

## ABSTRAK

Plagiarisme merupakan tindakan mengambil sebagian atau seluruh ide seseorang berupa dokumen maupun teks tanpa mencantumkan sumber pengambilan informasi. Oleh karena itu pendeteksian plagiarisme perlu dilakukan untuk mengurangi penjiplakan terhadap hasil karya orang lain. Penelitian ini bertujuan untuk mendeteksi kemiripan dokumen teks menggunakan metode *Word2vec* dan melakukan percobaan menggunakan TF-IDF untuk mengetahui perbedaan nilai. Dokumen yang digunakan untuk perbandingan teks ini adalah 116 abstrak bahasa Indonesia. Hasil penelitian yaitu saat dilakukan *stemming* nilai kemiripan lebih tinggi rata-rata 5% dari tidak dilakukan proses *stemming*. Menghasilkan nilai similaritas diatas 50% untuk dokumen yang tingkat kemiripannya tinggi. Sedangkan untuk dokumen dengan tingkat kemiripan rendah atau tidak berplagiat menghasilkan nilai similarity dibawah 30% Tahapan pembuatan adalah *preprocessing* yang terdiri dari *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Setelah proses *preprocessing* maka tahap selanjutnya dilakukan pembobotan TF-IDF dan *Word2vec*. Selanjutnya nilai kemiripan menggunakan *Cosine Similarity* sehingga mendapatkan nilai persentase kemiripan. Berdasarkan hasil percobaan jika menggunakan *Word2vec* nilai kemiripan yang dihasilkan lebih tinggi rata-rata 28% dibandingkan dengan nilai pembobotan TF-IDF.

Kata kunci: *Cosine Similarity*, Dokumen, plagiarisme, *preprocessing*, TF-IDF, *Word2vec*