

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Agar Sistem Pendeteksi Kalimat Bahasa Indonesia Berstruktur Rancu dapat melakukan pendeteksian kalimat rancu, ada beberapa tahap yang harus dilakukan, yaitu sebagai berikut.

- 1) Pengumpulan data dan *data preprocessing*.
- 2) Perancangan modul POS *tagger* menggunakan Hidden Markov Model dan Viterbi.
- 3) Perancangan modul *sequential pattern mining* untuk mengekstrak LSP dengan algoritma PrefixSpan.
- 4) Mentransformasikan LSP menjadi *feature* masukan pada *classifier* Weka.

Untuk menyesuaikan representasi LSP dengan bentuk data dan kebutuhan penelitian, maka dilakukan beberapa modifikasi pada algoritma PrefixSpan sebagai berikut:

- 1) Menambahkan nilai panjang maksimum yang diinginkan pada *input*, yaitu *maxLength*;
- 2) Menambahkan nilai *minimum confidence* pada *input*, sebagai *min_confidence*.
- 3) Modifikasi pada metode nomor 3; *database* proyeksi a' (S/a') merupakan data yang mengandung sub sekuen $\langle a b \rangle$ dengan urutan yang tepat sama, tidak sekadar mengandung *item a* dan *b* saja. Hal ini berarti data yang mengandung sub sekuen $\langle b a \rangle$ tidak akan dianggap *database* proyeksi.
- 4) Masih modifikasi pada metode nomor 3; menambahkan kondisi *PrefixSpan*(a' , $l+1, S/a'$) hanya akan dipanggil jika $l+1$ lebih kecil dari nilai *maxLength*.

Dari kedelapan eksperimen yang telah dilaksanakan dengan metode evaluasi *10 fold cross validation*, hasil klasifikasi terbaik adalah klasifikasi yang menggunakan metode Naïve Bayes, dengan konfigurasi LSP: *maxLength=2*, *min_support=0.0* dan *min_confidence=0.0*, yaitu 65.60% akurasi dan *recall* 49.70%. Masih banyak kalimat rancu yang tidak dapat dideteksi oleh sistem, dan setidaknya setiap satu kalimat rancu yang salah diprediksi mewakili satu jenis

penyebab kerancuan yang disebutkan pada instrumen daftar cek di Tabel 4.1. Dengan demikian, secara fungsional dan praktikal, belum dapat dipastikan garis tegas untuk menyebutkan kemampuan dari sistem yang dibangun. Oleh karena itu, masih diperlukan penelitian dan pengembangan lebih lanjut dengan modifikasi dan tambahan pada beberapa bagian.

5.2. Rekomendasi dan Saran

Berdasarkan hasil penelitian yang diperoleh, untuk perbaikan penelitian selanjutnya, ada beberapa poin yang dapat dipertimbangkan, yakni sebagai berikut.

- 1) Perlu dilakukan penelitian lebih lanjut mengenai kecocokan *labeled sequential pattern* dalam mendeteksi *grammatical error* pada bahasa Indonesia.
- 2) Diperlukan pengecekan berulang pada data latih (data kalimat rancu dan kalimat tidak rancu) yang dikumpulkan, bahkan jika perlu pengecekan dilakukan oleh ahli bahasa.
- 3) Untuk memperbaiki POS *tagger*, perlu dicoba *train corpus* bahasa Indonesia selain *corpus* yang digunakan pada penelitian ini.
- 4) Mengekstrak dan menambahkan nilai *computed linguistic* lain sebagai *feature* masukan, selain LSP. Misalnya probabilitas *parsing* untuk setiap kalimat, atau *language model*.
- 5) Memperkaya data latih dengan kalimat yang susunan katanya lebih bervariasi.
- 6) Mengkolaborasikan *sequential pattern mining* dengan metode *rule-based*.