

1. PENDAHULUAN

1.1 Latar Belakang

Semantic Role Labeling adalah proses untuk menentukan label peran atau *role* dari tiap argumen pada kalimat berdasarkan predikatnya. Proses ini menghasilkan lebih dari apa yang dapat dihasilkan oleh *parser* yang sudah ada saat ini. Sebelumnya jika menggunakan *parser* yang didapatkan adalah potongan per kata dari kalimat tanpa mengetahui makna dari kalimat, sementara *Semantic Role Labeling* dapat memisahkan kalimat menjadi argumen-argumen yang kemudian diberikan label peran berdasarkan predikatnya dan makna kalimat. Dari hasilnya pembaca masih dapat mengetahui siapa yang melakukan apa kepada siapa, bagaimana, di mana, dan kapan [8].

Semantic Role Labeling sangat berguna untuk mengekstrak informasi dari suatu kalimat yang kemudian dapat digunakan pada *question answering* sebagai contohnya. Namun bagaimana jika *dataset* yang akan digunakan belum memiliki label dan berasal dari domain yang berbeda dengan *dataset* yang sudah memiliki label? Salah satu solusi yang dapat digunakan adalah *Domain Adaptation*. Adaptasi domain adalah proses untuk mendapatkan sebuah *dataset* baru yang sudah memiliki *label* dengan menggunakan *dataset* lain dari domain yang berbeda namun sudah memiliki *label*. Seperti misalnya dengan menggunakan *data* yang sudah memiliki label dari *newswire domain* untuk memberi label pada kumpulan *data* dari *biomedical domain*. *Dataset* yang digunakan dibagi menjadi dua yakni, *training data (source domain)* dan *testing data (target domain)*. Adaptasi domain sangat membantu dalam pelabelan *dataset* karena dapat mengurangi *cost* yang seharusnya digunakan untuk membuat sebuah sistem *semantic role labeling*.

Salah satu metode dari adaptasi domain adalah *instance weighting*. Berdasarkan penelitian yang sebelumnya dilakukan oleh Jiang Jing metode ini memiliki performa yang baik dengan memberikan bobot kepada *instances* yang ada di *source domain* yang memiliki relasi dengan *instances* yang ada di *target domain*. Permasalahan dari adaptasi domain adalah pemrosesan data yang sangat besar. Pemrosesan ini akan membutuhkan banyak waktu untuk diselesaikan. Sementara tidak semua data yang diproses benar-benar diperlukan pada proses adaptasi domain. Cara yang digunakan pada penelitian sebelumnya adalah dengan menggunakan algoritma *instance pruning* pada *instance weighting* [6]. *Instance pruning* akan menghapus *misleading instances* dari *source domain*. *Instance* yang dihapus dianggap tidak terkait dengan *target domain* dan proses adaptasi domain. Sehingga jumlah data yang akan digunakan pada adaptasi domain akan berkurang dan mengurangi waktu yang dibutuhkan.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas maka perumusan masalah pada tugas akhir ini adalah:

1. Bagaimana hasil dari adaptasi domain menggunakan *instance weighting* dengan algoritma *instance pruning*?
2. Bagaimana pengaruh *instance pruning* pada adaptasi domain?

1.3 Tujuan

Tujuan yang ingin dicapai dalam tugas akhir ini adalah:

1. Menganalisis hasil adaptasi domain menggunakan *instance weighting* dengan algoritma *instance pruning*.
2. Menganalisis performa dari adaptasi domain sebelum dan setelah dilakukan *instance pruning*.

1.4 Batasan Masalah

Dalam menyelesaikan masalah yang telah dirumuskan di atas terdapat beberapa hal yang menjadi batasan masalah:

1. Metode yang digunakan pada adaptasi domain adalah *instance weighting* dengan menggunakan algoritma *instance pruning* berdasarkan penelitian sebelumnya.
2. *Dataset* yang digunakan dalam tugas akhir ini berupa korpus yang berasal dari PropBank dan korpus dari terjemahan Al-Qur'an dalam bahasa Inggris.
3. Semua *dataset* yang digunakan sudah memiliki label. *Dataset* terjemahan Al-Qur'an diberi label menggunakan *semantic role labeling* yang tersedia secara *online*.
4. *Evaluation metric* digunakan dalam menilai performa dari sistem adalah *F1 score*.
5. *Classifier* yang digunakan dalam tugas akhir ini adalah *Naïve Bayes*.

1.5 Metodologi Penyelesaian Masalah

Metode penyelesaian masalah untuk tugas akhir ini adalah:

1. Studi literatur
Pada tahap ini dilakukan pencarian maerial yang berkaitan dengan tugas akhir, baik itu studi kasus, algoritma, teori dan materi lainnya yang terkait dengan pelaksanaan tugas akhir.
2. Pengumpulan *data*
Pada tahap ini dikumpulkan *data* berupa teks yang diambil dari PropBank. *Data* yang diambil berupa korpus yang dijadikan *dataset* yang

berfungsi sebagai *input* dari sistem. *Dataset* berasal dari korpus *Propbank* dalam bahasa Inggris dan terjemahan Al-Qur'an dalam bahasa Inggris.

3. Perancangan dan implementasi sistem

Pada tahap ini dilakukan perancangan dan pengimplementasian sistem yang digunakan untuk melakukan adaptasi domain. Sistem yang dibangun menggunakan *dataset* yang sudah dikumpulkan.

4. Pengujian dan analisis hasil pengujian

Pada tahap ini dilakukan pengujian sistem dengan menggunakan *dataset* yang ada sebagai *input*. Hasil dari pengujian dianalisis dengan menitikberatkan pada kecepatan dan akurasi performa sistem.

5. Penyusunan buku tugas akhir

Pada tahap ini dilakukan penyusunan buku tugas akhir yang merangkul pengerjaan tugas akhir. Termasuk hasil pengujian dan analisis sistem serta kesimpulan dan saran dari tugas akhir ini.

1.6 Sistematika Penulisan

Sistematika penulisan dari tugas akhir ini adalah sebagai berikut:

1. Bab 1 Pendahuluan

Bab 1 berisi penjelasan tugas akhir secara umum. Mulai dari latar belakang, permasalahan, tujuan tugas akhir, dan batasan dari sistem yang dibuat. Termasuk juga metodologi penyelesaian masalah dan sistematika penulisan.

2. Bab 2 Kajian Pustaka

Bab 2 berisi penjelasan dasar-dasar teori yang digunakan pada tugas akhir yakni *Semantic Role Labeling*, *Domain Adaptation*, *Instance Weighting*, Korpus *Propbank*, Korpus Terjemahan Al-Qur'an, *Featue*, WEKA, dan *F1 Score*.

3. Bab 3 Perancangan Sistem

Bab 3 berisi perancangan sistem yang sudah dibangun mulai dari *flow diagram* gambaran kerja sistem, penjelasan *dataset* yang digunakan, ekstraksi fitur, *preprocessing* data untuk dijadikan *input* adaptasi domain, penjelasa metode adaptasi domain, *training classifier*, dan kebutuhan perangkat dari sistem yang dibangun.

4. Bab 4 Pengujian dan Anlisis

Bab 4 berisi pengujian yang dilakukan pada sistem dengan mengimplementasikan rancangan sistem yang dibuat pada bab 3. Pada bab ini terdapat tujuan dari pengujian, skenario pengujian, dan hasil dari pengujian. Hasil yang didapat dari pengujian kemudian dianalisis yang hasilnya akan menjadi acuan dalam pengambilan kesimpulan dan pemberian saran pada bab 5.

5. Bab 5 Kesimpulan dan Saran

Bab 6 berisi kesimpulan yang didapat dari analisis hasil pengujian pada bab 4. Kesimpulan menggambarkan bagaimana performa dari sistem yang dibangun secara umum. Dari kesimpulan dan analisis hasil pengujian kemudian diberikan saran yang dapat digunakan pada penelitian selanjutnya.