

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan dari penelitian yang sudah dilaksanakan dengan judul “Analisis sentimen pengguna twitter terhadap indeks harga saham gabungan menggunakan metode *Term Frequency - Inverse Document frequency* dan *Naïve Bayes classifier*” adalah sebagai berikut:

1. Aplikasi analisis sentimen pengguna twitter terhadap indeks harga saham gabungan menggunakan metode *Term Frequency - Inverse Document frequency* dan *Naïve Bayes classifier* berhasil dibuat. Website yang dibuat dari model analisis sentimen memiliki 2 tipe yakni *landing page* dan *admin page*. Fitur yang disediakan kedua website tersebut diantaranya prediksi tweet dengan input teks dan prediksi dengan input csv. Output dari fitur prediksi dengan input teks adalah prediksi dan probabilitas negatif, positif. Sedangkan output dari fitur input csv adalah *accuracy*, *precision*, *recall*, dan prediksi dalam bentuk tabel.
2. Akurasi, *recall*, dan *precision* terbaik yang didapatkan dari pengujian yang sudah dilakukan adalah sebagai berikut *accuracy* 90,3%, *precision* 90,3%, dan *recall* 95,9%, dengan total data *training* 4815 dan data *testing* 1203

5.2 Saran

Berdasarkan penelitian yang berhasil dilakukan, diberikan beberapa saran yang bisa meningkatkan nilai yang lebih baik dalam hal *accuracy*, *precision* dan *recall*, serta efisiensi dari website diantaranya sebagai berikut :

1. Ukuran dataset bisa dijadikan lebih besar, dan perlu konsistensi dalam pelabelan data. Perlu dilakukan pengecekan berkala pada dataset karena ada kemungkinan salah pada saat pelabelan, dan bisa juga dilakukan penghapusan tweet yang mengandung kata ambigu karena dapat mempengaruhi hasil prediksi.
2. Dapat dilakukan penelitian lain menggunakan algoritma klasifikasi yang lain seperti *Support Vector Machine*, *Random Forest Tree* atau *Logistic Regression*, untuk mengetahui algoritma apa yang paling baik untuk digunakan Bersama dengan TF-IDF.
3. Membuat kode program *preprocessing* se-efisien mungkin agar tidak memakan waktu terlalu lama pada saat *preprocessing*, dan pada saat model diimplementasikan pada website user juga nyaman menggunakannya karena proses prediksi bisa berjalan lebih cepat.
4. Pada saat pengumpulan dataset usahakan datanya seimbang, contohnya pada penelitian ini usahakan tweet dengan sentimen positif sama banyaknya atau tidak beda jauh dengan tweet negatif, agar hasil prediksinya tidak condong ke satu label.