

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pada zaman sekarang, manusia hampir tidak dapat lepas dengan teknologi. Hampir semua kegiatan dan produktivitas manusia bergantung pada teknologi-teknologi canggih seperti komputer, telepon genggam, internet, dan lain-lain. Salah satu teknologi yang paling banyak dipakai adalah internet. Badan Pusat Statistik menyatakan bahwa dari tahun 2011 sampai dengan tahun 2021, penggunaan internet warga Indonesia sendiri setiap tahun semakin meningkat dengan presentasi penduduk pengguna internet di atas usia 5 tahun adalah 62.10% pada tahun 2021 (Badan Pusat Statistik, 2021).

Maraknya penggunaan internet memicu peningkatan pelayanan dan fasilitas yang disediakan oleh perusahaan-perusahaan terutama yang banyak memanfaatkan internet sebagai media pelayanannya. Dari dorongan tersebut, melahirkan banyaknya penelitian dan cabang ilmu yang baru. Penelitian-penelitian tersebut tidak hanya bertujuan untuk peningkatan fasilitas saja, tetapi juga untuk meminimalisir *error* dan peningkatan efektivitas dari pelayanan yang ditawarkan. Salah satu penelitian yang cukup banyak dibicarakan adalah *Natural Language Processing* (NLP). NLP adalah cabang ilmu atau penelitian yang bertujuan supaya komputer atau sistem dapat mengerti bahasa yang digunakan manusia sehari-hari dalam berkomunikasi. Beberapa

aplikasi atau program yang menggunakan NLP adalah Google Translate, Google Assistant, dan Siri (Oktriwina, 2021). Aplikasi atau program tersebut sangat menitikberatkan pada pemahaman struktur suatu bahasa dalam melayani dan membantu manusia terhadap kebutuhannya.

Penelitian yang dilakukan pada NLP dibagi ke dalam beberapa fokus tahapan yang berbeda. Tahapan-tahapan yang dimaksud seperti pemahaman bahasa oleh sistem melalui suara, pemahaman *grammar* dalam suatu bahasa, kesimilaritas dari beberapa kata, dan lain-lain. Dari beberapa penelitian yang ada, terdapat penelitian yang cukup menarik perhatian penulis, yaitu *word embedding*. *Word embedding* merupakan salah satu metode dalam memetakan kata-kata ke dalam vektor representasinya. Penelitian mengenai *word embedding* banyak dilakukan oleh para peneliti dari berbagai belahan dunia yang melahirkan beberapa metode yang memiliki landasan atau tolak ukur yang berbeda-beda. Contoh metodenya adalah Word2Vec, Glove, dan FastText. Algoritma dari Word2Vec diciptakan pertama kali oleh Mikolov *et al.* (2013a). Algoritma dari metode lain yaitu GloVe, diciptakan dan diteliti oleh Pennington *et al.* (2014). Dan untuk algoritma dari FastText, pertama kali dikembangkan oleh Joulin *et al.* (2016) dan Bojanowski *et al.* (2017).

Penelitian dalam *word embedding* tidak hanya meneliti mengenai penciptaan metode baru, tetapi juga meneliti kinerja metode yang sudah ada dan mengembangkannya. Seperti contohnya adalah penelitian yang menganalisis teknik Word2Vec dengan pendekatan probabilistik kemunculan kata (Allen dan Hospedales, 2019). Penelitian juga dilakukan dengan membandingkan kinerja algoritma-algoritma

dalam *word embedding* yang dilakukan oleh Nurdin *et al.* (2020). Terdapat juga penelitian mengenai perkembangan dari metode dalam *word embedding* seperti yang dilakukan Bollegala dan O'Neill (2022), Singh *et al.* (2022), serta Goldberg (2022). Penelitian mengenai *word embedding* pun tidak hanya dilakukan dalam satu bahasa saja, tetapi juga antar bahasa. Contohnya seperti yang dilakukan Brychcín (2020), Bhowmik dan Ralescu (2021), serta Doval *et al.* (2018).

Peranan matematika dalam *word embedding* tidak hanya pada probabilitas kemunculan kata dan operasi vektor kata saja. Dalam beberapa penelitian dalam *word embedding*, terutama dalam pemetaan vektor kata antar bahasa, menggunakan prinsip transformasi linear dalam pengerjaannya. Transformasi linear digunakan pada pemetaan yang memetakan vektor kata dari bahasa pertama ke terjemahan vektor kata yang berada pada bahasa kedua. Transformasi linear tersebut diterapkan pada beberapa penelitian yang sudah dilakukan, seperti penelitian Xing *et al.* (2015), Brychcin *et al.* (2019), Doval *et al.* (2018), dan Bollegela *et al.* (2017).

Penelitian-penelitian mengenai *word embedding* antar bahasa menarik perhatian penulis terutama pada bagian transformasi linear yang diterapkan pada pemetaan antar vektor bahasa. Hal tersebut mendorong penulis untuk melakukan penelitian mengenai cara memperoleh matriks transformasi linear yang diterapkan pada salah satu penelitian tersebut. Selain itu, penulis juga terdorong untuk melakukan penelitian terhadap penerapan dari matriks yang sudah diperoleh. Penelitian tersebut dilakukan dengan menerapkan matriks pada contoh kasus dari pemetaan vektor kata dari suatu bahasa ke bahasa yang lain.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang yang dijelaskan pada subbab sebelumnya, identifikasi masalah yang akan dibahas pada skripsi ini adalah:

1. Bagaimana cara memperoleh transformasi linear yang terjadi pada pemetaan vektor representasi kata bahasa pertama ke vektor representasi bahasa kedua menggunakan *singular value decomposition*?
2. Bagaimana penerapan dari transformasi linear tersebut ke dalam contoh kasus masalah pemetaan antar vektor kata?

## 1.3 Batasan Masalah

Dalam skripsi ini penelitian dilakukan dengan beberapa batasan sebagai berikut

1. Bahasa yang digunakan dalam eksperimen pada penelitian ini menggunakan dua bahasa, yaitu Bahasa Inggris dan Bahasa Indonesia.
2. Model yang digunakan dalam memetakan kata ke dalam vektor representasinya menggunakan model FastText (Grave *et al.*, 2018).
3. Penerjemahan kata dari bahasa pertama ke bahasa kedua menggunakan bantuan Google Translate. Jika terdapat hasil terjemahan suatu kata yang menghasilkan sebuah frasa, maka frasa tersebut diasumsikan sebagai sebuah kata.
4. Setiap kata dalam bahasa pertama memiliki hasil pemetaan berupa kata terjemahan dalam bahasa kedua yang berbeda.

#### **1.4 Tujuan Penelitian**

Tujuan dari penelitian skripsi ini dilakukan adalah:

1. Memperoleh transformasi linear yang terjadi pada pemetaan vektor representasi kata bahasa pertama ke vektor representasi bahasa kedua menggunakan *singular value decomposition*.
2. Menerapkan transformasi linear tersebut ke dalam contoh kasus transformasi antar vektor bahasa.

#### **1.5 Kegunaan Penelitian**

Penulis skripsi mengharapkan bahwa skripsi ini dapat memberikan kegunaan atau manfaat diantaranya:

1. Memberikan wawasan mengenai penerapan ilmu dan konsep matematika di luar bidang matematika.
2. Menjadi referensi bagi peneliti lain terutama yang melakukan penelitian dalam bidang atau topik yang sama.

#### **1.6 Metodologi Penelitian**

Langkah-langkah yang dilakukan dalam melakukan penelitian dalam skripsi ini adalah

1. Studi Literatur

Studi literatur dilakukan untuk mempelajari referensi dari beberapa penelitian yang telah dilakukan yang berkaitan dengan transformasi linear *word*

*embedding*. Studi literatur juga dilakukan untuk mencari definisi, teorema, dan materi dari sumber literatur lain yang dapat mendukung pelaksanaan penelitian ini. Sumber-sumber yang diperoleh dari studi literatur digunakan untuk mengidentifikasi masalah dan tujuan penelitian, metode yang digunakan dalam memperoleh transformasi linear, serta penetapan batasan-batasan yang dibatasi dalam skripsi ini.

## 2. Eksperimental

Eksperimen pada penelitian ini dilakukan dengan mengambil beberapa kata dari bahasa pertama dengan terjemahannya pada bahasa kedua. Setelah itu, dapat diperoleh matriks transformasi linear dari vektor-vektor representasi kata-kata tersebut. Eksperimen dilakukan dengan perhitungan secara manual untuk contoh studi kasus sederhana dan dengan perhitungan menggunakan bantuan Python untuk studi kasus pada penerapan pemetaan vektor kata antar bahasa dalam *word embedding*.

## 3. Hasil dan Kesimpulan

Hasil yang diperoleh dari beberapa eksperimen yang telah dilakukan akan dibandingkan dan ditarik kesimpulan.

### 1.7 Sistematika Penulisan

Sistematika penulisan dalam skripsi ini adalah sebagai berikut

## BAB I – PENDAHULUAN

Pada bab ini berisi tentang gambaran umum dan pendahuluan dalam penulisan skripsi yang meliputi latar belakang, identifikasi masalah, tujuan masalah, kegunaan masalah, metodologi penelitian, dan sistematika penulisan.

## BAB II – LANDASAN TEORI

Pada bab ini berisi tentang teori-teori dasar yang digunakan dalam penulisan dan peniltian dalam pembuatan skripsi.

## BAB III – OBJEK DAN METODE PENELITIAN

Pada bab ini berisi tentang objek dan metode yang digunakan dalam penelitian skripsi.

## BAB IV – HASIL DAN PEMBAHASAN

Pada bab ini berisi hasil dan pembahasan dari penelitian yang dilakukan pada transformasi atau perubahan dari kata ke dalam bentuk vektor.

## BAB V – SIMPULAN DAN SARAN

Pada bab ini berisi tentang kesimpulan yang diperoleh dari hasil yang telah dijelaskan pada bab sebelumnya dan saran untuk pengembangan penelitian selanjutnya.