

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pesatnya perkembangan teknologi saat ini membuat proses-proses yang sebelumnya konvensional menjadi digital. Salah satunya adalah proses pencatatan katalog perpustakaan atau Online Public Access Catalog (OPAC). OPAC bisa diakses dan digunakan kapan dan dimana saja, hal ini memudahkan pengguna untuk mencari buku-buku yang terdapat pada koleksi suatu perpustakaan (Azzahra & Ramadhani, 2020). Namun hasil pencarian yang dihasilkan oleh katalog online tidak selalu memberikan hasil yang optimal. Hal ini disebabkan oleh beberapa hal mendasar, diantaranya adalah, pada proses pencarian pengguna tidak selalu menggunakan kata kunci yang sama dengan kata kunci pada dokumen yang tersimpan (Silva-Fuentes dkk., 2019), atau kata kunci yang pengguna gunakan sangat kompleks (Savitri dkk., 2021).

Saat ini, proses pencarian data pada basis data, khususnya basis data berbasis *Structured Query Language* (SQL) dilakukan dengan pendekatan pencocokan kata (*string matching*). Algoritma *string matching* adalah algoritma yang digunakan untuk mengidentifikasi apakah satu atau lebih bagian kata terdapat pada kumpulan kata-kata serta menemukan lokasinya dimana (Alqahtani dkk., 2021).

Metode-metode untuk melakukan pencocokan kata seperti *Exact String Matching*, *Approximate String Matching*, ataupun *Hybrid String Matching* banyak diusulkan oleh peneliti untuk mendapatkan hasil pencarian yang optimal (Hakak dkk., 2019). Metode-metode tersebut akan memberikan hasil yang optimal ketika hasil pencarian yang diinginkan hanya pada tingkat sintaksis saja tetapi tidak pada tingkat semantik. Hal ini menjadi alasan kenapa hasil pencarian yang diberikan tidak optimal ketika kata kunci pencarian yang digunakan kompleks atau ketika kata kunci yang digunakan tidak terdapat pada dokumen yang disimpan. Salah satu pendekatan yang bisa digunakan untuk hasil pencarian menjadi lebih optimal

adalah dengan melakukan pencocokan kata tidak hanya pada tingkat sintaksis saja, tetapi pencocokan kata juga dilakukan pada tingkat semantik.

Word embedding adalah teknik Natural Language Processing (NLP) untuk merepresentasikan kata-kata ke dalam bentuk vektor. Vektor hasil word embedding juga dipanggil dengan semantic vector space karena ia bisa merepresentasikan kata baik dari sisi semantik (makna) maupun dari sisi sintaksis (struktur) (Al-Saqqa & Awajan, 2019). Vektor hasil word embedding membuka kemungkinan untuk bisa melakukan operasi-operasi NLP pada kata-kata tersebut. Salah satunya adalah analisis semantik (Wang dkk., 2020).

Analisis semantik pada NLP adalah proses untuk mengevaluasi dan mewakili bahasa natural dengan interpretasi yang mirip dengan manusia (Salloum dkk., 2020). Salah satu algoritma word embedding yang terbukti efisien merepresentasikan vektor dari kata baik dari sisi sintaksis maupun semantik adalah word2vec (Mikolov dkk., 2013).

Word2vec adalah teknik word embedding yang menggunakan model jaringan syaraf tiruan untuk merepresentasikan kata kedalam bentuk vektor . Secara umum word2vec mempelajari makna semantik dari kata berdasarkan konteks kata. Sehingga, kata-kata yang ada pada konteks yang sama dianggap berhubungan (Al-Saqqa & Awajan, 2019). Berdasarkan uraian sebelumnya, penulis mengusulkan penelitian dengan judul "Implementasi Teknik Word Embedding Untuk Rekomendasi Hasil Pencarian Katalog Online Menggunakan Algoritma Word2Vec".

1.3 Rumusan Masalah

Adapun rumusan masalah pada penelitian ini adalah bagaimana mengaplikasikan teknik word embedding untuk rekomendasi hasil pencarian katalog online menggunakan algoritma word2vec.

1.4 Batasan Masalah

Penelitian ini memiliki beberapa batasan masalah, diantaranya adalah:

- a. Algoritma yang digunakan untuk merepresentasikan vektor dari kata hanya algoritma word2vec.
- b. Penentuan kata dengan konteks yang sama atau memiliki hubungan semantic berdasarkan teori *distributional hypothesis* yang algoritma *word2vec* gunakan.
- c. Penelitian ini bertujuan untuk mengimprovisasi hasil pencarian dengan memanfaatkan algoritma word2vec.
- d. Pengujian hasil pencarian dilakukan melalui uji presisi pada k *document* yaitu dengan melakukan pembagian total hasil pencarian yang relevan dengan total k hasil pencarian.
- e. Algoritma yang digunakan untuk menghitung *semantic similarity* antar vektor adalah *cosine similarity*.
- f. Data yang digunakan adalah data judul buku yang tersimpan pada katalog online (OPAC) Universitas Maritim Raja Ali Haji.
- g. Pendekatan word2vec yang digunakan pada proses pembelajaran adalah pendekatan *Skip-Gram* (SG).
- h. Pengujian rekomendasi hasil pencarian dilakukan uji presisi pada k *document* (P@k).

1.5 Tujuan Penelitian

Adapaun tujuan penelitian ini adalah untuk mengaplikasikan teknik *word embedding* supaya menghasilkan rekomendasi dari pencarian katalog online (OPAC) dengan menggunakan algoritma word2vec.

1.6 Manfaat Penelitian

Penelitian ini diharapkan dapat membantu penulis sendiri dan penelitian-penelitian yang berhubungan selanjutnya untuk lebih memahami tentang pengaplikasian teknik *word embedding* melalui algoritma *word2vec* untuk merekomendasikan hasil pencarian katalog online serta diharapkan juga dapat membantu membuat hasil pencarian katalog online Universitas Maritim Raja Ali Haji menjadi lebih variatif melalui hasil rekomendasi.

1.7 Sistematika Penulisan

Untuk mengetahui gambaran isi penelitian ini, maka penulis membuat sistematika penulisan skripsi. Secara garis besar penulisan skripsi ini terdiri dari tiga bagian, diantaranya adalah:

- a. Bagian awal yang berisi dari halaman judul, pernyataan, persetujuan, pengesahan, motto dan persembahan, abstrak, kata pengantar, daftar isi, daftar tabel, daftar gambar, dan daftar lampiran.
- b. Bagian utama skripsi terdiri dari lima bab, yaitu:
 1. BAB I PENDAHULUAN: Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penulisan, dan sistematika penulisan dalam pembuatan skripsi.
 2. BAB II TINJAUAN PUSTAKA: Bab ini menjelaskan tentang dasar teori yang berhubungan dengan topik skripsi. Dasar teori yang digunakan dalam penyusunan skripsi ini meliputi, *word embedding*, *word2vec* dengan pendekatan *skip gram*, *cosine similarity*, dan uji presisi pada *k document*.
 3. BAB III METODE PENELITIAN: Bab ini menjelaskan tentang dasar alur dan metode penelitian pada skripsi, data yang digunakan serta perancangan pembelajaran algoritma *word2vec* dan perancangan sistem rekomendasi hasil pencarian.
 4. BAB IV HASIL DAN PEMBAHASAN: Bab ini menjelaskan tentang pengujian sistem rekomendasi hasil pencarian serta hasil dari evaluasi sistem yang didapat.

5. BAB V PENUTUP: Bab ini menjelaskan tentang kesimpulan yang didapat dari penelitian ini serta saran yang mungkin bisa dilakukan dilakukan untuk penelitian-penelitian yang serupa selanjutnya.
- c. Bagian akhir yang berisi daftar Pustaka dan lampiran.

