

BAB 5 KESIMPULAN DAN SARAN

Bab ini membahas kesimpulan dan saran dari berdasarkan penelitian dan pengujian untuk sistem mendeteksi kata kasar di Twitter menggunakan *Information Gain* dan tidak menggunakan informasi klasifikasi dengan SVM.

5.1 Kesimpulan

Bab ini membahas kesimpulan dan saran dari berdasarkan penelitian dan pengujian untuk sistem mendeteksi kata kasar di Twitter disimpulkan sebagai berikut:

1. *Stemming* dan tidak menggunakan *Stemming* sangat mempengaruhi akurasi data tetapi saat menggunakan *Stemming* menghasilkan akurasi paling tinggi 93.1402% lebih baik dibanding tidak menggunakan *Stemming* dengan akurasi paling tinggi 92.3780%.
2. Dari hasil pengujian yang diuji, menggunakan *Information Gain* mendapatkan hasil akurasi dan nilai *f-measure* paling tinggi dengan menggunakan *kernel* linear . Nilai akurasi dan nilai *f-measure* paling tinggi yang didapat adalah sebesar 93.1402% dan *f-measure* 90.7787% sedangkan menggunakan TF - IDF mendapatkan akurasi sebesar 93.0640% dan *f-measure* 90.6858% linear perbedaannya hanya 0.0762% saja. Tetapi karena akurasi keduanya sudah mencapai 90% kenaikan 0.0762% saja sudah sangat besar.
3. Penggunaan *kernel polynomial* menjadi *kernel* yang memiliki nilai akurasi dan *f-measure* yang paling kecil untuk penelitian ini. Nilai akurasi dan nilai *f-measure* paling kecil di dapat oleh *kernel polynomial* bernilai 61.0518% dan 0% pada model Dataset Original No *Stemming* (TF-IDF) dan Dataset Original No *Stemming* (IG).
4. Kombinasi Parameter C terbaik adalah 3 karena jarak *hyperlane* tidak terlalu jauh dan tidak terlalu dekat. *Gamma* terbaik berada di 1 karena radius kesamaannya tidak terlalu kecil dan tidak terlalu besar.

5.2 Saran

Saran untuk pengembangan sistem mendeteksi kata kasar di Twitter kedepannya adalah:

1. Perbedaan jumlah dataset menggunakan kata kasar dan tidak menggunakan

kata kasar pada penelitian selanjutnya disarankan menggunakan dataset yang memiliki jumlah kelas yang setara.

2. Untuk mengatasi kata dengan banyak arti. Menggunakan jenis ekstraksi fitur yang lain seperti Algoritma *decision tree*, karena dapat merepresentasikan kata dengan lebih terstruktur dan multilabel [23].
3. Media sosial akan makin berkembang terus dan akan muncul kata atau istilah baru begitu juga penggunaan kalimat yang menggunakan satir dan sarkas yang bermaksud untuk mengumpat maka dari itu untuk penelitian selanjutnya harus memperbarui kata atau istilah dan mengklasifikasikan kalimat supaya bisa dimengerti sistem.
4. Untuk parameter *gamma* tidak ada aturan khusus untuk nilai parameter yang akan digunakan tetapi penggunaan nilai parameter disarankan untuk tidak terlalu berdekatan supaya tidak menyebabkan *overfitting* karena untuk nilai *gamma* yang tinggi akan membuat titik-titik *support vector* sangat dekat satu sama lain lebih baik seperti 0.001, 0.01, 0.1, 1, 10, 100.