

BAB V

PENUTUP

5.1 Kesimpulan

Penelitian ini menganalisis sentimen komentar YouTube terkait kebijakan Upah Minimum Indonesia menggunakan algoritma Support Vector Machine (SVM) dan Logistic Regression. Proses preprocessing mencakup text cleaning, case folding, normalisasi, tokenizing, stopwords removal, dan stemming untuk memastikan data siap untuk pemodelan. Data dibagi menggunakan dua metode, yaitu K-Fold Cross Validation dan split 80:20, untuk membandingkan konsistensi dan performa model. Hasil evaluasi berdasarkan akurasi, precision, recall, dan F1-score menunjukkan bahwa SVM dengan kernel RBF dan Logistic Regression memberikan performa klasifikasi yang paling optimal dan konsisten.

Berdasarkan hasil penelitian, dilakukan proses hyperparameter tuning menggunakan GridSearchCV untuk memperoleh kombinasi parameter terbaik dalam rangka meningkatkan performa model. Pada model SVM dengan kernel Radial Basis Function (RBF), tuning dilakukan terhadap parameter C dan gamma, dan diperoleh kombinasi terbaik dengan nilai $C = 1$ dan gamma = 1, yang menghasilkan akurasi tertinggi sebesar 80,47% pada skema K-Fold Cross Validation, serta 77,55% pada skema Split Data 80:20. Model SVM dengan kernel Sigmoid dan Polynomial juga melalui proses tuning serupa untuk mendapatkan parameter optimal masing-masing. Sementara itu, pada model Logistic Regression, tuning dilakukan terhadap parameter C dan penalty, dengan kombinasi terbaik $C = 1$ dan penalty = 'l2', yang menghasilkan akurasi sebesar 80,26% pada K-Fold Cross Validation dan 78,41% pada skema Split Data 80:20.

Secara keseluruhan, algoritma SVM, khususnya kernel RBF, menunjukkan performa terbaik dalam menganalisis sentimen publik terhadap kebijakan Upah Minimum Indonesia berdasarkan validasi K-Fold Cross Validation. Namun, pada metode pembagian data Split 80:20, Logistic Regression mencatat akurasi sedikit lebih tinggi dibandingkan SVM RBF. Ini menunjukkan bahwa meskipun SVM RBF unggul pada validasi silang, Logistic Regression tetap menjadi alternatif

kompetitif dan stabil dalam konteks pembagian data berbeda.

5.2 Saran

Meskipun algoritma SVM dengan kernel RBF memberikan hasil yang paling unggul dalam penelitian ini, masih terdapat ruang untuk peningkatan di masa mendatang. Penelitian lanjutan disarankan untuk mencoba berbagai algoritma lain seperti Random Forest, Naive Bayes, atau pendekatan deep learning seperti BERT dan LSTM, yang mungkin memberikan hasil yang lebih akurat. Penggunaan metode ensemble learning atau penggabungan beberapa model juga dapat menjadi strategi untuk meningkatkan performa klasifikasi. Dari sisi praproses data, metode lemmatization bisa dipertimbangkan sebagai pendekatan yang lebih baik dari stemming, dan analisis sentimen juga dapat dikembangkan lebih jauh dengan menambahkan fitur berbasis aspek atau emosi guna memperoleh wawasan yang lebih kaya terhadap opini publik. Di samping itu, penerapan rasio pembagian data yang berbeda, seperti 70:30 atau 90:10, serta penggunaan validasi silang dengan 10-Fold Cross Validation, berpotensi memberikan hasil evaluasi model yang lebih akurat dan menyeluruh. Ruang lingkup penelitian ini juga dapat diperluas dengan mengumpulkan data komentar dari berbagai video YouTube atau platform lain seperti Twitter dan Facebook, sehingga data yang dianalisis mencerminkan opini publik secara lebih luas. Dengan pendekatan ini, studi di masa mendatang diharapkan dapat menghasilkan temuan yang lebih komprehensif dalam menganalisis sentimen masyarakat terhadap isu-isu kebijakan nasional.