

## **BAB V**

### **Kesimpulan**

#### **5.1 Kesimpulan**

Kesimpulan dari analisis sentimen terhadap polusi udara di Jakarta menunjukkan bahwa mayoritas sentimen pengguna Twitter terkait isu ini bersifat negatif, baik dalam proses pelabelan otomatis berbasis Lexicon maupun dengan menggunakan metode SVM. Data preprocessing yang dilakukan, mencakup langkah-langkah seperti data cleaning, case folding, normalisasi, penghapusan stopword, stemming, dan pembobotan menggunakan TF-IDF, berhasil mempersiapkan dataset untuk analisis lebih lanjut.

Metode Support Vector Machine (SVM) terbukti efektif dalam klasifikasi sentimen dengan akurasi tinggi mencapai 94%. Namun, meskipun model menunjukkan performa yang sangat baik untuk sentimen negatif (precision 94%, recall 99%, F1-score 97%), performa terhadap sentimen positif masih rendah (precision 80%, recall 25%, F1-score 38%), yang menunjukkan adanya ketimpangan dalam deteksi sentimen positif. Distribusi sentimen secara keseluruhan memperkuat dominasi sentimen negatif (1795 tweet) dibandingkan sentimen positif (200 tweet), mencerminkan persepsi negatif yang kuat terhadap isu polusi udara di Jakarta.

#### **5.2 Saran**

Berikut saran yang bisa diterapkan untuk penelitian berikutnya:

1. Sebelum melanjutkan ke tahap pemrosesan, disarankan untuk menerjemahkan komentar bahasa asing. Langkah ini penting untuk memahami konten komentar yang menggunakan campuran bahasa, sehingga dapat meningkatkan ketepatan dalam analisis sentimen atau klasifikasi teks. Terjemahan ini akan membantu memastikan bahwa setiap kata atau frasa ditafsirkan dengan benar, yang pada akhirnya berkontribusi pada hasil analisis yang lebih akurat dan relevan.
2. Untuk meningkatkan kualitas model dan menghindari overfitting, disarankan untuk melakukan pengumpulan data dalam jumlah yang lebih besar. Hal ini

dapat membantu mengatasi ketidakseimbangan yang signifikan antara sentimen negatif dan positif dalam confusion matrix. Dengan memperbanyak dataset, peneliti dapat meningkatkan representativitas sampel dan memperluas variasi data yang digunakan dalam proses pelatihan dan pengujian model.

Peningkatan jumlah data akan berdampak positif pada tahap pembagian data (data splitting) dalam metode Support Vector Machine. Semakin besar volume data yang tersedia, semakin banyak pula data yang dapat dialokasikan untuk set pelatihan dan pengujian. Hal ini memungkinkan model untuk belajar dari pola yang lebih beragam dan kompleks, sekaligus memberikan evaluasi yang lebih akurat terhadap performa model pada data yang belum pernah dilihat sebelumnya.

Lebih lanjut, memperbesar dataset dapat berkontribusi pada pengurangan bias dan peningkatan kemampuan generalisasi model. Hasilnya, model yang dikembangkan diharapkan memiliki ketahanan lebih tinggi dan mampu menghasilkan prediksi yang lebih tepat untuk berbagai jenis masukan, termasuk skenario-skenario yang mungkin jarang ditemui dalam kumpulan data yang lebih terbatas. Sebelum diproses lebih lanjut, disarankan untuk menerjemahkan komentar yang berbahasa asing. Hal ini akan membantu memahami isi komentar yang menggunakan campuran bahasa dan meningkatkan akurasi dalam analisis sentimen atau klasifikasi teks.

3. Mencoba pendekatan alternatif dalam pembobotan kata atau klasifikasi menjadi langkah penting untuk mengeksplorasi berbagai metode analisis data yang lebih efisien dan akurat. Dengan mencoba berbagai strategi baru, penulis dapat menemukan cara-cara yang lebih optimal dalam memproses dan menganalisis data. Hal ini penting terutama dalam konteks data yang kompleks dan terus berkembang, di mana teknik konvensional mungkin tidak selalu memberikan hasil yang maksimal.