

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian analisis sentimen kasus Ferdy Sambo pada *Twitter* menggunakan algoritma *Naive Bayes* dengan seleksi fitur yang telah dilakukan, maka dapat menarik kesimpulan yaitu:

1. model *Naive Bayes* pada *split data* 90:10 menghasilkan akurasi 69%, *split data* 80:20 menghasilkan akurasi 73%, serta pada *split data* 70:30 menghasilkan akurasi 61%. Dapat disimpulkan bahwa model *Naive Bayes* merupakan hasil terbaik pada *split data* 80:20 dengan akurasi 73%;
2. model *Naive Bayes* kombinasi *Chi Square* pada *split data* 90:10 menghasilkan akurasi 62%, *split data* 80:20 menghasilkan akurasi 75%, serta pada *split data* 70:30 menghasilkan akurasi 62%. Dapat disimpulkan bahwa model *Naive Bayes* kombinasi *Chi Square* merupakan hasil terbaik pada *split data* 80:20 dengan akurasi 75%;
3. model *Naive Bayes* dengan *Information Gain* pada *split data* 90:10 menghasilkan akurasi 64%, *split data* 80:20 menghasilkan akurasi 74%, serta pada *split data* 70:30 menghasilkan akurasi 64%. Dapat disimpulkan bahwa model *Naive Bayes* kombinasi *Information Gain* merupakan hasil terbaik pada *split data* 80:20 dengan akurasi 74%;

4. perhitungan akurasi klasifikasi sentimen menghasilkan akurasi sebesar 73% *Naive Bayes*, 75% *Naive Bayes + Chi Square* dan 74% *Naive Bayes + Information Gain*. Metode klasifikasi terbaik menggunakan model *Naive Bayes + Chi Square* pada *split data* 80:20 dengan akurasi sebesar 75%;
5. tingkat ukuran performa model yang digunakan termasuk ke tingkat *Fair Classification* berdasarkan tingkat nilai diagnosa *Receiver Operating Characteristic (ROC)* yang sudah tersemat pada **sub-bab 2.6.4**.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, beberapa saran yang dapat menjadi perhatian bagi pembaca untuk menghadapi tantangan yang mungkin muncul seiring dengan topik yang telah dibahas sebelumnya, yaitu:

1. penelitian ini menggunakan proses *labelling* sentimen oleh pakar bahasa, sehingga penelitian selanjutnya diharapkan untuk menggunakan proses *labelling* dengan *lexicon based* untuk memperoleh efisiensi waktu yang optimal;
2. melakukan penelitian dengan algoritma klasifikasi yang lain seperti *Support Vector Machine (SVM)*;
3. melakukan uji coba rasio *split data* yang lain serta menambahkan *parameter random state* serta *cross-validation* dalam *code Python* pada proses *splitting dataset* untuk memperoleh hasil akurasi model yang terbaik.