

1. Pendahuluan

1.1 Latar Belakang

Bahasa Indonesia memiliki jenis kata - kata yang mempunyai makna lebih dari satu. Namun penggunaannya tergantung dari konteks kalimat yang menyertainya. Penggunaan kata yang memiliki makna lebih dari satu ini dapat menyebabkan ambiguitas kalimat. Ambiguitas kalimat ini dapat menyebabkan keraguan dalam penafsiran kalimat itu sendiri atau disebut dengan redundansi kalimat. Untuk mengatasi masalah ambiguitas tersebut diperlukan cara untuk memilih kata yang tepat sesuai dengan konteks kalimat. Cara atau teknik untuk menyelesaikan ambiguitas kata dalam kalimat dikenal sebagai *Word Sense Disambiguation*.

Word Sense Disambiguation (WSD) adalah proses untuk mengidentifikasi makna kata yang digunakan dalam kalimat tertentu ketika kata memiliki sejumlah makna yang berbeda [1]. Dalam *natural language processing*, penggunaan WSD ini bermanfaat untuk perbaikan kata dalam mesin penerjemah. Perbaikan kata ini dapat dilihat dalam *google translate*, dimana pengguna dapat membantu memperbaiki konteks kalimat yang telah diubah ke bahasa target dengan memilih kata yang dianggap ambigu. Selain itu dapat digunakan pula untuk *proof read*, dimana *proof read* adalah metode membaca ulang untuk memeriksa sebuah penulisan serta memastikan tidak ada penulisan yang salah, tidak konsisten dan tidak mengandung makna yang ambigu.

Saat ini, masih sedikit publikasi penelitian tentang *word sense disambiguation* untuk bahasa Indonesia. Salah satu penelitian *word sense disambiguation* bahasa Indonesia yang telah dipublikasikan oleh Edi Faisal, Farza Nurifan, dan Riyanarto Sarno dengan menggunakan algoritma *Support Vector Machine* (SVM) sebagai penyelesaiannya [9]. Dimana dalam penelitian ini menggunakan teks bahasa Indonesia dalam situs Wikipedia, dengan nilai akurasi yang didapat sebesar 87%. Akurasi ini dianggap bagus mengingat sedikitnya penelitian – penelitian yang membahas *word sense disambiguation* Bahasa Indonesia.

Dalam penyelesaian *word sense disambiguation*, secara umum penyelesaiannya dibagi menjadi tiga pendekatan yaitu pendekatan *supervised*, pendekatan *unsupervised* dan pendekatan *knowledge-based*. Dalam penelitian ini penulis menggunakan pendekatan *supervised* dengan menggunakan algoritma *naïve bayes* yang diterapkan kedalam teks Bahasa Indonesia.

1.2 Perumusan dan Batasan Masalah

Dalam permasalahan *word sense disambiguation* Bahasa Indonesia, penelitiannya masih sedikit dilakukan. Karena itu penulis akan membahas *word sense disambiguation* dalam Bahasa Indonesia dengan menggunakan algoritma *naïve bayes*. Pemilihan algoritma *naïve bayes* didasarkan penelitian oleh Thwet Aung, N. T., Soe, N. K., & Thein, N. L dengan menggunakan Bahasa Myanmar [12]. Penelitian tersebut menghasilkan akurasi sebesar 89%. Oleh karena itu, penulis akan menerapkan algoritma *naïve bayes* ke dalam *word sense disambiguation* Bahasa Indonesia.

Adapun rumusan masalah yang disusun:

1. Bagaimana mencari kata yang memiliki makna ambigu dan mengetahui makna yang tepat?
2. Bagaimana nilai akurasi yang didapat menggunakan algoritma *Naïve Bayes* bila diterapkan dalam *word sense disambiguation* Bahasa Indonesia?

Serta batasan masalah dalam penelitian ini:

1. Data yang digunakan dalam penelitian ini menggunakan teks Bahasa Indonesia menggunakan bahasa formal.
2. Menggunakan bahasa pemrograman *python*.

1.3 Tujuan

Tujuan penelitian ini adalah:

1. Menambah penelitian tentang *word sense disambiguation* dalam Bahasa Indonesia dengan menggunakan algoritma *naïve bayes*.
2. Menentukan *sense* dari suatu kata dengan menggunakan algoritma *naïve bayes* yang diterapkan dalam *word sense disambiguation*.
3. Menemukan nilai akurasi dari algoritma *naïve bayes* yang diterapkan pada *word sense disambiguation* dalam Bahasa Indonesia.

1.4 Organisasi Tulisan

Bagian selanjutnya, pada bab kedua akan menjabarkan tentang studi – studi terkait tentang penelitian *word sense disambiguation* Bahasa Indonesia dengan menggunakan algoritma *naïve bayes*. Setelah menjabarkan studi terkait, pada bab ketiga dijelaskan tentang sistem yang dirancang dari penelitian ini. Selanjutnya pada bab keempat dijabarkan tentang hasil penelitian serta analisis dari penelitian *word sense disambiguation* Bahasa Indonesia dengan menggunakan algoritma *naïve bayes* ini. Bab kelima menampilkan kesimpulan dari penelitian ini. Setelah kesimpulan ditampilkan daftar Pustaka yang digunakan dalam penelitian.