

ABSTRAK

Salah satu diantara teknik *word embedding* adalah model *word2Vec* yang dipopulerkan oleh Google. Hasil dari penerapan model *word2Vec* ini berupa *pre-trained* model yang sering digunakan dalam penyelesaian tugas-tugas terkait bidang NLP. Salah satu penerapannya adalah untuk pembentukan analogi kata dengan menggunakan dataset artikel Wikipedia Bahasa Indonesia. Analogi kata dapat terbentuk dengan melihat nilai vektor yang saling berdekatan atau dapat dikatakan memiliki kesamaan semantik. Hasil dari penelitian ini menunjukkan bahwa untuk kasus analogi kata masih harus diperlukan pengembangan penelitian. Karena hanya beberapa kasus analogi kata yang sudah tepat. Model *word2Vec* pada arsitektur CBOW menunjukkan keunggulannya daripada arsitektur *Skip gram*. Hal ini dibuktikan dengan hasil korelasi semantik bernilai 0.3702604044512485 (perbandingan dengan *gold standard* Wordsim353) daripada *Skip gram* yang bernilai 0.3671614655066779. Begitu pula tingkat *error rate* yang dihasilkan arsitektur CBOW bernilai 0.104719116 (perbandingan dengan SimLex999) sedangkan pada arsitektur *Skip gram* bernilai 0.1773278075.

Kata kunci : analogi kata, gold standard, NLP, semantik, word2Vec, vektor