), (101, 9), (45, 35), (4, 145), (17, 39), (11, 106), (37, 64), (43, 77), (7, 41), (45, 10), (4, 56), (8, 14), (72, 15), (38, 61), (4, 169), (23, 103), (33, 64), (14, 29), (8, 16), (39, 60), (9, 109), (51, 13), (2, 206), (19, 86), (7, 18), (40, 60), (17, 18), (32, 13), (5, 37), (14, 30), (35, 37), (7, 38), (2, 24), (76, 4), (22, 22), (6, 70), (7, 44), (52, 18), (39, 71), (18, 102), (88, 4), (59, 17), (7, 47), (25, 11), (41, 28), (40, 43), (3, 192), (39, 72), (38, 56), (26, 91), (74, 29), (40, 11), (67, 5), (22, 72), (11, 16), (3, 12), (25, 13), (17, 8), (18, 106), (40, 6), (38, 27), (46, 20), (11, 17), (24, 57), (9, 68), (59, 3), (8, 37), (72, 23), (25, 34), (55, 43), (58, 17), (5, 10), (5, 86), (57, 19), (3, 151), (2, 126), (20, 74), (11, 119), (9, 35), (9, 17), (8, 50), (5, 72), (17, 97), (85, 10), (44, 56), (38, 57), (7, 36), (46, 34), (2, 175), (66, 6), (4, 97), (43, 74), (32, 20), (3, 116), (22, 19), (6, 128), (67, 11), (74, 31), (48, 6), (2, 39), (9, 81), (40, 76), (16, 29), (13, 5), (10, 27), (42, 45), (7, 179), (34, 42), . . .	277 Surah dan Ayat

No	Pengelompokan	Surah dan Ayat	Jumlah
		(39, 32), (29, 68), (98, 6), (2, 81), (35, 36), (5, 29), (64, 10), (2, 257), (58, 8), (30, 16), (9, 113), (39, 8), (4, 14), (7, 50), (21, 98), (36, 63), (4, 121), (50, 30), (40, 50), (9, 49), (18, 100), (18, 29), (29, 54), (17, 98), (40, 10), (90, 19), (84, 13), (8, 36), (75, 24), (79, 10), (2, 217), (3, 197), (51, 14), (34, 5), (31, 7), (43, 75), (38, 58), (29, 55), (33, 68), (67, 6), (4, 93), (38, 60), (3, 162), (25, 66), (13, 25), (21, 29), (4, 115), (66, 10), (39, 16), (2, 275), (25, 65), (56, 94), (2, 161), (9, 73), (66, 9), (15, 43), (19, 68), (23, 107), (9, 95), (4, 55), (10, 4), (104, 6), (104, 8), (111, 3), (21, 100), (4, 140), (76, 31), (4, 151), (25, 69), (22, 57), (14, 28), (11, 82), (17, 10), (48, 17), (45, 33), (13, 18), (73, 13), (15, 44), (27, 5), (3, 112), (14, 16), (25, 77), (39, 40), (30, 10), (5, 80), (11, 107), (33, 65), (47, 12), (101, 11), (52, 16), (14, 17), (31, 6), (9, 101), (20, 101), (40, 52), (2, 162), (74, 46), (83, 11), (92, 16), (40, 72), (6, 129), (4, 37), (6, 157), (32, 14), (40, 39), (38, 55), (56, 92), (56, 51), (67, 9)	
2	Konsep 1	(84, 13), (17, 98), (90, 19), (79, 10), (40, 10), (75, 24), (8, 36), (18, 100), (9, 49), (40, 50), (36, 63), (50, 30), (21, 98), (4, 121), (29, 54), (67, 6), (2, 161), (3, 197), (33, 68), (43, 75), (38, 58), (31, 7), (51, 14), (34, 5), (29, 55), (18, 102), (38, 60), (66, 10), (25, 66), (13, 25), (3, 162), (9, 73), (66, 9), (4, 93), (23, 107), (10, 4), (4, 151), (14, 28), (4, 115), (4, 140), (21, 29), (56, 94), (15, 43), (19, 68), (9, 95), (35, 36), (25, 65), (22, 57), (3, 12), (4, 56), (8, 14), (38, 56), (11, 17), (38, 27), (2, 257), (39, 71), (18, 106), (45, 10), (46, 34), (3, 112), (85, 10), (25, 69), (76, 31), (13, 18), (11, 82), (47, 12), (73, 13), (45, 33), (17, 10), (32, 20), (4, 97), (46, 20), (48, 17), (74, 31), (15, 44), (14, 16), (24, 57), (22, 19), (98, 6), (40, 6), (57, 15), (48, 13), (27, 5), (5, 80), (29, 68), (39, 32), (4, 55), (17, 8), (38, 61), (17, 97), (3, 116), (43, 74), (38, 85), (52, 13), (89, 23), (78, 21), (17, 63), (9, 68), (30, 10), (25, 77), (14, 29), (8, 16), (8, 37), (22, 72), (2, 217), (4, 37), (3, 151), (14, 17), (2, 126), (18, 29), (59, 3), (48, 6), (39, 40), (2, 24), (14, 30), (2, 206), (76, 4), (4, 169), (23, 103), (4, 161), (39, 72), (9, 101), (31, 6), (25, 34), (39, 60), (9, 109), (7, 18), (40, 60), (33, 64), (9, 17), (40, 49), (8, 50), (39, 16), (16, 29), (40, 76), (57, 19), (5, 10), (5, 86), (64, 10), (5, 72), (19, 86), (32, 13), (17, 18), (2, 39), (40, 52), (72, 23), (72, 15), (55, 43), (6, 70), (13, 5), (30, 16), (58, 8), (9, 35), (20, 74), (38, 55), (11, 119), (20, 101), (6, 157), (52, 27), (40, 46), (41, 24), (39, 8), (7, 179), (5, 37), (47, 15), (33, 65), (11, 107), (35, 37), (7, 50), (9, 113), (9, 81), (42, 45), (22, 22), (58, 17), (2, 162),	207 Surah dan Ayat

No	Pengelompokan	Surah dan Ayat	Jumlah
		(10, 27), (32, 14), (52, 18), (34, 42), (83, 11), (74, 46), (92, 16), (44, 56), (6, 129), (23, 108), (88, 2), (38, 57), (9, 63), (10, 52), (89, 25), (50, 26), (7, 39), (67, 8), (40, 39), (51, 13), (3, 11), (67, 9), (56, 51), (56, 92), (83, 12), (38, 62), (57, 13), (104, 7), (34, 8), (44, 48), (43, 62), (37, 63)	
...
10	Konsep 9	(33, 67), (23, 106), (26, 97), (26, 94), (26, 91), (14, 30), (6, 128), (34, 8), (31, 6), (20, 100), (56, 46), (55, 41), (56, 92), (56, 51), (67, 9), (25, 34), (33, 65), (11, 107), (4, 115), (74, 31), (7, 179), (2, 175), (7, 38), (17, 97), (20, 101), (39, 8), (69, 37), (44, 44), (2, 81), (40, 39), (18, 53), (21, 99), (9, 63), (2, 217), (43, 74), (40, 11), (5, 29), (74, 44), (41, 28), (5, 37), (25, 69), (4, 93), (4, 169), (23, 103), (48, 17), (67, 11), (9, 17), (20, 74), (9, 68), (58, 17), (10, 27), (72, 23), (88, 2), (23, 108), (13, 5), (3, 116), (88, 6), (59, 17), (69, 36), (25, 65), (16, 29), (40, 76), (2, 257), (73, 13), (98, 6), (42, 45), (56, 52), (37, 66), (37, 62), (2, 206), (2, 275), (38, 62), (47, 12), (3, 112), (39, 72), (2, 162), (39, 40), (11, 16), (90, 19), (8, 36), (79, 10), (40, 10), (84, 13), (17, 98), (75, 24), (2, 161), (87, 13), (17, 10), (74, 28), (74, 35), (4, 14), (27, 5), (47, 15)	93 Surah dan Ayat
11	Konsep 10	(56, 55), (78, 24), (88, 5), (56, 54), (55, 44), (44, 46), (44, 45), (56, 93), (56, 42), (38, 57), (37, 67), (18, 29), (10, 4), (101, 11), (52, 16), (14, 16), (44, 48), (9, 81), (6, 70), (47, 15), (40, 72), (9, 35), (88, 4), (74, 44), (33, 65), (11, 107), (88, 6), (37, 66), (37, 62), (56, 52), (69, 36), (22, 19), (44, 43), (73, 13), (21, 99), (40, 39), (104, 7), (47, 12), (44, 44), (69, 37), (7, 36), (41, 28), (23, 103), (4, 169), (37, 64), (6, 128), (5, 37), (4, 93), (58, 17), (2, 39), (7, 50), (23, 106), (33, 67), (26, 97), (2, 275), (17, 60), (13, 5), (9, 17), (56, 94), (72, 23), (56, 92), (56, 51), (67, 9), (9, 68), (3, 116), (74, 46), (83, 11), (92, 16), (74, 29), (4, 161), (59, 17), (11, 82), (25, 65), (25, 11), (26, 94), (2, 119), (22, 51), (67, 10), (64, 10), (2, 257), (2, 24), (16, 62), (22, 22), (74, 28), (74, 35), (26, 91), (43, 74)	87 Surah dan Ayat

Pada tabel 4.10 diatas merupakan hasil dari pengelompokan surah dan ayat berdasarkan konsep yang didapat dari nilai matriks U, selanjutnya akan di evaluasi menggunakan data tematik Al-Qur'an Cordoba khusus tema neraka, sehingga nantinya dapat dilihat akurasi kecocokan antara konsep yang terbentuk dengan sub-sub tema neraka yang ada pada Al-Qur'an Cordoba.

4.9 Evaluasi Sistem

Pada penelitian ini tahap evaluasi yang dilakukan sebanyak dua kali yaitu, pertama evaluasi terhadap Al-Qur'an Cordoba dengan menggunakan sistem dan yang kedua yaitu, evaluasi menggunakan ahli tafsir Al-Qur'an dengan cara mengisi kuisioner terhadap data yang tidak tepat dari hasil evaluasi sistem.

4.9.1 Data Evaluasi

Setelah melakukan tahap pengelompokan dengan menggunakan algoritma SVD berdasarkan nilai dari matriks U, langkah selanjutnya yaitu mengevaluasi hasil pengelompokan tersebut dengan data tematik Al-Qur'an Amazing atau disebut juga dengan Al-Qur'an Cordoba khusus tema neraka.

Berikut ini data tematik Al-Qur'an Cordoba tema neraka yang digunakan sebagai data evaluasi dapat dilihat pada tabel 4.11 :

Tabel 4. 11 Inputan Data *Gold Standard*

No	Subtema	Surah dan Ayat	Jumlah
1	Nama-Nama Neraka	(2, 119), (2, 206), (3, 112), (3, 162), (3, 197), (4, 55), (4, 93), (4, 97), (4, 115), (4, 121), (4, 140), (4, 169), (5, 10), (5, 86), (7, 18), (7, 41), (7, 179), (8, 16), (8, 36), (8, 37), (9, 35), (9, 49), (9, 63), (9, 68), (9, 73), (9, 81), (9, 95), (9, 109), (9, 113), (11, 82), (11, 119), (13, 18), (14, 16), (14, 29), (15, 43), (16, 29), (17, 8), (17, 18), (17, 39), (17, 63), (17, 98), (18, 100), (18, 102), (18, 106), (19, 68), (19, 86), (20, 74), . . .	115 Surah & Ayat

No	Subtema	Surah dan Ayat	Jumlah
		(21, 29), (21, 98), (22, 51), (23, 103), (25, 34), (25, 65), (26, 91), (29, 54), (29, 68), (30, 10), (32, 13), (35, 36), (36, 63), (37, 23), (37, 63), (37, 163), (38, 56), (38, 85), (39, 32), (39, 60), (39, 71), (39, 72), (40, 7), (40, 49), (40, 60), (40, 76), (43, 62), (43, 74), (44, 47), (44, 56), (45, 10), (48, 6), (50, 24), (50, 30), (52, 13), (52, 18), (52, 27), (54, 48), (56, 93), (56, 94), (57, 19), (58, 8), (66, 9), (67, 5), (67, 6), (67, 10), (67, 11), (69, 31), (72, 15), (72, 23), (74, 26), (74, 27), (74, 42), (78, 21), (79, 10), (79, 14), (79, 36), (79, 39), (81, 12), (82, 14), (83, 16), (85, 10), (89, 23), (98, 6), (101, 9), (102, 6), (104, 4), (104, 5)	
2	Keabadian Neraka	(2, 39), (2, 81), (2, 162), (2, 167), (2, 217), (2, 257), (2, 275), (4, 14), (4, 169), (5, 37), (5, 80), (6, 128), (7, 36), (9, 17), (9, 63), (9, 68), (10, 27), (10, 52), (11, 107), (13, 5), (14, 17), (16, 29), (20, 74), (20, 101), (21, 99), (23, 103), (25, 66), (25, 69), (32, 14), (33, 65), (39, 40), (39, 72), (40, 11), (40, 39), (40, 76), (41, 28), (42, 45), (43, 74), (43, 77), (45, 35), (47, 15), (58, 17), (59, 17), (64, 10), (78, 23), (98, 6)	46 Surah & Ayat
3	Memasuki Neraka	(2, 24), (3, 192), (19, 86), (25, 13), (25, 14), (26, 94), (26, 95), (27, 90), (37, 68), (39, 71), (40, 60), (40, 71), (40, 72), (40, 76), (44, 47), (52, 13), (52, 16), (54, 48), (55, 41), (67, 7), (67, 8), (69, 30), (69, 31), (78, 21)	24 Surah & Ayat
4	Penjaga Neraka	(39, 71), (40, 49), (43, 77), (66, 6), (67, 8), (74, 30), (74, 31), (96, 18)	8 Surah & Ayat
5	Sifat Neraka	(2, 24), (7, 38), (7, 41), (9, 81), (11, 6), (13, 18), (14, 16), (14, 17), (14, 29), (15, 44), (16, 29), (17, 8), (17, 97), (18, 29), (18, 53), (21, 100), (22, 19), (22, 20), (22, 21), (22, 22), (23, 104), (24, 57), (25, 12), (25, 13), (25, 34), (25, 65), (25, 66), (25, 77), (29, 55), (37, 62), (37, 63), (37, 64), (37, 65), (37, 66), (37, 67), (38, 57), (38, 58), (39, 16), (40, 49), (40, 52), (43, 75), (44, 43), (44, 44), (44, 45), (44, 46), (44, 48), (47, 15), (50, 30), (54, 48), (55, 44), (56, 42), (56, 43), (56, 44), (56, 93), (56, 94), (58, 8), (66, 6), (67, 7), (67, 8), (69, 31), (69, 32), (70, 15), (70, 16), (73, 12), (74, 28), (74, 29), (74, 35), (76, 4), (77, 30), (77, 31), (77, 32), (77, 33), (78, 21), (88, 4), (88, 5), (90, 11), (90, 12), (90, 20), (92, 14), (101, 10), (101, 11), (104, 4), (104, 5), (104, 6), (104, 7), (104, 8), (104, 9), (111, 3), (111, 5)	89 Surah & Ayat

No	Subtema	Surah dan Ayat	Jumlah
6	Sifat Ahli Neraka dan Kejahatan Mereka	(2, 24), (2, 39), (2, 81), (2, 126), (2, 161), (2, 162), (2, 175), (2, 206), (2, 217), (2, 257), (2, 275), (3, 11), (3, 12), (3, 151), (3, 162), (4, 14), (4, 37), (4, 55), (4, 56), (4, 93), (4, 97), (4, 115), (4, 121), (4, 140), (4, 145), (4, 151), (4, 161), (4, 169), (5, 10), (5, 29), (5, 72), (5, 80), (5, 86), (6, 128), (6, 129), (6, 157), (7, 18), (7, 36), (7, 38), (7, 39), (7, 41), (7, 47), (7, 50), (7, 179), (8, 14), (8, 16), (8, 36), (8, 37), (8, 50), (9, 17), (9, 49), (9, 63), (9, 68), (9, 73), (9, 81), (9, 95), (9, 109), (9, 113), (10, 4), (10, 27), (10, 52), (11, 16), (11, 17), (11, 106), (11, 119), (13, 5), (13, 18), (13, 25), (14, 17), (14, 29), (14, 30), (15, 43), (16, 29), (16, 62), (17, 8), (17, 10), (17, 18), (17, 39), (17, 63), (17, 97), (18, 29), (18, 102), (18, 105), (18, 106), (19, 68), (19, 70), (19, 72), (20, 74), (20, 100), (21, 100), (22, 51), (22, 57), (22, 72), (23, 103), (23, 104), (25, 11), (25, 13), (25, 34), (25, 69), (25, 77), (26, 91), (26, 94), (26, 95), (26, 96), (26, 97), (26, 98), (26, 100), (26, 101), (27, 5), (29, 25), (29, 54), (29, 68), (30, 16), (31, 6), (31, 7), (32, 14), (32, 20), (33, 64), (33, 66), (34, 5), (34, 8), (34, 42), (35, 36), (35, 37), (37, 161), (38, 27), (38, 55), (38, 56), (39, 8), (39, 16), (39, 32), (39, 60), (39, 72), (40, 6), (40, 10), (40, 43), (40, 46), (40, 52), (40, 60), (40, 71), (40, 72), (40, 76), (41, 19), (41, 24), (41, 28), (41, 40), (43, 74), (43, 75), (45, 10), (45, 33), (45, 34), (45, 35), (46, 20), (46, 34), (47, 12), (48, 6), (48, 13), (48, 17), (50, 24), (50, 25), (50, 26), (51, 13), (51, 14), (52, 13), (55, 41), (55, 43), (55, 44), (56, 9), (56, 41), (56, 45), (56, 46), (56, 51), (56, 92), (57, 13), (57, 15), (57, 19), (58, 8), (58, 17), (59, 3), (59, 17), (64, 10), (66, 9), (66, 10), (67, 6), (67, 10), (67, 11), (70, 17), (70, 18), (71, 25), (72, 15), (72, 23), (74, 43), (74, 44), (74, 45), (74, 46), (75, 24), (76, 31), (78, 22), (78, 23), (78, 24), (82, 14), (83, 11), (83, 12), (84, 10), (84, 11), (84, 12), (84, 13), (85, 10), (88, 2), (88, 3), (88, 4), (89, 24), (89, 25), (89, 26), (90, 19), (92, 15), (92, 16), (98, 6), (101, 9), (104, 4), (104, 5), (111, 3)	222 Surah & Ayat
7	Siksaan Ahli Neraka Dilipatgandakan	(9, 101), (25, 69)	2 Surah & Ayat

No	Subtema	Surah dan Ayat	Jumlah
8	Mereka yang Kekal Dalam Neraka	(3, 116), (4, 56), (5, 37), (5, 80), (6, 128), (7, 36), (9, 63), (9, 68), (10, 27), (10, 52), (14, 16), (14, 17), (14, 28), (22, 22), (23, 103), (23, 108), (26, 102), (33, 65), (45, 35), (58, 17), (59, 17), (64, 10), (72, 23), (87, 13), (98, 6)	25 Surah & Ayat
9	Makanan dan Minuman Ahli Neraka	(6, 70), (10, 4), (14, 16), (18, 29), (22, 20), (37, 62), (37, 63), (37, 64), (37, 65), (37, 66), (44, 43), (44, 44), (44, 45), (44, 46), (56, 52), (56, 53), (56, 54), (56, 55), (69, 36), (69, 37), (73, 13), (88, 5), (88, 6), (88, 7)	24 Surah & Ayat
10	Pohon Zaquum	(17, 60), (37, 62), (37, 63), (37, 64), (37, 65), (37, 66), (37, 67), (44, 43)	8 Surah & Ayat
11	Percakapan Ahli Neraka	(7, 38), (7, 39), (7, 44), (7, 50), (23, 106), (23, 107), (25, 13), (25, 14), (26, 96), (26, 97), (26, 98), (33, 67), (33, 68), (35, 37), (38, 59), (38, 60), (38, 61), (38, 62), (38, 63), (39, 71), (40, 47), (40, 48), (40, 49), (40, 50), (43, 77), (67, 9), (67, 10)	27 Surah & Ayat

Pada tabel 4.11 merupakan data yang akan dijadikan sebagai data *gold standard* untuk tahap evaluasi menggunakan perhitungan *precision*, *recall*, dan *f1-score*. Data ini berasal dari Al-Qur'an amazing atau disebut juga dengan Al-Qur'an Cordoba yang terdiri dari 11 subtema dan 590 surah dan ayat.

4.9.2 Evaluasi Pengelompokan Berdasarkan Al-Qur'an Cordoba

Pada tahap ini pengelompokan surah dan ayat berdasarkan nilai matriks U yang diperoleh melalui algoritma *singular value decomposition*, akan di evaluasi dengan data tematik Al-Qur'an Cordoba yang bertema neraka sesuai dengan konsep yang terbentuk yaitu sebanyak 11 konsep dengan menggunakan perhitungan *precision*, *recall*, dan *f1-score*. Berikut ini hasil evaluasi data tematik Al-Qur'an Cordoba dapat dilihat pada tabel 4.12 :

Tabel 4. 12 Hasil Data Evaluasi Tematik Al-Qur'an Cordoba

Pengelompokan	No	Sub Tema	TP	FP	FN	Precision	Recall	F1-Score
Konsep 0	1	Nama-Nama Neraka	108	171	7	38%	93%	54%
	2	Keabadian Neraka	43	236	3	15%	93%	26%
	3	Memasuki Neraka	19	260	5	6%	79%	12%
	4	Penjaga Neraka	6	273	2	2%	75%	4%
	5	Sifat Neraka	57	222	32	20%	64%	30%
	6	Sifat Ahli Neraka dan Kejahatan Mereka	180	99	42	64%	81%	71%
	7	Siksaan Ahli Neraka Dilipatgandakan	2	227	0	0%	100%	1%
	8	Mereka yang Kekal Dalam Neraka	20	259	5	7%	80%	13%
	9	Makanan dan Minuman Ahli Neraka	6	273	18	2%	25%	3%
	10	Pohon Zaquum	1	277	7	0%	12%	0%
	11	Percakapan Ahli Neraka	19	260	8	6%	70%	12%
Konsep 1	1	Nama-Nama Neraka	86	124	29	40%	74%	52%
	2	Keabadian Neraka	33	177	13	15%	71%	25%
	3	Memasuki Neraka	8	202	16	3%	33%	6%
	4	Penjaga Neraka	4	206	4	1%	50%	3%
	5	Sifat Neraka	35	175	54	16%	39%	23%
	6	Sifat Ahli Neraka dan Kejahatan Mereka	148	62	74	70%	66%	68%
	7	Siksaan Ahli Neraka Dilipatgandakan	2	207	0	0%	100%	1%
	8	Mereka yang Kekal Dalam Neraka	19	191	6	9%	76%	16%
	9	Makanan dan Minuman Ahli Neraka	7	203	17	3%	29%	5%
	10	Pohon Zaquum	1	207	7	0%	12%	0%
	11	Percakapan Ahli Neraka	12	198	15	5%	44%	10%
...

Pengelompokan	No	Sub Tema	TP	FP	FN	Precision	Recall	F1-Score
Konsep 9	1	Nama-Nama Neraka	20	70	95	22%	17%	19%
	2	Keabadian Neraka	35	55	11	38%	76%	51%
	3	Memasuki Neraka	3	87	21	3%	12%	5%
	4	Penjaga Neraka	1	89	7	1%	12%	2%
	5	Sifat Neraka	14	76	75	15%	15%	15%
	6	Sifat Ahli Neraka dan Kejahatan Mereka	59	31	163	65%	26%	37%
	7	Siksaan Ahli Neraka Dilipatgandakan	1	89	1	1%	50%	2%
	8	Mereka yang Kekal Dalam Neraka	15	75	10	16%	60%	26%
	9	Makanan dan Minuman Ahli Neraka	8	82	16	8%	33%	14%
	10	Pohon Zaquum	2	88	6	2%	25%	4%
	11	Percakapan Ahli Neraka	7	83	20	7%	25%	11%
Konsep 10	1	Nama-Nama Neraka	19	71	96	21%	16%	18%
	2	Keabadian Neraka	19	71	27	21%	41%	27%
	3	Memasuki Neraka	4	86	20	4%	16%	7%
	4	Penjaga Neraka	0	87	8	0%	0%	0%
	5	Sifat Neraka	29	61	60	32%	32%	32%
	6	Sifat Ahli Neraka dan Kejahatan Mereka	44	46	178	48%	19%	28%
	7	Siksaan Ahli Neraka Dilipatgandakan	0	87	2	0%	0%	0%
	8	Mereka yang Kekal Dalam Neraka	12	78	13	13%	48%	20%
	9	Makanan dan Minuman Ahli Neraka	19	71	5	21%	79%	33%
	10	Pohon Zaquum	6	84	2	6%	75%	12%
	11	Percakapan Ahli Neraka	6	84	21	6%	22%	10%

Keterangan :

TP = *True Positif*

FP = *False Positif*

FN = *False Negatif*

Pada tabel 4.12 dapat dilihat bahwa hasil evaluasi yang dilakukan dengan menggunakan perhitungan *precision*, *recall*, dan *f1-score* masing-masing setiap konsep dan subtema memiliki nilai *f1-score* yang berbeda-beda. Dari keseluruhan data tersebut, masing-masing konsep selanjutnya di analisa kembali untuk mengambil konsep dan subtema berdasarkan hasil dari nilai *f1-score* yang terbaik atau besar dari 70% sehingga terbentuk sebuah pengelompokan yang memiliki akurasi yang lebih baik. Berikut nilai *f1-score* yang terbaik dapat dilihat pada gambar 4.7 dibawah ini :

Konsep 0 Sifat Ahli Neraka dan Kejahatan Mereka
Surat & Ayat Tepat (TP) 180
Surah & Ayat Tidak Tepat (FP) 99
Surah & Ayat Tidak Tepat (FN) 42
Precision 64.51612903225806 %
Racall 81.08108108108108 %
F1-Score 71.8562874251497 %

Gambar 4. 5 Hasil Evaluasi *F1-Score* Terbaik

Pada gambar 4.7 dapat dilihat bahwa hasil evaluasi dari pengujian *f1-score* terbaik hanya menghasilkan 1 buah konsep saja yaitu konsep 0 dengan subtema “Sifat Ahli Neraka dan Kejahatan Mereka” dengan bobot *f1-score* sebesar 71%. Selain itu evaluasi ini juga menghasilkan data yang tepat (TP) sebanyak 180 surah dan ayat, sedangkan data tidak tepat (FP) sebanyak 99 surah dan ayat. Untuk data yang tidak tepat (FP) ini akan di uji kembali oleh pakar tematik atau tafsir Al-

Qur'an dengan tujuan untuk mendapatkan kontribusi dari hasil evaluasi pakar tersebut.

4.9.3 Evaluasi Pakar Berdasarkan Hasil Data *False Positif* (FP)

Pada tahap ini evaluasi dilakukan oleh dua orang pakar atau tafsir Al-Qur'an yaitu bapak Dr. Hamzah, M.Ag selaku wakil dekan Fakultas Agama Islam di Universitas Islam Riau, dan bapak Musaddad Harahap M.Pd.I selaku sekretaris prodi Pendidikan Agama Islam di Fakultas Agama Islam Universitas Islam Riau. Data *false positif* yang akan di evaluasi sebanyak 99 surah dan ayat.

Berikut ini data *false positif* yang akan di evaluasi oleh kedua pakar :

Tabel 4. 13 Data *False Positif* (FP)

Surah & Ayat Tidak Tepat	Jumlah
(90, 20), (37, 163), (40, 48), (37, 23), (79, 36), (79, 39), (83, 16), (27, 90), (92, 23), (40, 49), (52, 27), (40, 47), (70, 15), (77, 31), (2, 167), (77, 32), (69, 31), (21, 99), (2, 119), (18, 53), (40, 7), (74, 26), (74, 42), (74, 27), (102, 6), (81, 12), (37, 6814), (67, 7), (44, 47), (25, 12), (67, 8), (54, 48), (38, 59), (78, 21), (38, 85), (89,), (73, 12), (101, 10), (37, 64), (43, 77), (38, 61), (19, 86), (32, 13), (5, 37), (76, 4), (22, 22), (6, 70), (7, 44), (52, 18), (39, 71), (3, 192), (74, 29), (40, 11), (67, 5), (24, 57), (9, 35), (44, 56), (38, 57), (66, 6), (3, 116), (22, 19), (74, 31), (42, 45), (21, 98), (36, 63), (50, 30), (40, 50), (18, 100), (79, 10), (17, 98), (3, 197), (29, 55), (33, 68), (38, 58), (38, 60), (25, 66), (21, 29), (25, 65), (56, 94), (23, 107), (104, 8), (104, 6), (14, 28), (11, 82), (73, 13), (15, 44), (3, 112), (14, 16), (39, 40), (30, 10), (33, 65), (11, 107), (101, 11), (52, 16), (9, 101), (47, 15), (20, 101), (40, 39), (67, 9)	99 Surah & Ayat

Pada tabel 4.13 merupakan data yang akan di evaluasi oleh pakar tematik tujuannya untuk menganalisa apakah surah dan ayat tersebut termasuk kedalam pengelompokan konsep 0 dengan subtema "Sifat Ahli Neraka dan Kejahatan Mereka". Selain itu hasil dari evaluasi yang yang termasuk kedalam pengelompokan nantinya akan mendapatkan tambahan ataupun kontribusi surah dan ayat pada subtema tersebut.

Berikut hasil evaluasi yang telah dilakukan oleh pakar tematik atau tafsir

Al-Qur'an dapat dilihat pada tabel 4.14 dibawah ini :

Tabel 4. 14 Hasil Evaluasi Pakar Tematik

No	Subtema	Surah	Ayat	Pakar 1		Pakar 2		Hasil Evaluasi
				Ya	Tidak	Ya	Tidak	
1	Sifat Ahli Neraka dan Kejahatan Mereka	79	39	√			√	
2		90	20	√			√	
3		92	14	√			√	
4		44	47	√			√	
5		27	90	√		√		√
6		67	7	√			√	
7		25	12	√			√	
8		79	36	√			√	
9		83	16	√			√	
10		37	163	√			√	
11		40	48	√		√		√
12		37	23		√		√	
13		67	8	√		√		√
14		54	48	√			√	
15		38	59	√		√		√
16		38	85		√		√	
17		78	21	√			√	
18		89	23	√		√		√
19		40	49	√			√	
20		52	27		√		√	
21		40	47	√		√		√
22		2	167	√		√		√
23		77	32	√			√	
24		77	31	√			√	
25		69	31	√			√	
26		70	15	√			√	
27		21	99		√	√		
28		2	119		√		√	
29		40	7	√			√	
30		18	53	√		√		√
31		74	27	√			√	
32		74	26	√			√	
33		74	42	√			√	
34		102	6	√			√	

No	Subtema	Surah	Ayat	Pakar 1		Pakar 2		Hasil Evaluasi
				Ya	Tidak	Ya	Tidak	
35	Sifat Ahli Neraka dan Kejahatan Mereka	81	12	√			√	
36		37	68	√			√	
37		73	12	√			√	
38		101	10	√			√	
39		37	64	√			√	
40		43	77	√			√	
41		38	61	√			√	
42		19	86	√		√		√
43		32	13		√		√	
44		5	37	√			√	
45		76	4	√		√		√
46		22	22	√			√	
47		6	70	√		√		√
48		7	44	√		√		√
49		52	18		√		√	
50		39	71	√		√		√
51		3	192	√		√		√
52		74	29	√			√	
53		40	11	√		√		√
54		67	5	√			√	
55		24	57	√		√		√
56		9	35	√		√		√
57		44	56	√			√	
58		38	57	√			√	
59		66	6	√			√	
60		3	116	√		√		√
61		22	19	√		√		√
62		74	31	√		√		√
63		42	45	√		√		√
64		21	98	√		√		√
65		36	63	√			√	
66		50	30	√			√	
67		40	50	√		√		√
68		18	100		√	√		
69		17	98	√		√		√
70		79	10	√		√		√
71		3	197	√			√	
72		38	58	√			√	
73		29	55	√			√	

No	Subtema	Surah	Ayat	Pakar 1		Pakar 2		Hasil Evaluasi
				Ya	Tidak	Ya	Tidak	
74	Sifat Ahli Neraka dan Kejahatan Mereka	33	68		√		√	
75		38	60	√			√	
76		25	66	√			√	
77		21	29	√		√		√
78		25	65	√			√	
79		56	94	√			√	
80		23	107	√		√		√
81		104	6	√			√	
82		104	8	√			√	
83		14	28		√		√	
84		11	82		√	√		
85		73	13	√			√	
86		15	44	√			√	
87		3	112		√	√		
88		14	16	√			√	
89		39	40	√			√	
90		30	10		√	√		
91		11	107	√			√	
92		33	65	√			√	
93		101	11	√			√	
94		52	16	√			√	
95		9	101		√	√		
96		47	15		√		√	
97		20	101	√			√	
98	40	39		√		√		
99	67	9	√		√		√	

Keterangan :

Pakar 1 = Dr. Hamzah, M.Ag

Pakar 2 = Musaddad Harahap M.Pd.I

Pada tabel 4.14 merupakan hasil evaluasi dari data *false positif* yang didapat dari sistem. Dasar-dasar atau indikator yang digunakan oleh pakar 1 untuk mengevaluasi hasil dari sistem adalah buku “Ulumul Qur’an” sedangkan pakar 2 menggunakan buku “Indeks Al-Qur’an Azharuddin Sahil. Dari hasil evaluasi yang

dilakukan oleh kedua pakar tersebut maka didapatkan nilai *F1-Score* yang berbeda. Perbedaan hasil evaluasi tersebut dapat dilihat pada tabel berikut :

Tabel 4. 15 Perbedaan Hasil Evaluasi Kedua Pakar

Pengelompokan Konsep 0				
Subtema Sifat Ahli Neraka dan Kejahatan Mereka				
Pakar	F1-Score	Cordoba	Selisih	Peningkatan
Pakar 1	88.92 %	71.85 %	17.07%	23.75%
Pakar 2	77.24 %		5.39%	7.5%

Keterangan :

Pakar 1 = Dr. Hamzah, M.Ag

Pakar 2 = Musaddad Harahap M.Pd.I

Pada tabel 4.15 dapat dilihat bahwa nilai *F1-Score* tertinggi adalah hasil evaluasi yang dilakukan oleh pakar 1 yaitu sebesar 88.92% sedangkan pakar 2 hanya 77.24%.

4.9.4 Perbedaan Hasil Evaluasi Al-Qur'an Cordoba dengan Hasil Evaluasi Pakar

Pada tahap ini hasil evaluasi yang dilakukan oleh kedua pakar selanjutnya diseleksi kembali menggunakan logika AND, dimana evaluasi yang diambil hanya evaluasi yang bernilai “Ya” pada kedua pakar. Dari hasil evaluasi tersebut terdapat 28 surah dan ayat yang menjadi tambahan atau kontribusi terhadap pengelompokan tematik Al-Qur'an Cordoba pada konsep 0 dan subtema “sifat ahli neraka dan kejahatan mereka”.

Berikut ini perbedaan hasil evaluasi setelah mendapatkan kontribusi dari kedua pakar dapat dilihat pada tabel 4.16 dibawah ini :

Tabel 4. 16 Perbedaan Hasil Evaluasi Al-Qur'an Cordoba dengan Pakar

Pengelompokan	F1-Score		Selisih	Peningkatan
	Cordoba	Pakar		
Konsep 0 Subtema Sifat Ahli Neraka dan Kejahatan Mereka	71.85 %	78.55 %	6.7%	9.32 %

Pada tabel 4.16 dapat dilihat bahwa hasil evaluasi *f1-score* pada pengelompokan konsep 0 subtema “Sifat Ahli Neraka dan Kejahatan Mereka” meningkat setelah melakukan pengujian terhadap pakar tematik. Dari yang sebelumnya nilai *f1-score* berdasarkan Cordoba 71,85% menjadi 78,55% setelah mendapatkan kontribusi dari pakar tematik. Hasil tersebut memiliki selisih sebesar 6,7% yang di dapat dari pengurangan nilai *f1-score* dari pakar dengan nilai *f1-score* Cordoba. Selain itu hasil evaluasi tersebut juga mengalami peningkatan sebesar 9,32%. Hasil peningkatan tersebut di dapat dari angka selisih di bagi dengan nilai *f1-score* Cordoba kemudian di kali dengan 100%.

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan

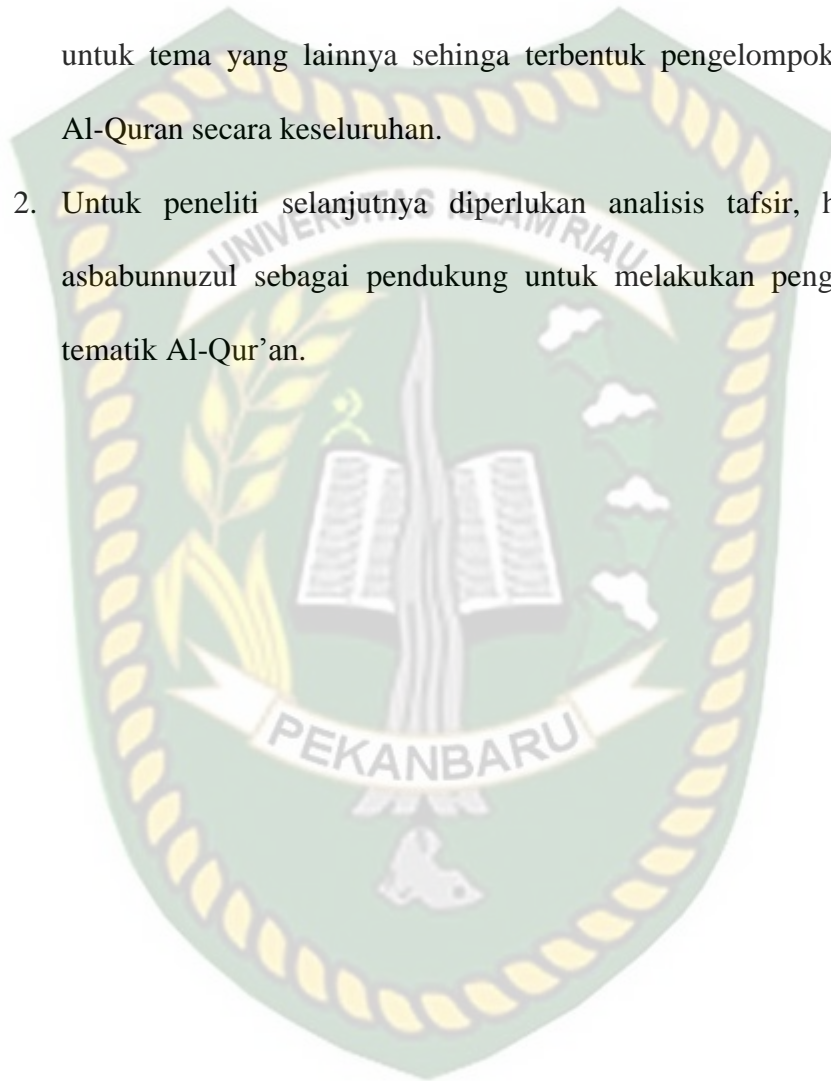
Berdasarkan hasil penelitian yang telah dilakukan oleh peneliti melalui analisis yang telah dilakukan pada bab-bab sebelumnya maka dapat diambil kesimpulan beberapa hal sebagai berikut, yaitu :

1. Terciptanya kumpulan kata kunci atau *bag of words* tema neraka yang nantinya bisa digunakan oleh penelitian lebih lanjut.
2. Terciptanya sebuah pengelompokan tematik Al-Qur'an secara otomatis menggunakan metode *latent semantic analysis* dengan tema neraka.
3. Pada tahap pengelompokan tematik Al-Qur'an tema neraka memperoleh hasil persentasi evaluasi menggunakan metode pengujian *Precision, Recall, F1-Score* besar dari 70%.
4. Terdapat penambahan kontribusi sebanyak 28 surah dan ayat dari hasil evaluasi pakar atau tafsir Al-Qur'an sehingga nilai *F1-Score* mendapatkan peningkatan sebesar 9,32%.

5.2 Saran

Berdasarkan hasil penelitian, agar tercapainya efektifitas dan efisiensi sesuai dengan tujuan peneliti, maka peneliti merekomendasikan atau menyarankan beberapa hal mengenai pengelompokan tematik Al-Qur'an sebagai berikut :

1. Pada penelitian ini hanya membahas tentang pengelompokan tematik Al-Qur'an tema neraka saja, penulis berharap kedepannya ada pengembang yang melakukan pengelompokan tematik Al-Qur'an untuk tema yang lainnya sehingga terbentuk pengelompokan tematik Al-Quran secara keseluruhan.
2. Untuk peneliti selanjutnya diperlukan analisis tafsir, hadist, dan asbabunnuzul sebagai pendukung untuk melakukan pengelompokan tematik Al-Qur'an.



Dokumen ini adalah Arsip Miik :

Perpustakaan Universitas Islam Riau

DAFTAR PUSTAKA

- A. S., Rosa dan Shalahuddin, M. 2013. Rekayasa Perangkat Lunak Terstruktur Dan Berorientasi Objek. Informatika. Bandung.
- Al-Bahra Bin Ladjamudin. 2006. Analisis dan Desain Sistem Informasi. Graha Ilmu. Yogyakarta.
- Dary, M. I. (2015). Analisis dan Implementasi Short Text Similarity dengan Metode Latent Semantic Analysis Untuk Mengetahui Kesamaan Ayat al-Quran. Universitas Telkom. Bandung
- Fitri, M. (2013). Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi *TF-IDF* Untuk Pencarian Dokumen Berbahasa Indonesia. Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi *TF-IDF* Untuk Pencarian Dokumen Berbahasa Indonesia. Universitas Tanjungpura. Pontianak.
- Hilwah, N., Kudus, A., & Sunendiari, S. (n.d.). Klasifikasi Text Mining untuk Terjemahan Ayat-Ayat Al- Qur ' an menggunakan Metode Klasifikasi Naive Bayes. Universitas Islam Bandung.
- Jadhira, A. A., Bijaksana, M. A., & Wahyudi, B. A. (2018). Deteksi Kemiripan Bagian-bagian Terjemah Al-Qur'an dengan Menggunakan Metode Latent Semantic Analysis. Universitas Telkom. Bandung.
- Luthfiarta, A., Zeniarja, J., & Salam, A. (2013). Algoritma Latent Semantic Analysis (*LSA*) Pada Peringkat Dokumen Otomatis Untuk Proses Clustering Dokumen. Universitas Dian Nuswantoro. Semarang.
- Mardiana, T., Nyoto, R. D., Studi, P., & Informatika, T. (2015). *Kluster Bag-of-Word Menggunakan Weka*. Universitas Tanjungpura. Pontianak
- Mujilahwati, S., Studi, P., Informatika, T., Teknik, F., Lamongan, U. I., & Mining, D. (2016). PRE-PROCESSING TEXT MINING PADA DATA TWITTER. Universitas Lamongan.
- Norman E.Gronlund, (1985). Measurement and Evaluation : Macmillan New York.
- Nur Ichwan, Mohammad, 2004. Tafsir 'Ilmiy Memahami Al-Qur'an Melalui Pendekatan Sains Modern. Yogyakarta: Menara Kudus.
- Pangestu, A. Gelar B., Irma S. (2017). Analisis Image Watermarking Menggunakan Compressive Sensing Algoritma Orthogonal Matching Pursuit dengan Pendekatan Berbasis Discrete Cosine Transform

Menggunakan Singular Value Decomposition. Bandung: Universitas Telkom. Bandung.

Polettini, N. (2004). *The Vector Space Model in Information Retrieval - Term Weighting Problem Local Term-Weighting*. Universitas of Trento.

Prasetya, D. A., & Nurviyanto, I. (2012). Deteksi wajah metode viola jones pada opencv menggunakan pemrograman python. Universitas Muhammadiyah Surakarta.

Siallagan, Sariadin. 2009. Pemrograman Java. Yogyakarta: Andi Yogyakarta.

Sukmana, H. T. (2018). Development of Semantic Web of Indonesian Qur ' an INTERNATIONAL COLLABORATIVE RESEARCH TAHUN ANGGARAN 2015 DEVELOPMENT OF SEMANTIC WEB. Jakarta.

Wahyuni, R. T., Prastiyanto, D., & Supraptono, E. (2017). Penerapan Algoritma Cosine Similarity dan Pembobotan *TF-IDF* pada Sistem Klasifikasi Dokumen Skripsi. Jurnal Teknik Elektro Universitas Negeri Semarang.

Winata, F., & Rainarli, E. (2016). Implementasi Cross Method Latent Semantic Analysis untuk Meringkas Dokumen Berita Berbahasa Indonesia. Universitas Komputer Indonesia. Bandung.

Yamani, M. T. (2015). Memahami Al-Qur'an Dengan Metode Tafsir Maudhu'I. UIN Maulana Malik Ibrahim. Malang.