

1. PENDAHULUAN

1.1 Latar Belakang

Informasi atau *knowledge* dapat diambil dari sebuah dokumen, artikel, bahkan dalam sebuah kalimat. Argumen semantik adalah salah satu bidang linguistik dalam mempelajari makna di dalam sebuah kalimat. Argumen semantik merupakan bagian dari teknik dalam mengembangkan solusi dari bidang *text mining* [5]. *Text mining* adalah upaya pencarian dan penambangan data yang berupa teks di dalam sebuah dokumen [6]. Salah satu bidang *text mining* yang memerlukan semantik adalah *task* pada *Natural Language Processing* (NLP) seperti *Information Extraction* (IE), *Question Answering* (QA), dan *Summarization* [8].

Representasi argumen semantik kalimat dapat disampaikan dalam bentuk ekstraksi informasi dan menjawab pertanyaan, seperti *Who*, *What*, *Whom*, *When*, *Where*, *Why*, dan *How* [10]. Representasi argumen semantik dijelaskan dalam bentuk peran semantik dan proses penentuan peran semantik tersebut dikenal sebagai pelabelan peran semantik atau *semantic role labeling*. Pelabelan peran semantik dibagi dalam dua cara, yaitu identifikasi argumen semantik dan klasifikasi argumen semantik. Identifikasi argumen semantik mengklasifikasi apakah setiap elemen sintaksis merupakan sebuah argumen atau tidak. Sedangkan, klasifikasi argumen semantik mengidentifikasi argumen semantik ke dalam peran semantik yang lebih spesifik, seperti *ARG0*, *ARG1*, *ARG2*, *ARG3*, *ARG4*, dan *ARGM* [7]. Pelabelan peran semantik ini merupakan salah satu tahap sebelum memasuki proses *text mining* [6]. Dengan melakukan klasifikasi argumen semantik, maka akan membantu tahap untuk melakukan *text mining* nantinya, sehingga dapat menjawab pertanyaan *Who*, *What*, *Whom*, *When*, *Where*, *Why*, dan *How*.

Pada penelitian ini, lebih menerapkan klasifikasi argumen semantik pada *database* PropBank. Dalam melakukan klasifikasi argumen semantik digunakan

feature konteks semantik, yang terdiri dari *feature baseline* dan *feature* argumen semantik tetangganya sebagai *feature* tambahan. *Feature* argumen semantik tetangga dapat dimanfaatkan untuk melihat ketergantungan dari semua argumen-argumen di dalam sebuah kalimat berdasarkan predikat pada masing-masing kalimatnya, sehingga dapat meningkatkan akurasi dalam klasifikasi argumen semantik dengan menggunakan metode *classifier Support Vector Machine* (SVM). Metode ini dipilih karena mampu mengklasifikasikan data berdimensi tinggi yang dalam konteks tugas akhir ini adalah berupa teks. Prinsip dari SVM adalah *Structural Risk Minimization* (SRM) yaitu meminimalkan *error* pada *training-set*.

1.2 Perumusan Masalah

Rumusan masalah yang dikaji dalam Tugas Akhir ini adalah *feature* apa yang mempengaruhi dalam memanfaatkan ketergantungan argumen pada klasifikasi argumen semantik?

1.3 Tujuan

Tujuan penelitian Tugas Akhir ini adalah menganalisa *feature* konteks semantik, yaitu *feature baseline* dan *feature* argumen semantik tetangga dalam memanfaatkan ketergantungan argumen pada klasifikasi argumen semantik.

1.4 Batasan Masalah

Adapun batasan masalah pada Tugas Akhir ini adalah :

1. Melakukan klasifikasi argumen semantik dengan menggunakan *feature baseline* dan *feature* argumen semantik tetangga.
2. Penelitian ini hanya melakukan klasifikasi argumen semantik, sehingga dengan asumsi proses identifikasi argumen telah dilakukan.
3. Label pada setiap kata dalam sebuah kalimat didasarkan pada label yang ada pada *database* PropBank.
4. Dataset yang digunakan adalah kalimat berbahasa Inggris pada *database* PropBank.
5. Metode *Support Vector Machine* (SVM) tidak dibahas secara teknis.

1.5 Metodologi Penyelesaian Masalah

Metodologi yang digunakan untuk menyelesaikan masalah dalam Tugas Akhir ini adalah :

1. Studi Literatur

Mencari referensi dan sumber-sumber baik dari internet, buku, ataupun jurnal ilmiah mengenai permasalahan yang berkaitan dengan Tugas Akhir, seperti klasifikasi argumen semantik berdasarkan ketergantungan argumen tetangganya, *feature-feature* yang mendukung, dan metode *Support Vector Machine* (SVM).

2. Pencarian dan Pengumpulan Data

Dataset yang digunakan merupakan kalimat yang berbahasa Inggris pada *database* PropBank.

3. Perancangan Sistem

Model sistem akan dirancang berdasarkan hasil analisis pada studi literatur. Perancangan sistem dimulai dari inputan dataset, melakukan *preprocessing*, pelabelan argumen semantik dengan menggunakan *feature* yang telah digunakan, melakukan klasifikasi argumen semantik, dan rancangan hasil outputnya.

4. Implementasi dan Pengujian

Melakukan implementasi dan pengujian berdasarkan rancangan yang telah dibuat sebelumnya, kemudian dilihat tingkat akurasi dari *feature-feature* yang telah ditentukan dengan menggunakan SVM.

5. Analisis Hasil Pengujian

Pada tahap ini dilakukan analisa berdasarkan hasil pengujian sistem, dengan membandingkan hasil klasifikasinya dan hasil ketergantungan argumen pada *feature-feature* yang telah digunakan.

6. Penyusunan Laporan

Pengambilan kesimpulan dan melaporkan semua yang telah dilakukan selama implementasi dan pengujian dalam penulisan Tugas Akhir ini.