**PEMODELAN TOPIK TERJEMAH AL-QUR’AN INDONESIA DENGAN MENGGUNAKAN BERTOPIC**

**1Asep Ridwan Hidayat**

1Universitas Pamulang

**Abstrak**

**Kata kunci**: *Cisco Packet Tracer,* *Simulasi jaringan,* *Analisa perangkat jaringan,*

***Abstract***

**Keywords**: *Cisco Packet Tracer, Network simulation, Network device analysis,*

# PENDAHULUAN

Al-quran adalah salah kitab suci orang muslim dengan banyak surat 114 surat dengan jumlah ayat 6127 ayat.

Bertopic modeling adalah Teknik pemodelan topik yang memanfaatkan transormator dan c-ItF-IDF untuk membuat cluster padat yang memungkinkan topik mudah diinterpretasikan sambal memperthankan kata-kata penting dalam deskripsi topik.

Pada pemodelan kali ini selain untuk pemenuhan tugas ujian akhir semerter, sekaligus untuk menambah pemahaman penulis terhadap terjemah ayat-ayat al-qur’an, juga agar menjadi referensi tambahan penerapan topik modeling dengan model BERTopik.

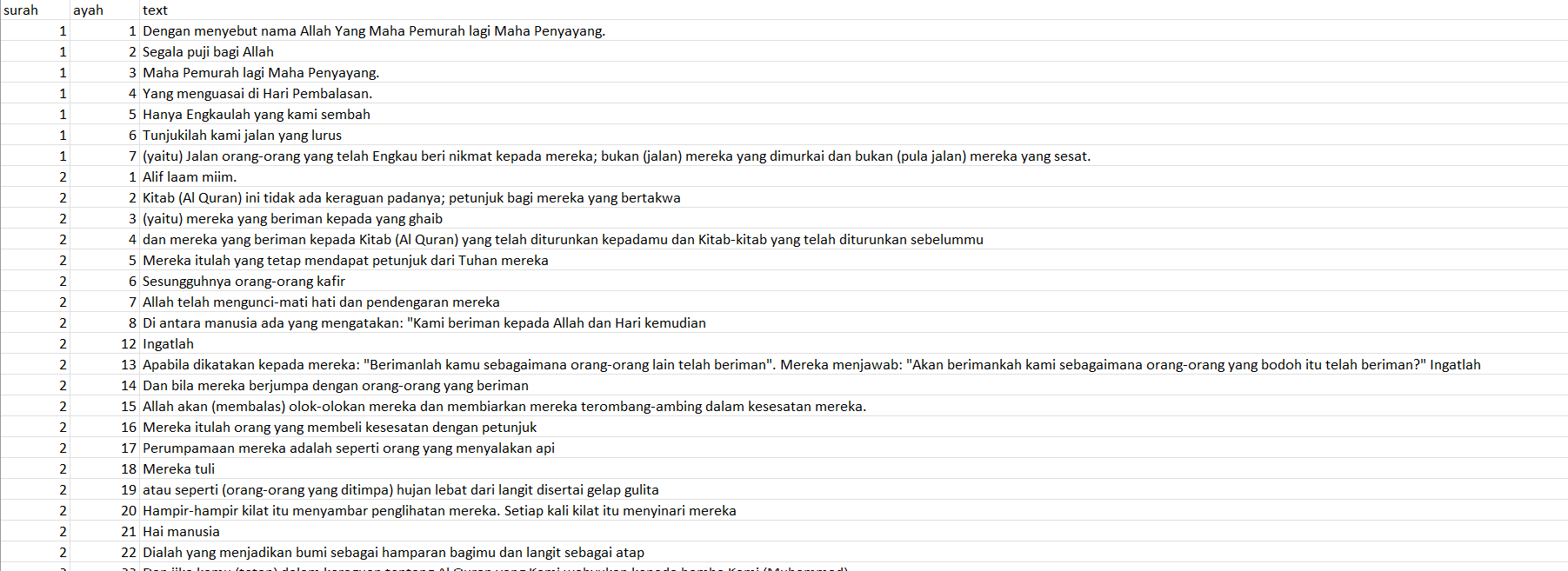
# METODOLOGI

# Pengumpulan data

Dalam penelitian ini dataset yang digunakan adalah Al-Qur’an terjemah bahasa indonesia yang didapat dari quranenc.com.

Selanjutnya setelah pengambilan data, ada beberapa data yang dirapihkan seperti hanya kolom text terjemah bahasa indonesia saja yang digunakan untuk penomoran surat dan ayat tidak diikut sertakan. Setelah pemilihan kolom text dipilih maka dilakukan preprocesing terhadap data.

Berikut capture dataset yang didapat:



Gambar 1.1 Capture Dataset Al-Qur’an

# Untuk pemodelan yang digunakan yaitu pemodelan BERTopik dengan menggunakan pemograman python, library phyton dan tools googlecolab.

# 2.2 Proses dan Preprocesing Teks

Procesing data

Berikut scrypt pegolahan data pada python.

|  |
| --- |
|  |

# Preprocesing Teks

Proses dalam preprocessing data, antara lain sebagai berikut:

1. Remove symbol dan Number

Langkah yang sering dilakukan adalah menghapus simbol dan angka yang terdapat dalam data. Hal ini dilakukan karena simbol dan angka biasanya tidak memiliki makna yang signifikan dalam proses analisis dan dapat mengganggu pemodelan topik atau analisis lainnya. Untuk menghapus simbol dan angka, kita dapat menggunakan fungsi seperti re sub () atau string punctuation pada Python.

berikut script yang digunakan dalam pengolahan data

|  |
| --- |
| text = re.sub(r'\d+', '', text) # Menghapus angka  text = re.sub(r'\W', ' ', text) # Menghapus karakter non-alfanumerik  text = re.sub(r'\s+', ' ', text) # Menghapus spasi berlebih |

1. Lower case

Lower case adalah proses mengubah semua teks menjadi huruf kecil (lower case). Proses ini dilakukan untuk menghindari perbedaan bentuk yang mungkin terjadi akibat adanya huruf besar.

|  |
| --- |
| text = text.lower() # Lower |

1. Tokenizing
2. Remove Stopword
3. Join Case
4. Model Generation and Visualization (Model BERTOPIC)

# KESIMPULAN DAN SARAN

# DAFTAR PUSTAKA

[1] Astuti Indah Kusuma, “Jaringan Komputer”, 2020.

[2] M. Y. M. Simargolang, A. S. M. Widarma And M. D. S. M. Irawan, “Jaringan Komputer”, Yayasan Kita Menulis,

2021.

<https://www.kaggle.com/code/yhirakawa/bertopic-visualization-of-topic-modeling#Bertopic>

https://maartengr.github.io/BERTopic/getting\_started/clustering/clustering.html#agglomerative-clustering