

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan tahapan analisis yang sudah dilakukan dan hasil yang sudah didapatkan, diperoleh hasil sebagai berikut :

1. Penelitian ini berhasil menerapkan analisis sentimen terhadap komentar pengguna *YouTube* mengenai pendaftaran CPNS tahun 2024 melalui pendekatan *text mining* dan pembelajaran mesin. Tahapan prapemrosesan teks yang meliputi *cleaning*, *case folding*, *normalisasi*, *tokenisasi*, *stopword removal*, dan *stemming* mampu meningkatkan kualitas data, sehingga komentar yang dianalisis menjadi lebih terstruktur dan representatif untuk proses klasifikasi sentimen positif dan negatif.
2. Berdasarkan hasil pengujian dan evaluasi model, terdapat perbedaan kinerja antara algoritma *Naïve Bayes Classifier* dan *Support Vector Machine* dalam mengklasifikasikan sentimen komentar *YouTube*. *Support Vector Machine* dengan kernel *Radial Basis Function (RBF)* menunjukkan performa yang lebih baik dibandingkan *Multinomial Naïve Bayes*, dengan nilai akurasi masing-masing sebesar 85,16% dan 74,48%. Hasil ini mengindikasikan bahwa *SVM* memiliki kemampuan yang lebih optimal dalam menangani pola data teks yang bersifat kompleks dan tidak linier.
3. Penerapan pembobotan fitur menggunakan *TF-IDF* serta penanganan ketidakseimbangan kelas melalui teknik *Synthetic Minority Over-sampling Technique (SMOTE)* memberikan kontribusi positif terhadap peningkatan stabilitas dan keandalan model. Kombinasi metode tersebut membantu model dalam merepresentasikan kata-kata yang relevan dan mengenali sentimen pada kelas minoritas secara lebih proporsional, sehingga hasil klasifikasi yang diperoleh menjadi lebih objektif dan konsisten.

5.2 Saran

Berdasarkan hasil penelitian dan evaluasi model, terdapat sejumlah rekomendasi yang dapat dipertimbangkan pada penelitian selanjutnya maupun untuk pengembangan sistem analisis sentimen yang lebih adaptif dan akurat, sebagai berikut :

1. Penelitian berikutnya disarankan untuk menggunakan variasi teknik prapemrosesan yang lebih mendalam, seperti stemming berbasis machine learning, deteksi sarkasme, serta normalisasi kata tidak baku pada bahasa informal yang sering muncul pada komentar YouTube. Teknik tersebut berpotensi meningkatkan representasi fitur dan mengurangi noise yang memengaruhi performa model.
2. Akan lebih baik jika penelitian lanjutan melibatkan lebih banyak sumber data dari platform digital lain agar cakupan analisis menjadi lebih luas dan beragam. Data dari *Twitter*, *TikTok*, atau forum publik lainnya dapat memberikan gambaran sentimen masyarakat yang lebih komprehensif dan tidak terbatas pada satu kanal interaksi.
3. Model dapat dikembangkan menggunakan algoritma berbasis deep learning, seperti LSTM, BiLSTM, atau Transformer (BERT, IndoBERT, IndoBART) yang memiliki kemampuan mengekstraksi konteks secara lebih mendalam dibandingkan model berbasis fitur seperti TF-IDF. Teknologi tersebut sangat relevan untuk data berbahasa Indonesia yang kaya variasi morfologi dan konteks.