

BAB VI

KESIMPULAN DAN SARAN

6.1 Pendahuluan

Bab ini akan membahas perihal kesimpulan dan saran berdasarkan uraian yang telah dibahas pada bab sebelumnya. Kesimpulan serta saran yang diberikan pada bab ini diharapkan dapat menjadi acuan bagi peneliti untuk melanjutkan penelitian ini.

6.2 Kesimpulan

Berdasarkan hasil uraian yang telah didapatkan maka penulis dapat menarik beberapa kesimpulan sebagai berikut:

1. Proses pada system pengklasifikasian pertanyaan berbahasa Indonesia menggunakan algoritma Support Vector Machine (SVM) dan metode seleksi fitur *Information Gain*, *Chi Square* dan *Mutual Information* berhasil diimplementasikan.
2. Penggunaan metode seleksi fitur *Information Gain* yang dikombinasikan dengan algoritma klasifikasi SVM pada kernel Rbf dengan parameter berupa nilai C :10 dan *threshol* 1.67 berhasil mendapatkan hasil kinerja terbaik dengan nilai *accuracy* 0.9, *precision* 0.91, *recall* 0.86, *f-measure* 0.88, waktu komputasi 14 s dan jumlah fitur 511. Kemudian, penggunaan metode seleksi fitur *Chi Square* yang dikombinasikan dengan algoritma klasifikasi SVM pada kernel *linear*

dengan parameter berupa nilai $C :1$ dan *threshold* 3.5 berhasil mendapatkan hasil kinerja terbaik dengan nilai *accuracy* 0.92, *precision* 0.93, *recall* 0.89, *f-measure* 0.91, waktu komputasi 8 s dan jumlah fitur 432. Selanjutnya, penggunaan metode seleksi fitur *Mutual Information* yang dikombinasikan dengan algoritma klasifikasi SVM pada kernel *linear* dengan parameter berupa nilai $C :1$ dan *threshold* 3.5 berhasil mendapatkan hasil kinerja terbaik dengan nilai *accuracy* 0.92, *precision* 0.93, *recall* 0.89, *f-measure* 0.9, waktu komputasi 7 s dan jumlah fitur 240.

6.2 Saran

Adapun saran yang dapat digunakan pada penelitian selanjutnya, yaitu:

1. Melakukan penambahan data pada dataset penelitian serta membuat jumlah fitur yang ada pada data penelitian menjadi seimbang untuk tiap label agar mencegah terjadi *imbalance data*.
2. Melakukan kombinasi pada penelitian menggunakan metode *resampling* untuk mengatasi terjadinya *imbalance data*.
3. Melakukan perbandingan terhadap metode seleksi fitur dengan tipe lain seperti metode seleksi fitur bertipe *wrapper* dan *embedded selector*.
4. Melakukan penelitian terkait klasifikasi pertanyaan hanya untuk kata tanya berupa “apa”. Dikarenakan kata tanya tersebut memiliki tingkat ambiguitas tinggi dibandingkan dengan kata tanya lain.