

## **BAB 4**

### **PENUTUP**

#### **4.1 Kesimpulan**

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penggunaan algoritma *Naïve Bayes Classifier* dengan pembobotan fitur menggunakan TF-IDF (*Term Frequency – Inverse Document Frequency*) dapat mengklasifikasikan sentimen publik terhadap peristiwa Bitcoin Halving secara efektif. Dengan menggunakan data dari 538 tweet yang diperoleh melalui proses *crawling* di media sosial *Twitter*, dan setelah dilakukan berbagai tahap preprocessing serta pembagian data untuk pelatihan dan pengujian model, penelitian ini menunjukkan hasil yang signifikan. Tiga rasio data latih dan data uji yang berbeda 90:10, 80:20, dan 80:20 digunakan dalam pengujian model. Hasilnya menunjukkan bahwa model *Naïve Bayes* dengan rasio 80:20 memberikan akurasi terbaik yaitu 74% dari beberapa model rasio lainnya. Ini menunjukkan bahwa rasio data latih dan uji yang cukup beragam memberikan model untuk membaca dan mempelajari dataset dengan baik sehingga performa yang dihasilkan akurasi yang cukup optimal. Dan pada penelitian ini berhasil mendapatkan akurasi serta menunjukkan bahwa metode *text mining*, khususnya algoritma *Naïve Bayes* dengan TF-IDF, dapat diterapkan dengan baik dalam analisis sentimen publik terhadap topik yang sedang trending di media sosial. Hasil ini dapat digunakan sebagai dasar untuk penelitian lebih lanjut dalam bidang analisis sentimen serta pengembangan model yang lebih kompleks untuk meningkatkan akurasi dan performa klasifikasi.

#### **4.2 Saran**

Berdasarkan hasil penelitian, beberapa rekomendasi dapat dipertimbangkan untuk meningkatkan kinerja model klasifikasi teks peristiwa bitcoin halving di *Twitter*. Pertama, Mengatasi ketidakseimbangan kelas melalui teknik *oversampling* dan *undersampling* mungkin bermanfaat. Kedua, memperbesar dataset sehingga model klasifikasi dapat belajar lebih banyak lagi dalam memahami suatu pola untuk melakukan klasifikasi serta. Ketiga, Penggunaan

metode ekstraksi fitur alternatif seperti *word embeddings* bisa membantu menangkap makna kata dengan lebih baik. Keempat, Membandingkan performa dengan algoritma klasifikasi lain seperti *Gradient Boosting* atau *Random Forest* dapat memberikan wawasan baru. Kelima, Peningkatan proses preprocessing, terutama untuk bahasa informal, juga bisa menjadi area perbaikan. Memperbesar dataset dan menerapkan *cross-validation* mungkin akan meningkatkan generalisasi dan akurasi estimasi performa model. Implementasi dari saran-saran ini diharapkan dapat meningkatkan akurasi dan keseimbangan model secara signifikan.