

## **Abstrak**

Makalah ini membahas perhitungan kesamaan antara kata-kata dalam bahasa Indonesia dengan menggunakan teknik representasi Word2Vec. Word2Vec adalah model yang digunakan untuk merepresentasikan kata ke dalam bentuk vektor. Model dalam percobaan ini dibentuk menggunakan corpus Wikipedia bahasa Indonesia 4GB dan kemudian menggunakan metode perhitungan kemiripan cosinus untuk menentukan nilai kemiripannya. Model ini kemudian diuji dengan set standar emas WordSim-353 dan SimLex-999 yang telah diberi label dengan nilai kesamaan menurut peringkat manusia. Untuk mengetahui akurasi korelasinya menggunakan korelasi Pearson. Hasil korelasi dari penelitian ini adalah 0,5663 untuk data tes WordSim-353 menggunakan ukuran jendela 14 dan dimensi vektor 150, dan 0,3472 untuk data uji SimLex-999 menggunakan ukuran jendela 2 dan dimensi 300. Hasil percobaan menunjukkan bahwa korelasi antara standar emas dan teknik sistem masih relatif lemah.