

# BAB I PENDAHULUAN

Dalam mencari nilai kesamaan dan keterkaitan perlu diketahui informasi berupa makna dari kata dalam bahasa. Dimana makna sangat bersifat arbitrer (berubah-ubah, tidak tetap), berbeda dengan morfem atau kata, sebagai sasaran dalam studi morfologi yang strukturnya tampak jelas dan dapat disegmen-segmenkan [1]. Oleh karena itu makna yang terkandung pada suatu bahasa, kode, atau jenis representasi lain yang dapat diketahui dan dipelajari oleh cabang ilmu linguistik [4].

Di era teknologi yang telah maju seperti sekarang ini, pemrosesan bahasa khususnya mencari nilai kesamaan dan keterkaitan kata telah banyak diminati. Namun, pemrosesan bahasa ini memerlukan komputasi pengolahan data yang cukup berat dikarenakan data teks yang tidak terstruktur serta mempunyai informasi yang kaya, dalam arti mempunyai fitur yang banyak dan berdimensi tinggi [13]. Bukan hanya itu, nilai untuk kesamaan dan keterkaitan suatu pasang kata atau analogi akan beragam karena dinilai dari sudut pandang yang berbeda dimana analogi kata merupakan cara untuk menunjukkan dua situasi yang didalamnya terdapat struktur relasional. Selain itu analogi kata memerlukan kemampuan kognitif yang lebih sedikit dan dapat digunakan pada bidang ekonomi, mengukur keterampilan untuk keberhasilan di perguruan tinggi, sekolah pasca sarjana dan bekerja [7]. Walaupun secara manual, analogi sebuah kata dapat dicari dengan menggunakan kamus sebagai informasi [2] namun perlu adanya sebuah penelitian untuk membuktikan bahwa analogi kata dapat dicari dengan sebuah sistem dengan metode tertentu.

Oleh karena itu dalam penelitian ini dibangun sebuah model untuk mengetahui nilai analogi kata dalam bahasa. Model yang dibangun yaitu sebuah word2vec untuk merubah suatu kata lalu digambar dengan vector (sesuatu yang berarah dalam dimensi). Representasi dari word2vec ini dapat menemukan arti pada kata dengan menyamakan arah dan panjang vektor. Model ini telah banyak digunakan untuk mempresentasikan kata dalam penelitian terbaru di bidang linguistik komputasi [6].

## 1.1 Latar Belakang

Semantik adalah cabang ilmu linguistik yang mempelajari arti atau makna dalam suatu bahasa. Semantik mempelajari arti atau makna yang terkandung pada suatu Bahasa, kode, atau jenis representasi lain [4]. Penilaian untuk semantik pada umumnya disebut *gold standart* yaitu nilai yang diberikan ahli pada bidangnya.

Nilai untuk kesamaan dan keterkaitan suatu pasang kata akan beragam karena dinilai dari sudut pandang berbeda. Nilai yang diberikan dapat dibantu dengan adanya kamus sebagai informasi dari kata yang akan dinilai [2]. Informasi yang didapat dari kamus akan berupa makna dan perlu dipelajari dengan salah cabang ilmu linguistik (ilmu bahasa) yaitu semantik.

Model word2vec merubah suatu kata lalu digambar dengan vector (sesuatu yang berarah dalam dimensi). Dengan representasi dari word2vec dapat menemukan arti pada kata dengan menyamakan arah dan panjang vektor. Metode ini banyak digunakan untuk mempresentasikan kata dalam penelitian terbaru di bidang linguistik komputasi atau text mining [6]. Sehingga metode ini perlu di eksplorasi agar dapat mencapai nilai akurasi dan kecepatan yang maksimal.

## 1.2 Topik dan Batasannya

Topik yang dibahas adalah sistem yang dapat menyelesaikan soal analogi yang terdapat pada tes psikotes dengan batasan sebagai berikut :

1. Analisis efektifitas Word2Vec
2. Korpus yang digunakan bahasa Indonesia

## 1.3 Tujuan

Adapun tujuan yang ingin dicapai pada penelitian ini adalah analisis konfigurasi *windows size* dan dimensi pada model Word2vec dalam menyelesaikan soal analogi

## 1.4 Organisasi Tulisan

Pada jurnal ini mencakup pendahuluan, studi terkait, sistem yang dibangun, evaluasi, dan kesimpulan. Pada bagian pendahuluan terdapat latar belakang, topik dan batasannya serta tujuan dari penelitian ini. Bagian studi terkait berisi tentang materi yang mendukung pada jurnal ini. Bagian selanjutnya berisi tentang sistem yang dibuat untuk mencapai tujuan jurnal ini. Adapun hasil dari sistem yang dibangun akan dibahas pada bagian evaluasi. Pada bagian evaluasi akan terdapat dua sub-bagian yaitu hasil pengujian dan analisis hasil pengujian. Untuk bagian kesimpulan bersisi tentang pekerjaan selanjutnya untuk meningkatkan capaian dari jurnal ini.