

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pemahaman makna teks Al-Quran merupakan tantangan kompleks yang memerlukan pendekatan mendalam. Sebagai kitab suci umat Islam, Al-Quran mengandung nilai-nilai spiritual, moral, dan hukum yang kaya. Teks Al-Quran sering memiliki makna mendalam dan konteks beragam, sehingga pemahaman akurat menjadi sangat penting [1]. Salah satu permasalahan utama ialah kesulitan mengidentifikasi kesamaan makna antarkata atau frasa dalam Al-Quran, yang membutuhkan analisis linguistik cermat dan tepat.

*Word2Vec*, yang dikembangkan oleh Mikolov et al. (2013), merupakan teknik representasi kata yang berhasil menangkap hubungan semantik antarkata dengan sangat baik. Melalui pelatihan yang tepat, *Word2Vec* dapat memahami konteks dan makna sebuah kata berdasarkan kata-kata di sekitarnya [2]. Teknik ini menggunakan model pembelajaran terawasi untuk mempelajari vektor representasi kata yang dapat digunakan untuk berbagai aplikasi *NLP*, seperti analisis sentimen, penerjemahan mesin, dan pemrosesan teks lainnya. Hal ini sangat relevan untuk analisis teks terjemahan Al-Quran yang memiliki struktur dan konteks unik.

Penggunaan *Word2Vec* dalam analisis teks terjemahan Al-Quran Bahasa Inggris memungkinkan peneliti menangkap nuansa dan variasi makna yang mungkin tidak terdeteksi dengan metode tradisional. Misalnya, dalam ayat-ayat yang memiliki kata-kata bermakna ganda atau bersinonim, *Word2Vec* dapat membantu mengidentifikasi konteks spesifik yang memengaruhi makna kata tersebut. Penelitian Goldberg dan Levy (2014) menunjukkan bahwa teknik ini mampu meningkatkan akurasi dalam memahami hubungan semantik antarkata dalam berbagai jenis teks [3].

Dalam penelitian ini penulis memilih menggunakan pendekatan *Continuous Bag of Words (CBOW)*. Dengan merujuk dari penelitian Peijun Li (2023) menunjukkan bahwa hasil dari membandingkan elemen efisiensi termasuk total waktu pelatihan, jumlah kata efektif, waktu pelatihan, dan kata efektif dalam hasil

eksperimen, model *CBOW* dapat dimanfaatkan dengan lebih baik dibandingkan model *Skip-gram*.

Di sisi lain, *lexicon* merupakan kumpulan kata atau frasa yang telah diberi anotasi makna tertentu, yang dapat membantu mengklarifikasi interpretasi makna teks. *Lexicon* dapat berfungsi sebagai sumber referensi penting dalam analisis linguistik, karena menyediakan informasi tentang makna kata, sinonim, antonim, serta hubungan semantik lainnya. Dalam konteks Al-Quran, penggunaan *lexicon* dapat membantu memahami makna kata-kata yang mungkin memiliki interpretasi berbeda bergantung pada konteksnya. Bond dan Paik (2012) menunjukkan bahwa *lexicon* dapat digunakan untuk meningkatkan akurasi dalam pemrosesan teks dengan menyediakan anotasi makna yang lebih tepat [4].

Perhitungan *semantic similarity* berbasis *lexicon* umumnya dihitung dengan memanfaatkan kesamaan topologis yang terdapat dalam ontologi seperti *WordNet*[5]. Lingling Meng (2013) dalam jurnalnya menemukan bahwa dalam metode perhitungan *semantic similarity* berbasis *WordNet (lexicon)* lebih akurat ketika menggunakan penggabungan metode[6]. Dalam penelitian ini penulis mengusulkan untuk menggunakan metode dari 3 kelompok. Metode yang diusulkan pertama adalah metode perhitungan *semantic similarity* berdasarkan jarak (Wu & Palmer)[7], yang kedua metode perhitungan berdasarkan *Information Content* (Jiang & Conrath)[8] dan yang ketiga metode mencari nilai *semantic similarity* berdasarkan jumlah seberapa banyak kata yang relevan terhubung melalui jalur yang tidak terlalu panjang dan yang tidak sering berubah arah (Hirst St-Onge)[9].

Dengan mengintegrasikan *Word2Vec* dan *lexicon*, diharapkan tercipta model yang lebih akurat dalam mengidentifikasi kesamaan makna dalam teks Al-Quran. Integrasi ini memungkinkan pemanfaatan keunggulan kedua pendekatan tersebut, yaitu kemampuan *Word2Vec* dalam menangkap hubungan semantik dari data teks besar dan kemampuan *lexicon* dalam menyediakan anotasi makna yang tepat. Penelitian ini bertujuan mengeksplorasi dan mengimplementasikan keterpaduan antara *Word2Vec* dan *lexicon* dalam konteks analisis kesamaan makna pada teks Al-Quran. Melalui penelitian ini, diharapkan diperoleh pemahaman yang lebih mendalam tentang cara teknologi pengolahan bahasa alami dapat digunakan untuk meningkatkan kualitas interpretasi teks-teks keagamaan, khususnya Al-Quran.

## 1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah disampaikan sebelumnya, maka, terdapat beberapa permasalahan utama yang harus diselesaikan. Permasalahan tersebut adalah sebagai berikut :

1. Bagaimana cara mengintegrasikan model *Word2Vec* dengan *lexicon* untuk analisis teks terjemahan Al-Quran dalam Bahasa Inggris?
2. Seberapa akurat model yang dihasilkan dalam mengidentifikasi kesamaan makna antar kata dalam teks terjemahan Al-Quran dalam Bahasa Inggris?
3. Bagaimana perbedaan kinerja model *Word2Vec* dengan model *Word2Vec* yang dikombinasikan dengan *lexicon* dalam analisis kesamaan makna kata pada teks terjemahan Al-Quran Bahasa Inggris?

## 1.3 Batasan Masalah

Untuk mengarahkan pembahasan pada penelitian ini, maka diperlukan adanya batasan masalah sehingga pembahasan tidak menyimpang dari sasaran yang dituju, maka Batasan masalah penelitian ini yaitu:

1. Penelitian ini hanya akan menganalisis teks Al-Quran dalam bahasa Inggris. Terjemahan ini dipilih untuk memastikan konsistensi data dan memudahkan penggunaan alat NLP yang umumnya lebih berkembang untuk bahasa Inggris.
2. Model *Word2Vec* yang digunakan dalam penelitian ini adalah varian *Continuous Bag of Words (CBOW)*.
3. Penelitian ini akan menggunakan beberapa *lexicon* yang sudah dikenal luas dan sering digunakan dalam penelitian *NLP*, yaitu metode *Wu Palmer*, *Jiang Conrath*, dan *Hirst St. Onge*. Dengan memanfaatkan kesamaan topologis dari *WordNet*.
4. Data yang digunakan dalam penelitian ini dibatasi pada *corpus* teks Al-Quran terjemahan bahasa Inggris yang tersedia secara publik.
5. Evaluasi kinerja model akan difokuskan pada pengukuran akurasi dalam mengidentifikasi kesamaan makna antar kata
6. Software yang digunakan adalah *Phyton*.

Dengan batasan-batasan ini, diharapkan penelitian dapat dilakukan dengan lebih terarah dan menghasilkan temuan yang signifikan dan relevan dalam konteks analisis kesamaan makna dalam teks Al-Quran.

#### 1.4 Tujuan Penelitian

Tujuan dari skripsi ini adalah :

1. Mengembangkan metode yang mengintegrasikan model *Word2Vec* dengan berbagai metode *lexicon* untuk analisis teks terjemahan Al-Quran dalam bahasa Inggris.
2. Mengevaluasi akurasi model yang dihasilkan dalam mengidentifikasi kesamaan makna antar kata dalam teks terjemahan Al-Quran dalam bahasa Inggris.
3. Membandingkan kinerja model *Word2Vec* dengan model yang mengkombinasikan *Word2Vec* dan *lexicon* dalam analisis kesamaan makna kata pada teks terjemahan Al-Quran bahasa Inggris.

#### 1.5 Metode Penelitian

Metode penelitian pada tugas akhir ini yaitu sebagai berikut:

1. Studi Literatur.

Tahap studi literatur dilakukan pengumpulan data dan referensi literatur berupa buku, jurnal, karya ilmiah, dan artikel yang berkaitan dengan metode-metode yang digunakan yaitu, *word2vec* dan berbagai *lexicon*.

2. Analisis

Pada tahap ini dilakukan pengkajian dan analisis terhadap hasil referensi literatur yang diperoleh yang sesuai dengan topik masalah yang akan diteliti dalam skripsi ini.

3. Simulasi

Pada tahap ini penulis melakukan simulasi model “keterpaduan dari *word2vec* dan *lexicon*” menggunakan data *training* dan data *testing* dengan konteks alquran menggunakan bahasa pemrograman *python*. Kemudian dari hasil simulasi tersebut akan dihitung nilai akurasi ketika diuji dengan data *testing* sebagai evaluasi.

## 1.6 Sistematika Penulisan

Dalam skripsi ini terdapat lima bab utama dan beberapa sub bab, serta daftar pustaka dan lampiran. Berikut adalah penjabarannya :

### BAB I PENDAHULUAN

Bab ini berisi Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Metode Penelitian, dan Sistematika Penulisan.

### BAB II LANDASAN TEORI

Landasan teori berisi tentang penjelasan teori-teori yang berkaitan dengan pembahasan dari penelitian ini. Didapatkan dari sumber yang terdapat pada buku, artikel, dan penelitian terdahulu.

### BAB III KETERPADUAN DARI WORD2VEC DAN LEXICON PADA KESAMAAN MAKNA KATA DALAM DOMAIN AL-QURAN

Bab ini berisi tentang pembahasan utama mengenai metode yang digunakan, berupa pembahasan rinci tentang pengumpulan data, *pre-processing*, metode *word2vec* dan *lexicon*. serta upaya bagaimana pendekatan untuk mengsinergikan keduanya.

### BAB IV STUDI KASUS DAN ANALISIS

Pada bab ini akan dijelaskan mengenai analisis hasil pengaplikasian dan percobaan kombinasi dari *word2vec* dan *lexicon* untuk mencari *semantic similiarity* pada terjemahan Al-Qur'an dalam bahasa Inggris. Pada penelitian ini, percobaan dilakukan dengan menggunakan bahasa pemrograman Python dan dijalankan di Jupyter Notebook.

## BAB V

### KESIMPULAN DAN SARAN

Pada bab ini, berisi tentang kesimpulan yang dapat diambil dari hasil analisis yang telah dilakukan di bab sebelumnya. Dan juga terdapat saran untuk pengembangan penelitian lanjutan yang dapat dilakukan dari penelitian ini

