

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini bertujuan membantu memprediksi keaslian akun Twitter tertentu dengan mengembangkan sebuah *website* yang dapat diakses oleh pengguna umum. Setelah menyelesaikan seluruh tahapan yang ada mulai dari *scrape* data, pengolahan data dengan menggunakan Panda, eksplorasi data, melihat distribusi data dengan visualisasi, pelabelan kelas, perancangan model, pembobotan klasifikasi, hingga pembuatan *website* pengklasifikasi *buzzer*, diperoleh beberapa poin sebagai berikut :

1. Dengan pendekatan *data mining*, yaitu klasifikasi dengan menggunakan Algoritma Naïve Bayes, didapatkan pola atau *pattern* dari akun-akun *buzzer* dengan karakteristik umur akun, sentimen akun, jumlah *following* dan jumlah *followers*.
2. Algoritma Naïve Bayes varian Gaussian digunakan karena dataset memiliki sifat *continuous* dan klasifikasi yang dilakukan pada penelitian ini merupakan klasifikasi biner untuk kelas *buzzer* dan *non-buzzer*.
3. Atribut yang digunakan dalam penelitian ini antara lain, umur akun, nilai *mean* dari nilai *compound* akun, jumlah *following*, dan jumlah *followers*.
4. Model Naïve Bayes yang dibuat telah diuji dengan metode pengujian yang sesuai. Hasil pengujian menunjukkan bahwa model Naïve Bayes yang dibuat memiliki nilai *accuracy* sebesar 0.80 atau 80% sesuai Gambar 4.27.
5. Berdasarkan hasil grafik *ROC curve*, *confusion matrix*, dan *precision recall value* visualisasi hasil pengujian model, ditentukan bahwa skenario 2 yang

telah dilakukan penyeimbangan kelas terlebih dahulu pada dataset memiliki performa model lebih baik.

6. Uji validasi yang dilakukan pada *website* pengklasifikasi *buzzer* menghasilkan kesalahan dalam prediksi sebanyak 7 dari 70 data sehingga hasil akurasi dari uji validasi pada *website* sebesar 0.90 atau 90%.

5.2 Saran

Dalam Penelitian ini terdapat beberapa hal yang dapat dijadikan bahan pembelajaran, hal tersebutlah yang dapat dilakukan agar hasil dari penelitian ini menjadi lebih baik. Berikut adalah beberapa saran yang direkomendasikan dari penelitian kali ini.

1. Bagi Penelitian Selanjutnya

Gunakan atribut lain dalam dataset, seperti frekuensi *tweet* akun per hari, jumlah URL yang ada pada tiap *tweet*, jumlah *retweet* dari 100 *tweet* terakhir (Ibrahim, dkk 2015). Penggunaan atribut yang lebih beragam akan lebih baik untuk meningkatkan akurasi model. Hal ini dikarenakan atribut yang lebih beragam akan memberikan informasi yang lebih banyak dan dapat meningkatkan kemampuan model dalam menangani data yang lebih kompleks. Selain itu, variasi dataset *buzzer* dan *non-buzzer* dari topik lain juga dapat digunakan dalam proses *training* untuk mencegah terjadinya *overfitting* terhadap satu topik tertentu.

2. Bagi organisasi bisnis

Penelitian yang sudah dilakukan menunjukkan bahwa *buzzer* dapat dikenali melalui karakteristik atau perilakunya dalam *posting* pada Twitter. Sehingga, organisasi bisnis harus lebih bijak untuk mempertimbangkan

penggunaan strategi *buzz marketing* dalam menjalankan kampanye organisasi bisnisnya, alih-alih digemari masyarakat, bisa saja masyarakat malah meninggalkan *brand* karena tingkat kepercayaannya menurun.