

ABSTRAK

Nama : Suci Eka Safitri

NIM : 1197010072

Judul : Klasterisasi Tema Eskatologi pada Teks Al-Qur'an Berbahasa Indonesia Menggunakan *Hierarchical Agglomerative Clustering* (HAC) Berbasis *Word2Vec*

Eskatologi merupakan tema penting dalam kajian Islam, tetapi kalimat-kalimat yang membahasnya tersebar di berbagai ayat sehingga hubungan semantiknya sulit dipetakan secara manual. Penelitian ini bertujuan membentuk representasi vektor kalimat, mengelompokkan kalimat bertema eskatologi, mengevaluasi kualitas kluster, dan menginterpretasikan tema yang terbentuk. Dari 6.230 ayat, penyaringan berbasis 13 kategori dan 142 kata kunci bahasa Arab menghasilkan 1.348 ayat bertema eskatologi, yang disegmentasi menjadi 2.292 kalimat Kemenag dan 3.170 kalimat Tafsir Jalalayn. Korpus gabungan 5.462 kalimat digunakan untuk melatih model *Word2Vec* berarsitektur *Skip-gram* berdimensi 300 dengan *window* 5 dan *min count* 1. Vektor kalimat Kemenag dibentuk menggunakan SIF, CCR, dan normalisasi norma- L_2 , kemudian dikelompokkan menggunakan HAC dengan *average linkage* dan *cosine distance*. Hasil optimisasi menghasilkan 19 kluster pada *threshold* 0,6392, dengan *Silhouette Score* 0,2375, CCC 0,6479, rata-rata kohesi 0,6112, dan rata-rata *inter-cluster similarity* -0,0383. Tema kluster ditentukan menggunakan TF-IDF sebagai dasar utama, dengan frekuensi kata dan konteks kalimat sebagai pendukung. Tema yang teridentifikasi meliputi kekuasaan Allah, kisah dan tanda kekuasaan-Nya pada alam, kematian dan kehidupan, azab, pahala dan balasan, gambaran surga dan neraka, hari kiamat dan kebangkitan, serta relasi dan ketentuan antarmanusia. Hasil penelitian menunjukkan bahwa kombinasi *Word2Vec*, SIF, CCR, normalisasi L_2 , dan HAC dapat memetakan struktur semantik kalimat bertema eskatologi secara eksploratif.

Kata Kunci: Eskatologi Islam, kalimat, *Word2Vec*, *Smooth Inverse Frequency*, *Common Component Removal*, TF-IDF, *Hierarchical Agglomerative Clustering* (HAC)

ABSTRACT

Name : Suci Eka Safitri

Student ID : 1197010072

Title : *Clustering of Eschatological Themes in Indonesian Qur'anic Text Using Hierarchical Agglomerative Clustering (HAC) Based on Word2Vec*

Eschatology is an important theme in Islamic studies, yet sentences discussing it are scattered across many verses, making their semantic relationships hard to map manually. This study constructs sentence vector representations, clusters eschatology-themed sentences, evaluates cluster quality, and interprets the resulting themes. From 6,230 verses, rule-based filtering using 13 categories and 142 Arabic keywords yielded 1,348 eschatology-themed verses, segmented into 2,292 Kemenag sentences and 3,170 Tafsir Jalalayn sentences. The combined corpus of 5,462 sentences trained a Word2Vec model with a Skip-gram architecture, 300 dimensions, window size 5, and minimum count 1. Kemenag sentence vectors were built using SIF, CCR, and L_2 -norm normalization, then clustered using HAC with average linkage and cosine distance. Optimization produced 19 clusters at a threshold of 0.6392, with a Silhouette Score of 0.2375, CCC of 0.6479, average cohesion of 0.6112, and average inter-cluster similarity of -0.0383 . Cluster themes were determined mainly through TF-IDF, supported by word frequency and sentence context. Identified themes include God's power, signs of His power in nature, death and life, punishment, reward, depictions of heaven and hell, the Day of Judgment and resurrection, and relations among people. The results show that combining Word2Vec, SIF, CCR, L_2 -norm normalization, and HAC can map the semantic structure of eschatology-themed sentences in an exploratory manner.

Keywords: *Islamic Eschatology, sentence, Word2Vec, Smooth Inverse Frequency, Common Component Removal, TF-IDF, Hierarchical Agglomerative Clustering (HAC)*