

BAB V

SIMPULAN DAN SARAN

5.1. Simpulan

Berdasarkan penelitian yang telah dilakukan dalam melakukan pembangunan model klasifikasi bot sosial *twitter/x* dengan *semi-supervised machine learning* dan *support vector machine*, terdapat beberapa hal yang dapat disimpulkan:

- 1) Pada penelitian ini telah dibangun model klasifikasi yang dapat mengklasifikasikan apakah sebuah *post twitter/x* adalah sebuah *post* dari bot sosial atau pengguna asli. Model klasifikasi yang dibangun dilakukan pengukuran kinerjanya dengan penilai performa algoritma *confusion matrix*. Skor *accuracy*, *precision*, *recall*, dan *f1 score* digunakan untuk mengukur kinerja dari model. Setelah dilakukan pengukuran kinerja terhadap model tersebut, ditemukan bahwa dalam pelatihan model, model klasifikasi SVM yang dilatih dengan data *pseudo-labeled*, saat diuji berperforma lebih baik dari model klasifikasi SVM yang hanya dilatih dengan data berlabel. Model klasifikasi SVM yang dilatih oleh data *pseudo-labeled* tersebut menghasilkan skor akurasi 92%. Sebuah peningkatan dari model klasifikasi SVM yang dilatih hanya dengan data berlabel, dengan nilai skor akurasi 83%.
- 2) Pengujian lebih lanjut dilakukan. Model klasifikasi SVM diuji dengan pelatihan model klasifikasi menggunakan pembagian rasio data yang berbeda-beda. Dari pengujian tersebut ditemukan bahwa performa model klasifikasi SVM yang dilatih dengan dataset *pseudo* mengalami peningkatan tingkat akurasi. Dengan model tersebut menghasilkan nilai skor *accuracy* tertinggi 93%, pada pembagian data 60-40. Model klasifikasi SVM yang dilatih dengan dataset berlabel pun mengalami kenaikan tingkat akurasi dengan pembagian data 90-10, dimana model klasifikasi tersebut terjadi kenaikan nilai skor *accuracy* dari 83% menjadi 91%. Selanjutnya dilakukan pengujian dengan penggunaan *hyperparameter C* dari algoritma SVM. Setelah pengujian telah dilakukan, ditemukan bahwa model klasifikasi SVM yang dilatih dengan dataset *pseudo* secara keseluruhan berkinerja lebih baik daripada model klasifikasi SVM yang dilatih dengan data

berlabel, dengan penyetelan *hyperparameter* $C=100$ menghasilkan skor akurasi tertinggi. Dengan skor *accuracy* dari model klasifikasi SVM yang dilatih dengan dataset *pseudo* menghasilkan skor *accuracy* 94%, yaitu nilai skor *accuracy* tertinggi dari semua pengujian.

5.2. Saran

Berdasarkan penelitian yang telah dilakukan, berikut merupakan beberapa saran yang penulis berikan terhadap penelitian ini:

- 1) Hasil dari *confusion matrix*, kinerja, dan performa dari model yang dibangun masih belum sempurna. Performa dari model klasifikasi masih bisa dioptimalkan dengan memaksimalkan kualitas dataset yang digunakan. Kualitas dataset yang bagus dan penambahan fitur yang relevan dapat meningkatkan kinerja dari model.
- 2) Pada penelitian ini konten fitur untuk mengklasifikasikan apakah sebuah *tweet/post* berasal dari sebuah pengguna asli atau bot sosial masih bersifat kurang. Penelitian berikutnya dapat ditingkatkan dengan mengfaktorkan faktor lain pada *twitter/x* untuk digunakan sebagai fitur seperti usia akun, foto profil, jumlah *tweet/post* dalam suatu jangka waktu, dan lain-lainnya.
- 3) Pada penelitian ini, digunakan satu algoritma klasifikasi yaitu *support vector machine*. Untuk penelitian kedepannya dapat digunakan algoritma-algoritma lainnya dan juga memanfaatkan metode *semi-supervised machine learning* lainnya.
- 4) Penelitian ini melakukan klasifikasi yang bersifat dua dimensi yaitu bot atau bukan bot. Penelitian selanjutnya bisa dilakukan klasifikasi yang bersifat multi-dimensi untuk meningkatkan ruang lingkup penelitian.