

Text Mining Approaches for Analyzing an Indonesian Tafseer and Translation
of The Holy Quran

Information Extraction

Oleh

Kelompok 4 & 6

Anggota

Text Mining Approaches for Analyzing an Indonesian Tafseer and Translation of The Holy Quran

Kelompok 4	Kelompok 6
5200411042 - Afifah Inas Hanifah	5200411345 - Alexander Romian Simarmata
5200411196 - Nasmah Nur Amiroh	5200411090 - Nurcholis Majid
5200411233 - Siti Nurazila	5200411578 - Ilham muhammad furqon

Materi Pembahasan

Text Mining Approaches for Analyzing an Indonesian Tafseer and Translation of The Holy Quran

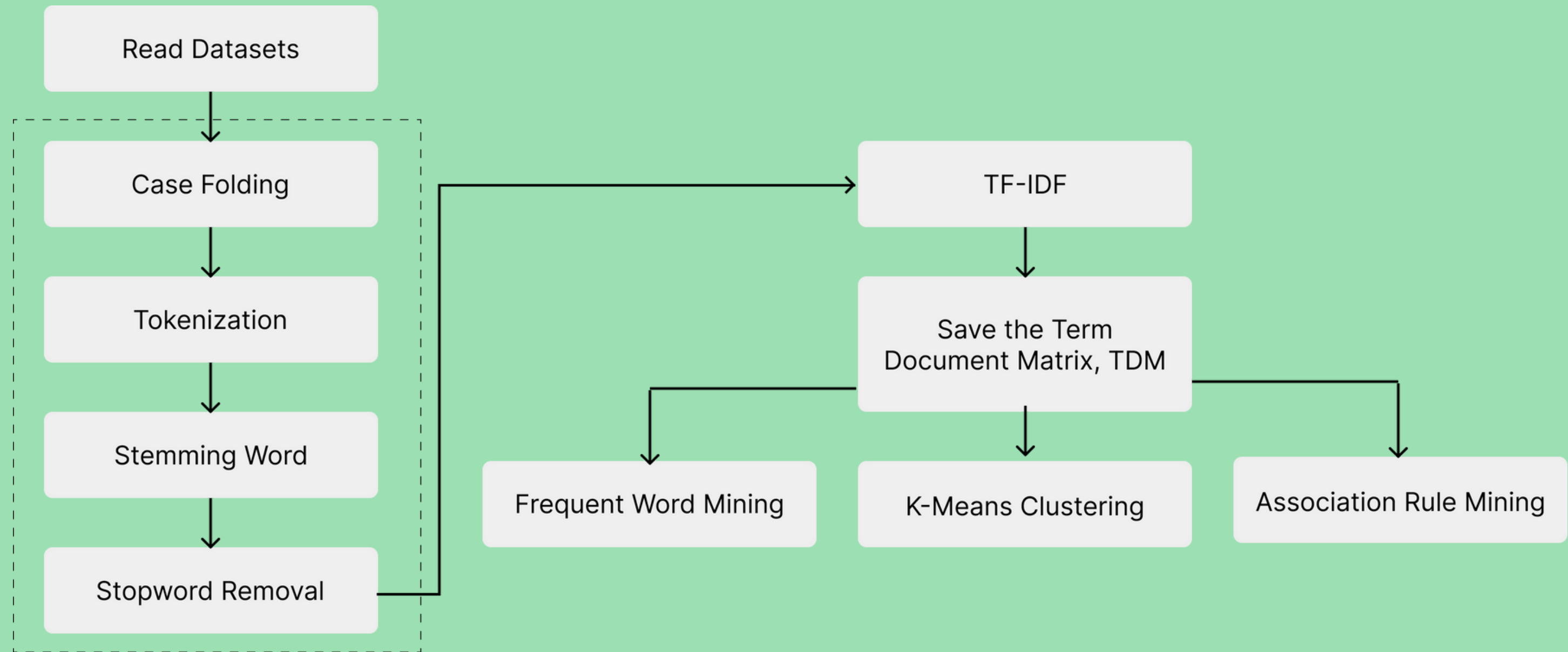


Latar Belakang

Al-Quran merupakan kitab suci bagi umat muslim yang dituliskan menggunakan bahasa Arab. Namun, tidak semua orang muslim bisa berbahasa Arab untuk mengerti maksud dari isi Al-Quran. Oleh karena itu, pada analisis studi kasus kali ini, penulis (jurnal) berusaha menyelesaikan permasalahan tersebut agar lebih efektif dengan cara mengklasifikasikan suatu kata dengan merujuk pada tafsir dan terjemahan.



Metodelogi



Step-Step Penyelesaian Masalah

Step Pertama: Preprocessing

Pada, tahap ini dilakukan pengurangan kata untuk menghasilkan kata yang bermakna dan mencegah adanya redundansi kata.

Step Kedua: Proses TF-IDF

Pada tahap ini dilakukan perhitungan TF-IDF dengan menggunakan rumus :

$$W_{i,j} = t_{fi,j} \times \log(N/df_i)$$

ket:

$t_{fi,j}$ = Jumlah kejadian dalam i dan j

df_i = Jumlah dokumen yang berisi i

N = Total dokumen

Step Ketiga: Most Frequent Word Mining

Tahap ini adalah tahap untuk memvisualisasikan kata dalam bentuk word cloud untuk melihat kemunculan kata yang terbanyak hingga paling sedikit. Sehingga, setelah diperoleh kata yang muncul terbanyak akan mempermudah pengilustrasian hasil.

Step-Step Penyelesaian Masalah

Step Keempat: K-Means clustering

Alogaritma yang digunakan dalam proses ini adalah K-Means Clustering. Pengelompokan dengan menentukan jarak Euclidian antar kata.

Dengan rumus:

$$\begin{aligned} \|A-B\| &= \sqrt{\sum (a_i - b_i)^2} \\ &= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} \\ &= \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \end{aligned}$$

Step Kelima: Association Rules

Association Rules dalam *Text Mining* dilakukan untuk mengetahui hasil kata positif dan negatif yang dihasilkan dari proses-proses sebelumnya. Tujuan pada tahap ini adalah untuk mengetahui korelasi antar kata yang telah dihasilkan serta melihat pola yang terbentuk dalam pemrosesan teks ini.

Hasil: *Preprocessing*

Contoh hasil preprocessing atau seleksi fitur tafsir (a) dan terjemahan (b)

SURAH	PENDAPAT	ULAMA	ABU	AHLI	ALLAH	AYAT
134.71901	134.11199	66.08931	158.08545	98.83268	808.36051	553.46830
MEDINAH	SALAT	HADIS	NABI	ALBUKHARI	MUSLIM	RIWAYAT
50.42218	196.86536	362.28900	527.68617	115.59964	154.48807	218.49652
UMAR	MEKAH	ABBAS	RASULULLAH	AHMAD	HURAIRAH	SAHABAT
75.60364	125.21297	75.93314	393.16143	76.56092	82.72732	63.75376
ALQURAN	MELARANG	MANUSIA	MUHAMMAD	PERINTAH	TUHAN	WAHYU
322.96000	58.01810	437.15789	342.15315	236.57726	332.61349	87.30771

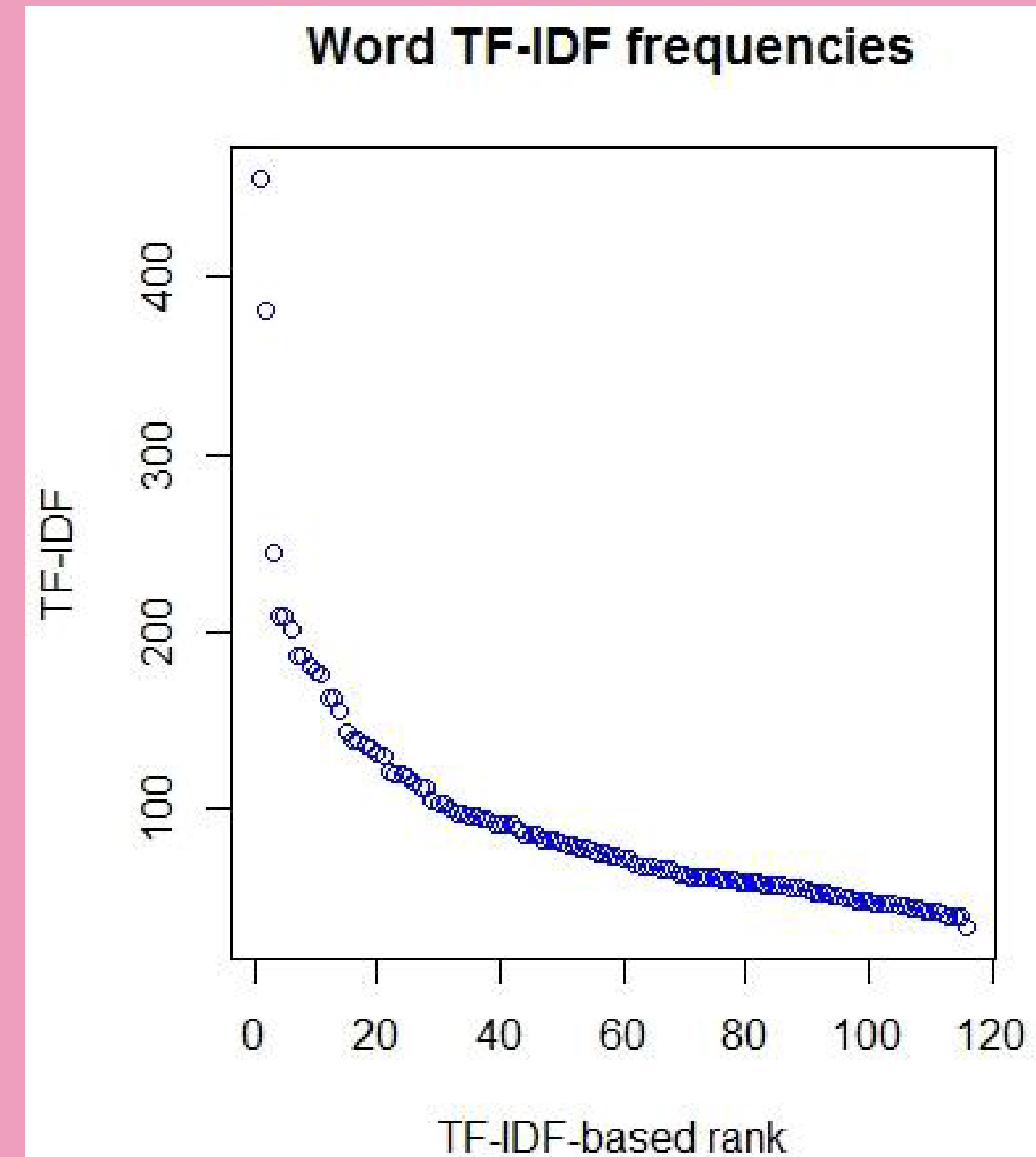
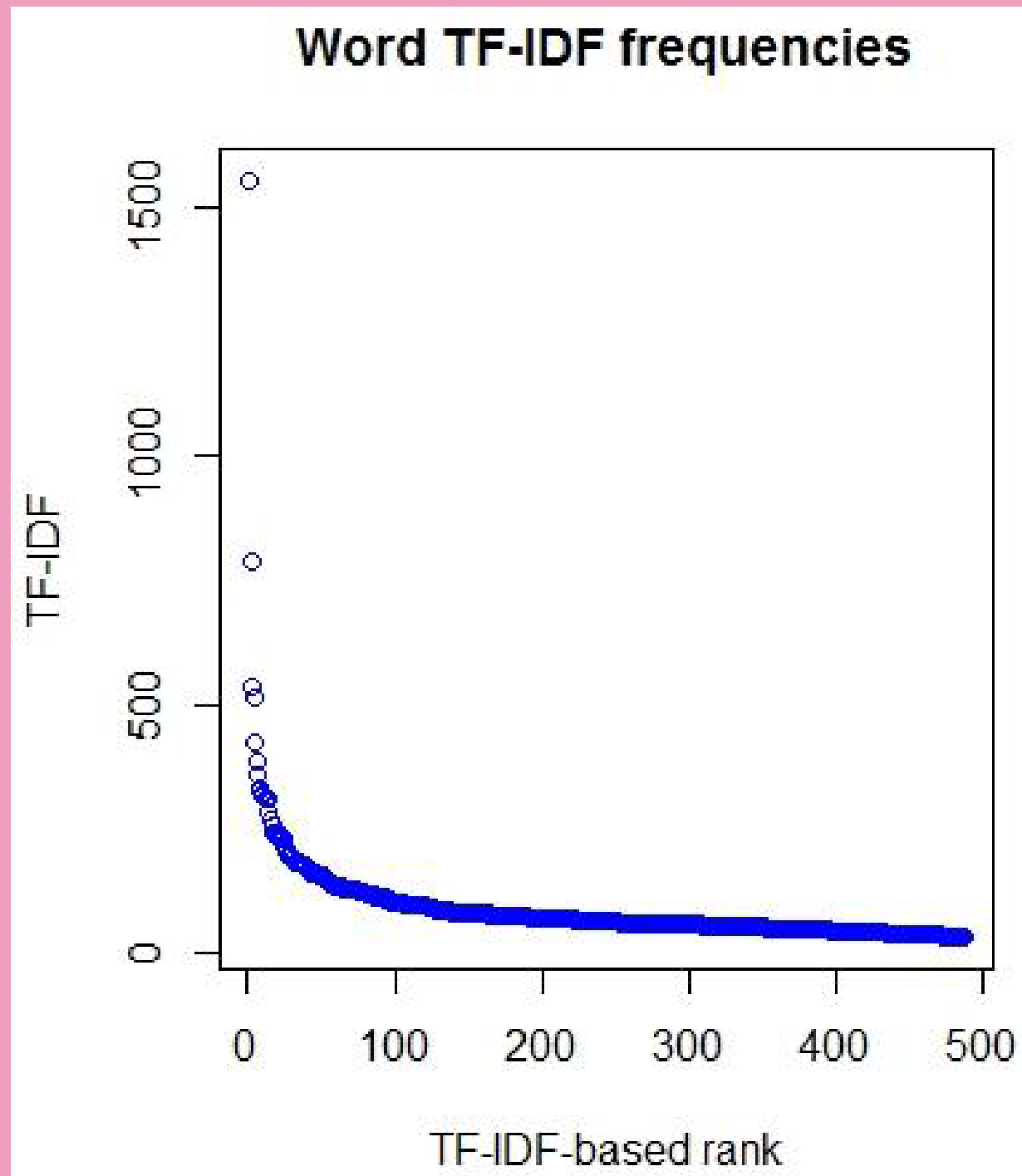
(a)

DOSA	NYATA	BESAR	KIAMAT	KITAB	MENGIKUTI	TAKUT	MENDUSTAKAN	DUNIA
107	139	175	174	249	110	146	138	181
KAFIR	LAKILAKI	RAHMAT	KEHIDUPAN	AIR	MALAM	SURGA	KEKAL	MELIHAT
414	136	118	177	143	116	164	102	232
PAHALA	NERAKA	CIPTA	MATI	REZEKI	TAQWA	BERPALING	MENGATAKAN	KEBAJIKAN
101	274	162	119	109	221	108	115	129
MENDENGAR	HATI	AKHIRAT	ANAK	GOLONGAN	KAUM	FIRAUN	MUSA	NEGERI
119	194	139	150	120	207	111	235	128
SALAT	HARTA	FIRMAN	PERBUATAN	MALAIKAT	IBRAHIM	SETAN	PEREMPUAN	
100	133	154	107	133	126	110	176	

(b)

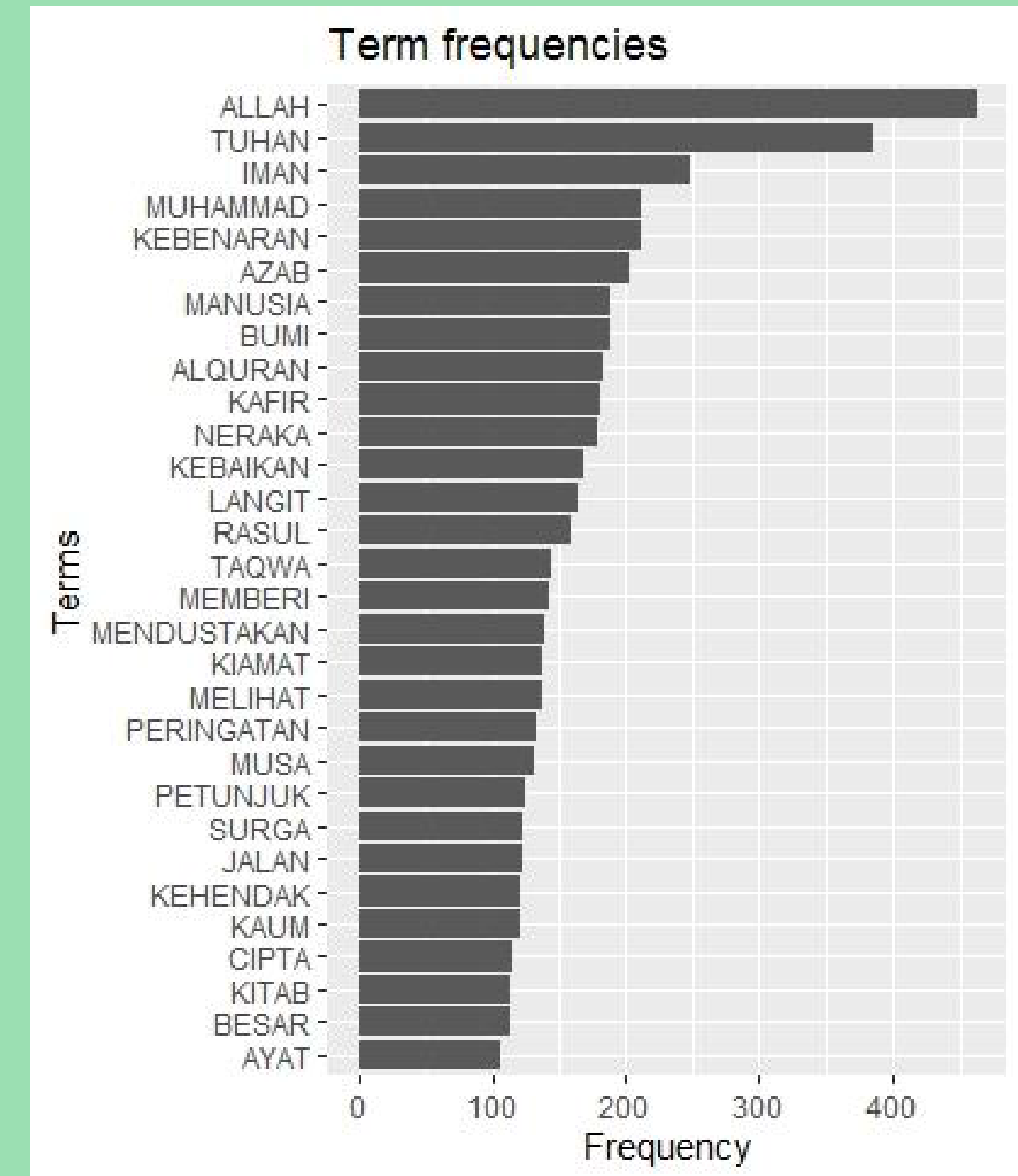
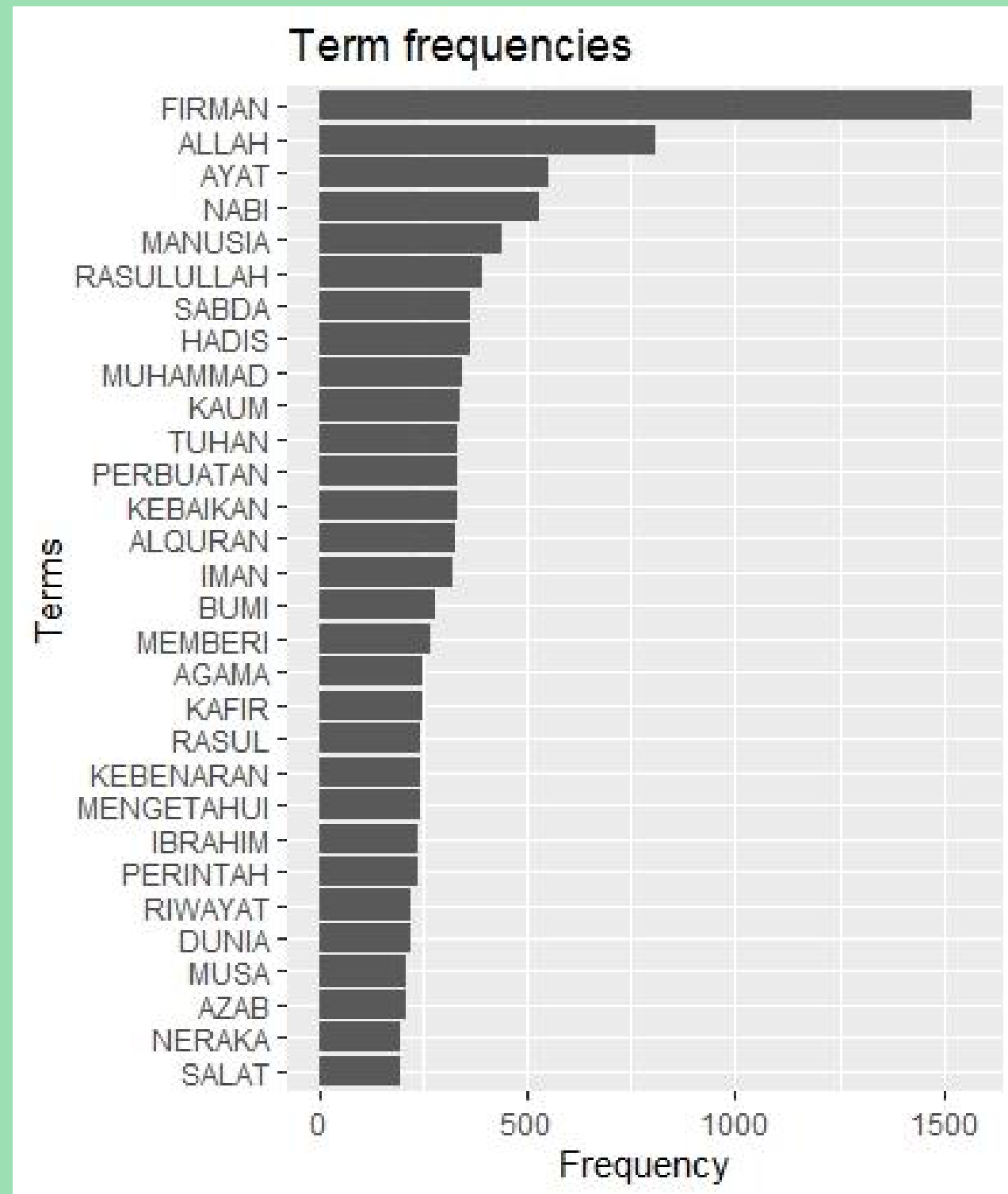
Hasil: TF-IDF

TF-IDF term plot frekuensi tafsir dan terjemahannya



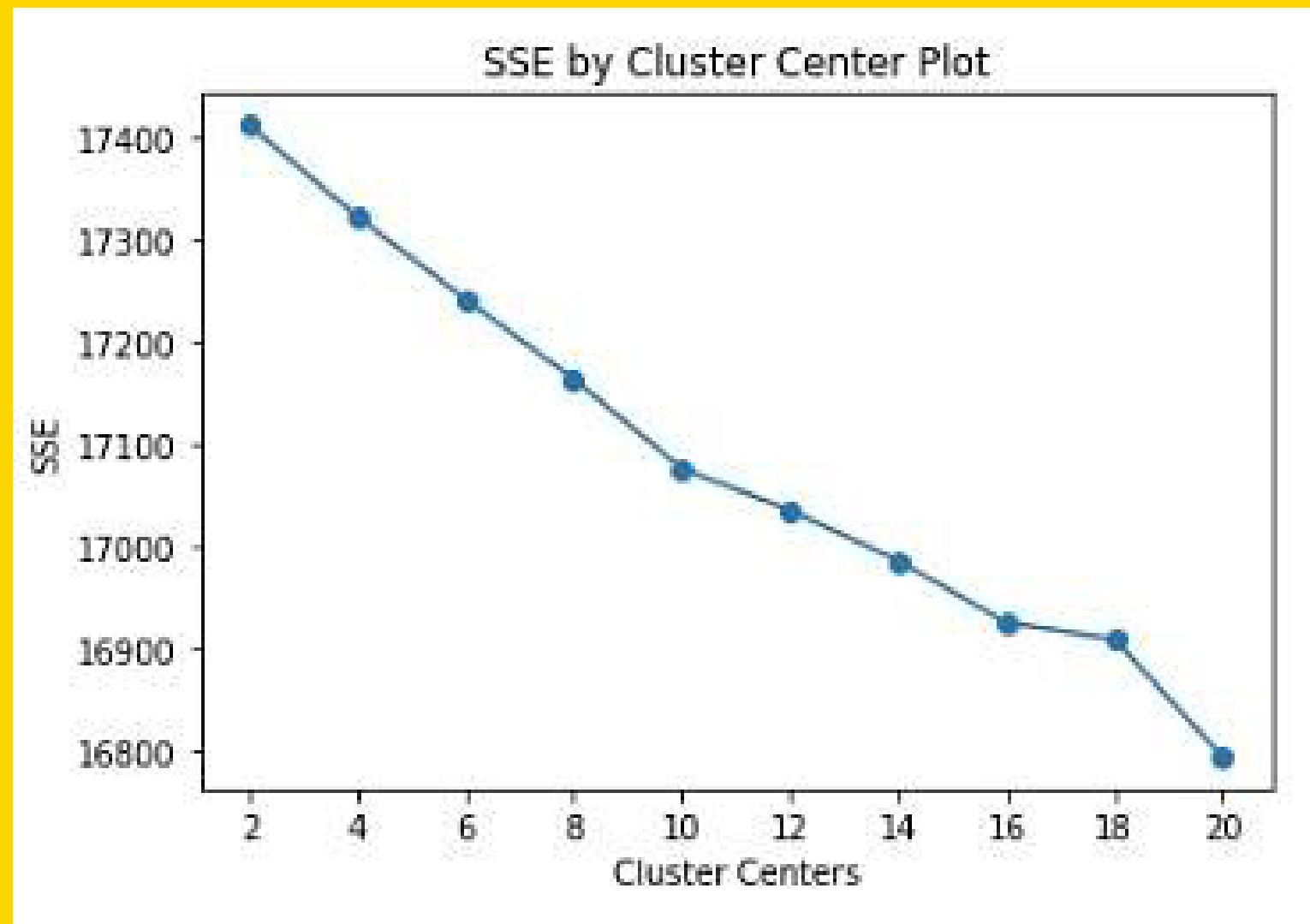
Hasil: *Most Frequent Words*

30 kata yang paling sering muncul dalam tafsir dan terjemahan

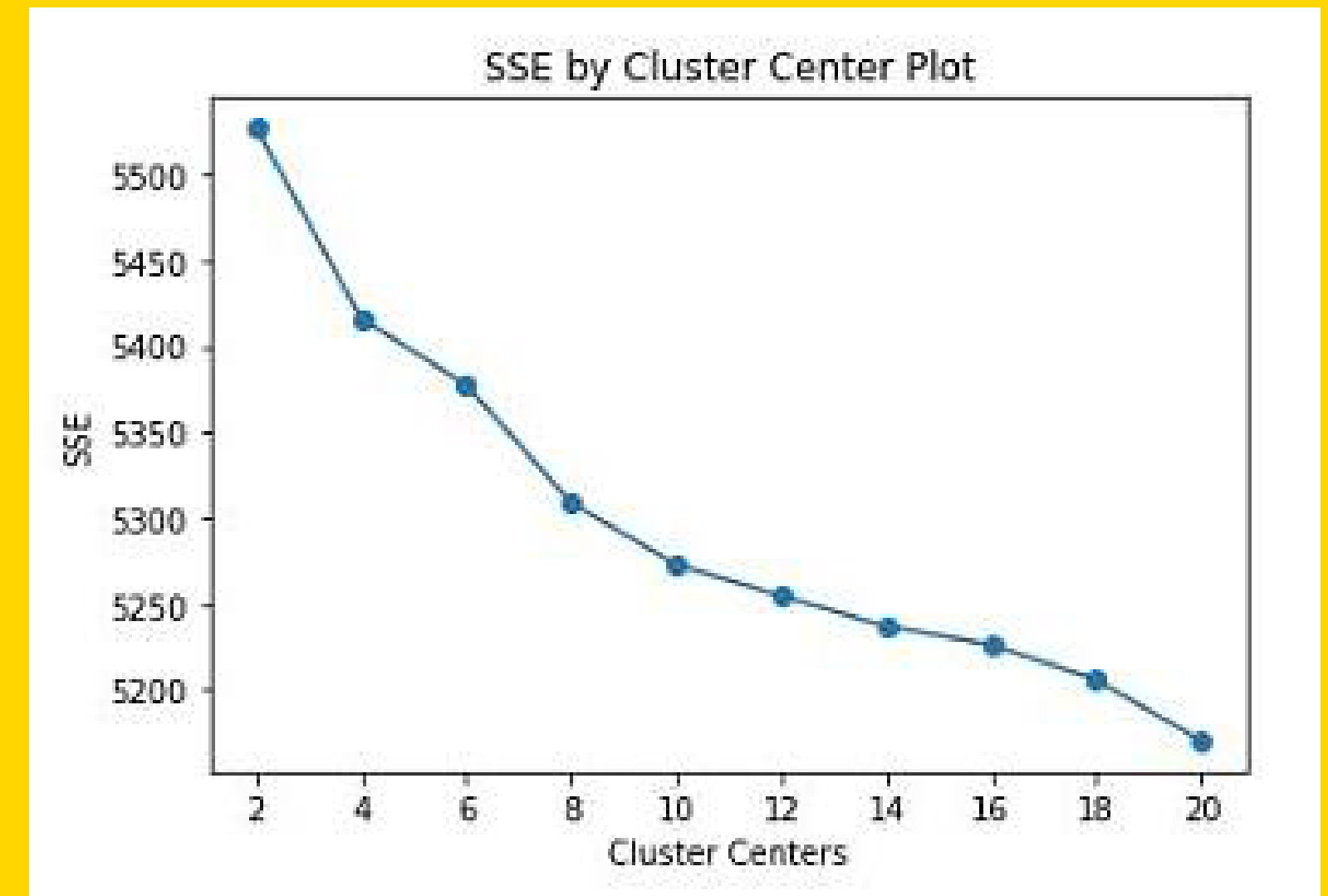


Hasil: *K-Means*

Plot pusat SSE (*Sum of Squared Error*) vs cluster untuk tafseer dan terjemah (Elbow Methods)



tafsir



terjemahan

Hasil: *K-Means*

Tabel hasil clustering dengan K=8

Table 2. Clustering results with K=8 clusters

Cluster	Tafsir	Translation
1	Kitab, Anak, Mukjizat, Sihir, Mesir, Agama, Harun, Tuhan, Firaun, Israil, Bani, Musa	Petunjuk, Rasul, Benar, Mengingkari, Kafir, Azab, Mendustakan, Kitab, Quran.
2	Hati, Peringatan, Ajaran, Hukum, Petunjuk, Agama, Kitab, Quran	Kitab, Kiamat, Peringatan, Rasul, Musa, Kafir, Neraka, Azab.
3	Tanah, Planet, Gunung, Siang, Hujan, Kekuasaan, Bulan, Malam, Benda, Alam, Tanda, Matahari, Air, Menciptakan, Bintang, Mahluk, Malaikat, Langit, Bumi	Baik, Kebesaran, Hujan, Tanda, Air, Gunung, Langit, Bumi
4	Tirmidzi, Imam, Ibnu, Ismail, Ahmad, Bukhari, Hurairah, Abu.	Karunia, Kafir, Hati, Petunjuk, Muhammad, Hamba, Rasul, Beriman.
5	Balasan, Pahala, Kehidupan, Berhala, Dosa, Hamba, Kafir, Nikmat, Neraka, Surga, Amal, Azab, Akhirat, Dunia.	Kenikmatan, Bertaqwa, Kebajikan, Kekal, Petunjuk, Lurus, Sungai, Surga, Manusia.
6	Hati, Peringatan, Ajaran, Hukum, Petunjuk, Agama, Kitab, Quran	Petunjuk, Janji, Jalan, Azab, Firaun, Kitab, Rasul, Muhammad, Tanda, Benar
7	Dosa, Kiamat, Perempuan, Laki, Tempat, Kafir.	Berdoa, Rahmat, Quran, Azab, Pengampun, Pengasih, Penyayang.
8	Istidraj, Istiadat, Israfil, Israil, Istana, Istilah, Isteri, Petunjuk, Malaikat	Beriman, Istri, Azab, Akhirat, Yatim, Dunia, Harta, Nikmat, Perempuan, Laki

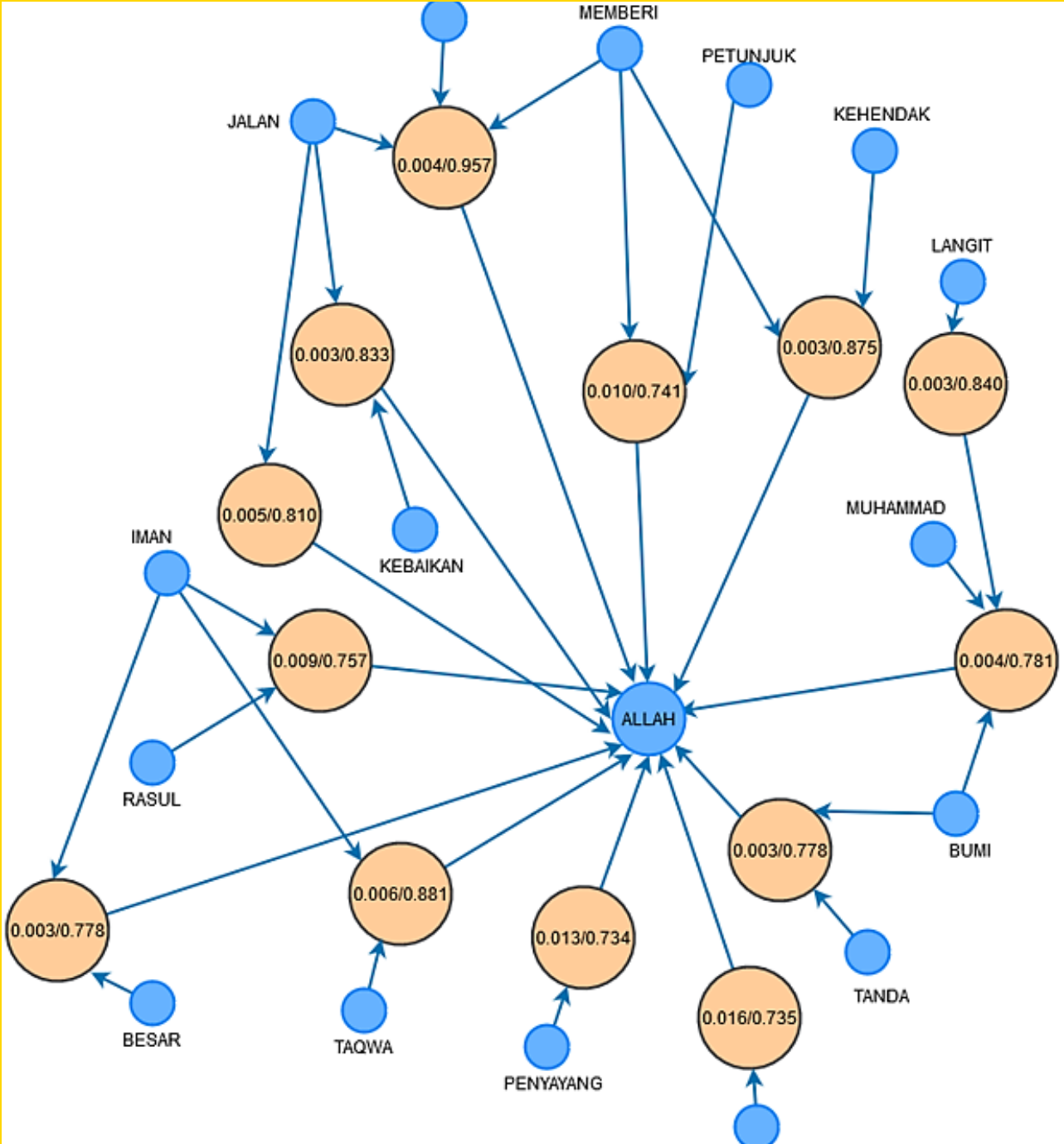
Hasil: *K-Means*

Tabel hasil clustering dengan K=10

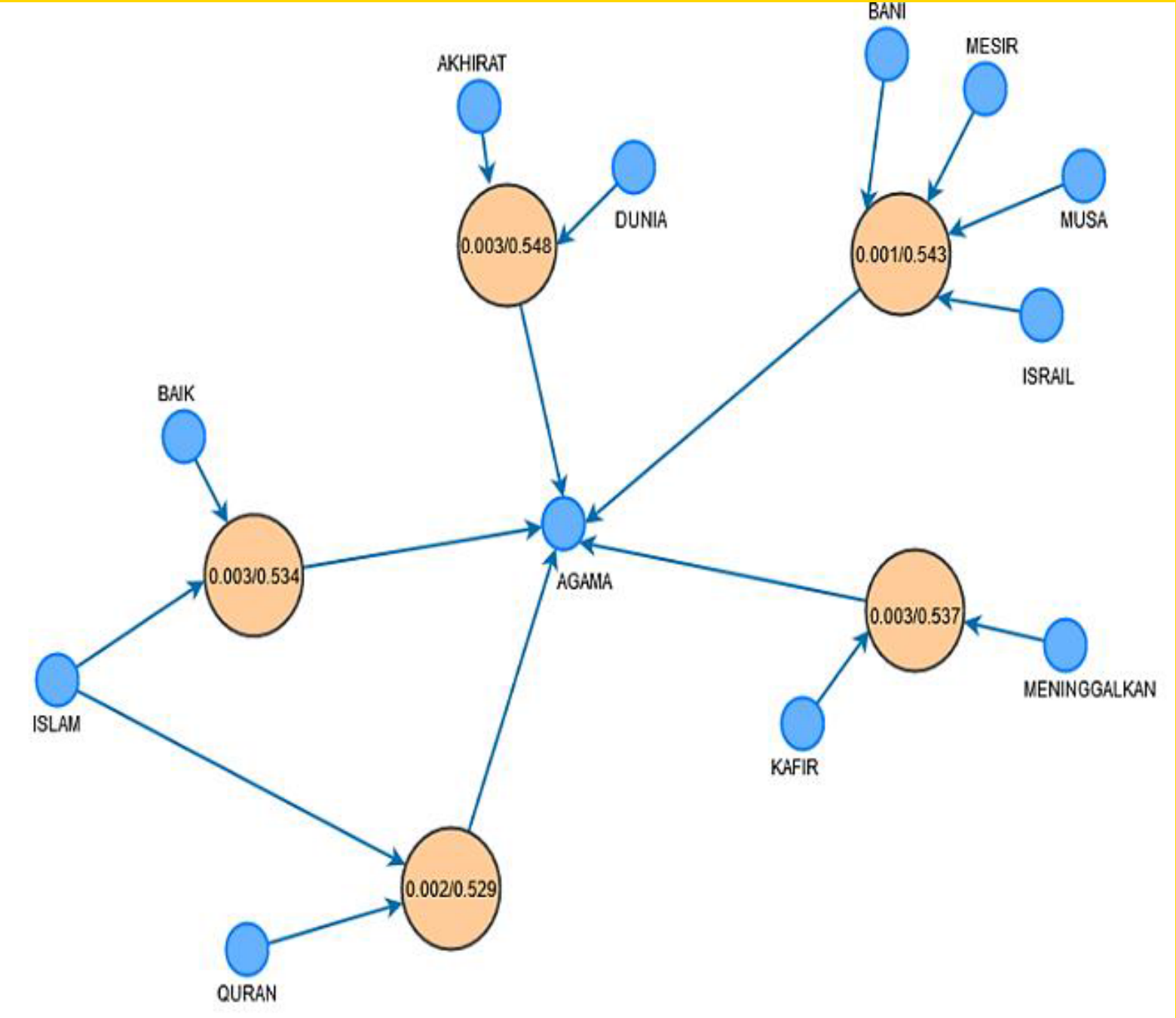
Cluster	Tafseer	Translation
1	Munafik, Larangan, Kemenangan, Musuh, Yahudi, Ibrahim, Mekah, Perang, Musyrik, Kafir.	Hati, Langit, Hamba, Tobat, Bumi, Quran, Muhammad, Beriman
2	Istidraj, Isteri, Israfil, Israil, Istana, Petunjuk, Malaikat	Gembira, Perjalanan, Kafir, Celakalah, Manusia, Beriman, Kebenaran, Peringatan.
3	Negeri, Tanda, Nuh, Setan, Nikmat, Hati, Kiamat, Kafir, Neraka.	Malaikat, Hamba, Benar, Sahaya, Istri, Anak, Perempuan, Laki.
4	Nikmat, Hamba, Quran, Ajaran, Surga, Umat, Baik, Kesenangan, Kebahagiaan, Neraka, Kafir, Kehidupan, Azab, Hidup, Akhirat, Dunia.	Sapi, Takut, Bumi, Malam, Firaun, Harun, Kekuasaan, Kebesaran, Tanda.
5	Isa, Hud, Esa, Musyrik, Sembah, Berhala, Patung, Tuhan.	Diutus, Azab, Umat, Yatim, Nuh, Harta, Anak, Rasul.
6	Dawud, Abdullah, Umar, Tirmidzi, Imam, Ahmad, Ibnu, Bukhari, Muslim, Abu, Hurairah, Sabda.	Mata, Air, Balasan, Baik, Taman, Buah, Penghuni, Kenikmatan, Mengalir, Kekal, Sungai, Surga.
7	Jalan, Kebajikan, Buruk, Balasan, Sifat, Isteri, Ibu, Surga, Saleh, Pahala, Perempuan, Laki, Hamba, Harta, Amal, Dosa, Anak.	Air, Golongan, Dunia, Gunung, Waktu, Negeri, Muhammad, Malaikat, Baik, Kiamat, Manusia, Azab, Neraka.
8	Harun, Hati, Sihir, Umat, Petunjuk, Kaum, Mukjizat, Kebenaran, Taurat, Firaun, Bani, Israil, Kitab, Musa.	Petunjuk, Nikmat, Muhammad, Dustakan, Azab, Kafir.
9	Tumbuhan, Planet, Bulan, Kekuasaan, Tanda, Benda, Tanah, Gunung, Alam, Matahari, Hujan, Cipta, Mahluk, Binatang, Air, Langit, Bumi.	Azab, Kafir, Kerajaan, Tanda, Janji, Rasul, Bumi, Langit, Besar
10	Ibrahim, Menyampaikan, Hamba, Mahluk, Utusan, Laki, Wahyu, Jibril, Lut, Adam, Malaikat.	Puji, Zalim, Disembah, Esa, Langit, Bumi, Azab, Pengasih.

Hasil: *Association Rule*

Hasil dari proses aturan asosiasi pada dataset terjemahan



Hasil dari proses aturan asosiasi pada dataset tafsir



Kesimpulan

Penelitian ini menggunakan 3 metode, yang hasilnya terdapat kata mutual bernilai 0,5306 dengan frekuensi yang sama tinggi. Hasil dataset tafsir lebih baik daripada dataset terjemahan, karena data pada tafsir dapat menyelesaikan dataset yang tidak terstruktur dan bias. Adanya hal ini dapat membantu seseorang membedakan topik berdasarkan informasi yang relevan untuk pencariannya.

TERIMA KASIH!