

BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurut Abriani, G. U., & Yaqin, M. A. (2019), kalimat terdiri dari serangkaian kata dengan beberapa jenis struktur dasar yang membentuknya. Dalam menentukan nama sebuah kalimat, diperlukan setidaknya susunan struktur dasar yang terdiri dari *noun* dan *verb*. Sementara itu, kesamaan makna kalimat merupakan hasil dari kesamaan antara kata-kata yang digunakan. Untuk mengetahui seberapa dekat makna antara dua kalimat, dapat dilakukan dengan menghitung nilai kesamaan makna antara keduanya.

Memahami makna antar kalimat memang merupakan tantangan yang kompleks dan kontroversial dalam teori bahasa. Meskipun makna berkaitan dengan bahasa, namun memiliki hubungan erat dengan kehidupan manusia. Keberadaannya sangat terikat dan terkait erat dengan kompleksitas dan luasnya berbagai aspek kehidupan manusia (Ramadani, F. 2020).

Pemahaman antara kalimat sering ditemui dalam pengerjaan skripsi. Salah satu permasalahan yang sering dihadapi mahasiswa dalam menentukan kemiripan antara dokumen teks. Mahasiswa perlu membandingkan beberapa dokumen dan mengidentifikasi kemiripan pada tingkat kata, kalimat, atau paragraf. Pendekatan ini terbukti tidak efektif dan efisien. Untuk mengembangkan aplikasi pendeteksian kemiripan pada dokumen teks, diperlukan pemahaman tentang pengolahan teks. *Text Mining* adalah penerapan konsep data *mining* untuk mencari pola dalam teks. Proses analisis

teks ini digunakan untuk menemukan informasi yang relevan dengan tujuan tertentu. *Text mining* merupakan teknik penambangan data yang fokus pada teks, biasanya didapatkan dari dokumen, dengan tujuan menganalisis isi dokumen tersebut. Secara lebih spesifik, text mining dapat diartikan sebagai penggunaan konsep dari teknik data mining untuk menemukan pola dalam teks guna mencari informasi yang berguna untuk tujuan tertentu (Fauzi, A., 2022).

Penelitian lain telah menyimpulkan bahwa Word Embedding merupakan bagian dari NLP (*Natural Language Processing*) yang baru-baru ini telah menjadi topik menarik untuk penelitian. Salah satu metode word embedding yang diusulkan oleh *Facebook AI research* adalah *FastText*, yang memiliki keunggulan dalam menemukan kata-kata yang jarang ditemui atau OOV (*Out of Vocab*). Dalam penelitian ini, dilakukan proses pengujian menggunakan 500 data, di mana 75% data digunakan sebagai data training (375 data) dan 25% data digunakan sebagai data testing (125 data). Untuk membandingkan nilai Y aktual dengan Y prediksi, digunakan metode MAPE (*Mean Absolute Percentage Error*). Hasil penelitian menunjukkan bahwa terdapat kombinasi terbaik dengan nilai MAPE terbaik pada proses training dan testing, yaitu masing-masing 5% dan 8%. Hal ini mengindikasikan bahwa model yang dihasilkan menggunakan algoritma *backpropagation* untuk melakukan penilaian/koreksi jawaban secara otomatis memiliki kualitas yang sangat baik (Sani, D. A., & Sarwani, M. Z., 2022).

Dan pada penelitian “Rancang Bangun Aplikasi Pengecekan Kemiripan Judul Skripsi Dengan Metode *Cosine Similarity*” oleh Apriyanto, I. A., & Aribowo, E. (2018) menyatakan bahwa Metode *Cosine Similarity* merupakan metode yang dapat melakukan pendeteksian secara cepat dan tepat, *cosine similarity* juga merupakan salah satu metode yang dapat digunakan untuk mengukur tingkat kemiripan dokumen.

Berdasarkan beberapa referensi diatas dapat diperoleh data bahwa metode *Cosine Similarity* mempunyai kemampuan untuk mendeteksi dua kalimat secara cepat, tepat dan metode tersebut sangat cocok untuk digunakan pengecekan kalimat referensi kesamaan makna. Oleh karena itu, penelitian ini berjudul “Pengecekan Kesamaan Makna Kalimat Dari Referensi Jurnal dengan Metode *Cosine Similarity*”.

1.2 Rumusan Masalah

Berdasarkan latar belakang sebelumnya, perumusan masalah pada penelitian ini adalah sebagai berikut :

1. Bagaimana melakukan pengecekan kalimat referensi berdasarkan kesamaan makna?
2. Bagaimana implementasi metode *Cosine Similarity* untuk pengecekan kalimat referensi berdasarkan kesamaan makna?

1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. *Dataset* yang di gunakan penelitian ini berfokus pada jurnal (Abstrak dan Pendahuluan).
2. Fokus ditujukan untuk melakukan pengecekan kalimat referensi kesamaan makna dan hasil akhir dari *Cosine Similarity*.
3. Pengecekan kalimat referensi kesamaan makna hanya mencocokkan kalimat yang sama.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Untuk melakukan pengecekan kalimat referensi berdasarkan kesamaan makna.
2. Untuk mengetahui implementasi metode *Cosine Similarity* dalam pengecekan kalimat referensi berdasarkan kesamaan makna.

1.5 Manfaat Penelitian

Adapun manfaat yang diharapkan oleh penulis dari penelitian ini adalah sebagai berikut :

1. Bagi Penulis
 - a. Dapat memahami proses pengecekan kalimat referensi kesamaan makna dengan metode *Cosine Similarity*.

- b. Dapat menjadikan bahan referensi untuk penelitian berikutnya yang berkaitan dengan pengecekan kalimat referensi kesamaan makna ataupun penelitian tentang metode *Cosine Similarity*.
2. Bagi Universitas Sebagai bahan referensi yang dapat digunakan oleh pihak yang membutuhkan serta mampu menghasilkan calon-calon sarjana yang handal khususnya dalam perkembangnya teknologi informasi.

1.6 Sistematika Penulisan

Sistematika penulisan pada penelitian ini disusun untuk memudahkan pembaca dalam memahami pembahasan yang ada dalam penulisan skripsi. Dalam sistematika penulisan skripsi Fakultas Teknologi Informasi Universitas Merdeka Pasuruan, terdapat lima bab penulisan yaitu :

BAB I PENDAHULUAN

Pada bab pendahuluan berisi tentang apa saja yang menjadi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan pada skripsi.

BAB II TINJUAN PUSTAKA

Pada bab dua akan membahas mengenai teori-teori yang diperoleh dari literatur yang menjadi dasar dari penulisan dan penyelesaian masalah yang diambil dalam penyusunan tugas akhir.

BAB III METODOLOGI PENELITIAN

Pada bab metodologi penelitian membahas tentang teori dan konsep rancangan yang berhubungan dengan penelitian yang dilakukan dan mendukung dalam pemecahan permasalahan yang diambil.

BAB IV HASIL DAN PEMBAHASAN

Pada bab empat menjelaskan mengenai penerpan serta hasil dari pengujian yang sudah dilakukan terhadap suatu kalimat referensi dengan menggunakan metode yang telah ditentukan sebelumnya.

BAB V PENUTUP

Pada bab penutup akan membahas tentang kesimpulan dari hasil Analisa data yang didapatkan dari penelitian ini, serta berikan saran-saran yang bisa digunakan untuk penelitian selanjutnya.

