

ABSTRAK

Nama : Imam Jaenuri

NIM : 1197010033

Judul : Penerapan Algoritma *OPTICS Clustering* Dalam Menangkap Hubungan Semantik Kata Menggunakan Data Al-Qur'an Terjemah Bahasa Indonesia

Perkembangan teknologi mendorong kebutuhan informasi yang relevan. *Data Mining* mampu melakukan ekstraksi informasi tersembunyi dari berbagai jenis data, misalnya penambangan data berbasis teks (*text mining*). Berbagai jenis metode penambangan data teks berfungsi untuk menangkap hubungan semantik kata melalui proses *embedding*. *Word2Vec* merupakan salah satu metode *embedding* yang terbukti efektif dalam analisis similaritas kata. Berdasarkan hal tersebut, penelitian ini dilakukan dengan menggunakan dataset terjemahan Al-Qur'an dalam bahasa Indonesia. Representasi vektor numerik yang dihasilkan oleh *Word2Vec* digunakan dalam proses pengelompokan menggunakan algoritma *OPTICS clustering*. Dengan kemampuannya dalam mengatasi variasi kepadatan data, *OPTICS clustering* cocok digunakan untuk menganalisis Al-Qur'an, yang memiliki banyak kesamaan karakteristik kata. Evaluasi *cluster* dilakukan melalui perhitungan *silhouette score* dan *cosine similarity* untuk melihat kemiripan kata. Sedangkan hasil sebaran *cluster* dianalisis menggunakan *Calinski-Harabasz Index* dan *Density-Based Clustering Validation* untuk menunjukkan tingkat kepadatan yang tinggi. Hasil menunjukkan Algoritma *OPTICS* efektif dalam mengelompokkan kosakata Al-Qur'an dengan kepadatan beragam. *Metric Euclidean* membantu identifikasi jarak, sementara *epsilon* dan *xi* berperan dalam pembentukan *cluster*. *Silhouette score* kurang relevan untuk hubungan semantik, sedangkan *cosine similarity* lebih sesuai. Validasi *clustering* menunjukkan *DBCV* 0,7550 lebih akurat dalam menilai kualitas *cluster OPTICS* dibandingkan *Calinski-Harabasz Index* 19,62.

Kata kunci : *Data Mining, Text Mining, Word Embedding, Word2Vec, OPTICS clustering, Silhouette Score, Calinski-Harabasz Index, Density-Based Clustering Validation.*

ABSTRACT

Name : Imam Jaenuri

NIM : 1197010033

Title : Penerapan Algoritma *OPTICS Clustering* Dalam Menangkap Hubungan Semantik Kata Menggunakan Data Al-Qur'an Terjemah Bahasa Indonesia

The advancement of technology drives the need for relevant information. Data Mining enables the extraction of hidden information from various types of data, including text-based data mining (text mining). Various text mining methods function to capture semantic relationships between words through the embedding process. Word2Vec is one of the embedding methods proven to be effective in word similarity analysis. Based on this, this study utilizes an Indonesian translation dataset of the Al-Qur'an. The numerical vector representations generated by Word2Vec are clustered using the OPTICS clustering algorithm. With its ability to handle variations in data density, OPTICS clustering is well suited for analyzing the Al-Qur'an, which contains many words with similar characteristics. Cluster evaluation is conducted using the silhouette score and cosine similarity to assess word similarity. Meanwhile, cluster distribution is analyzed using the Calinski-Harabasz Index and Density-Based Clustering Validation (DBCV) to measure density levels. The results indicate that the OPTICS algorithm is effective in clustering Al-Qur'an vocabulary with varying densities. The Euclidean metric aids in distance identification, while the epsilon and xi parameters play a crucial role in cluster formation. The silhouette score is less relevant for semantic relationships, whereas cosine similarity is more appropriate. Clustering validation shows that DBCV (0.7550) is more accurate in assessing the quality of OPTICS clusters compared to the Calinski-Harabasz Index (19.62).

Keyword : Data Mining, Text Mining, Word Embedding, Word2Vec, OPTICS clustering, Silhouette Score, Calinski-Harabasz Index, Density-Based Clustering Validation.