

ABSTRAK

Latent Semantic Analysis (LSA) adalah salah satu metode yang populer digunakan dalam tugas klasifikasi. LSA digunakan untuk mengesktrak dan mewakili makna penggunaan kata secara kontekstual dalam dokumen. Umumnya, TF-IDF digunakan sebagai metode untuk membangun matriks istilah – dokumen atau untuk menghasilkan fitur yang digunakan sebelum menerapkan *Singular Value Decomposition (SVD)* di LSA. Berdasarkan percobaan awal, fitur TF-IDF pada LSA belum bekerja dengan baik untuk mengklasifikasikan teks yang terkait erat seperti pada klasifikasi artikel akidah dan ibadah serta klasifikasi artikel nasional dan regional. Hal ini dapat terjadi karena keterbatasan TF-IDF untuk menangkap informasi semantic dalam artikel. Mengacu pada masalah tersebut, penelitian ini berkontribusi dalam penggunaan vektor sebagai representasi kata pada proses klasifikasi teks Bahasa Indonesia dengan menggunakan word2vec yang kemudian digunakan pada metode LSA. Dalam proses klasifikasi antara artikel nasional dan regional, penggunaan fitur word2vec di LSA berhasil meningkatkan nilai f-score dari 74% (LSA dengan TF-IDF) menjadi 75% (LSA dengan word2vec) serta nilai akurasi dari 74% (LSA dengan TF-IDF) menjadi 78% (LSA dengan word2vec), sementara itu dalam klasifikasi artikel akidah dan ibadah, nilai f-score juga berhasil meningkat dari 64% (LSA dengan TF-IDF) menjadi 73% (LSA dengan word2vec), serta nilai akurasi yang meningkat secara signifikan dari 49% (LSA dengan TF-IDF) menjadi 72% (LSA dengan word2vec).

Keywords: Klasifikasi, *Latent Semantic Analysis*, *word2vec*, Artikel Islami Bahasa Indonesia, Artikel Berita Bahasa Indonesia, *Support Vector Machine*