

BAB V PENUTUP

5.1 Kesimpulan

Penelitian ini berhasil menganalisis sentimen masyarakat terhadap Kementerian PUPR pasca ditetapkannya UU tapera dengan menggunakan algoritma *BERT*. Data yang digunakan dalam penelitian ini diperoleh dari beberapa *channel* pada platform *YouTube*. Data tersebut dikumpulkan melalui proses *scraping* menggunakan *YouTube Data API* dan *Google Colab*, menghasilkan 15.984 baris komentar. Setelah melewati berbagai tahapan, mulai dari pembersihan data (*cleaning*) hingga praproses (*preprocessing*), jumlah data yang siap digunakan untuk pelatihan model adalah sebanyak 14.332 baris komentar. Penelitian ini memiliki fokus utama yaitu untuk mengeksplorasi bagaimana pendekatan *NLP* dapat dimanfaatkan untuk memahami persepsi publik terhadap kebijakan pemerintah yang berdampak signifikan. Penelitian ini menggunakan dua skema dan dengan jumlah dataset 6.000, 10.000, dan 14.000 dalam tahapan klasifikasi. Pada skema pertama, tidak diterapkan metode *SMOTE* maupun *k-fold cross-validation*. Hasil dari penerapan skema pertama menunjukkan bahwa performa model pada tahap pelatihan mengalami *overfitting*, dengan tingkat kesalahan yang cukup tinggi dalam membedakan antara label negatif dan positif dengan tingkat akurasi sebesar 88%, 86%, dan 85%. Sementara itu, pada skema kedua, diterapkan metode tambahan berupa *SMOTE* dan *k-fold cross-validation* untuk meningkatkan kualitas model serta mengatasi tantangan akibat ketidakseimbangan data. Penerapan skema ini berhasil mengatasi *overfitting* dan secara signifikan meningkatkan akurasi model dalam membedakan label negatif dan positif, dengan tingkat akurasi yang mencapai 99% secara keseluruhan. Dari hasil penelitian memberikan pandangan bahwa penerapan algoritma *BERT* yang dikombinasikan dengan teknik pengolahan data seperti *SMOTE* dan *k-fold cross validation* dalam melakukan analisis sentimen terhadap kebijakan pemerintah sangatlah efektif. Melalui pendekatan yang komprehensif ini, diharapkan penulis tidak hanya memberikan kontribusi pada pemahaman mengenai sentimen masyarakat terhadap kinerja kementerian PUPR dan kebijakan tapera, tetapi juga memberikan partisipasi bagi pengembangan metode analisis yang lebih efisien dan efektif dalam konteks penelitian terhadap

sosial. Serta diharapkan juga bisa menjadi referensi bagi peneliti lain dalam bidang analisis sentimen lebih lanjut.

5.2 Saran

Berdasarkan hasil dan kesimpulan dari penelitian ini, peneliti memberikan beberapa saran yang diharapkan dapat diterapkan dalam penelitian selanjutnya untuk meningkatkan kualitas penelitian, beberapa saran yang dapat penulis berikan sebagai berikut:

1. Disarankan dapat menggunakan dataset yang lebih beragam dengan sumber data dari berbagai platform media sosial lainnya. Hal ini bertujuan untuk bisa mendapatkan persepsi dari publik atau masyarakat yang lebih luas dan representatif terhadap sentimen masyarakat terhadap kebijakan pemerintah.
2. Untuk memastikan bahwa hasil penelitian lebih generalis, disarankan pada penelitian selanjutnya dapat menggunakan data tambahan untuk validasi model secara eksternal. Hal ini dapat membantu untuk mengevaluasi performa dari model pada data yang baru.
3. Untuk mendukung penelitian dengan dataset yang lebih besar, disarankan untuk memanfaatkan infrastruktur komputasi berbasis *cloud* dengan *GPU* yang lebih kuat agar proses pelatihan model lebih efisien.
4. Mempertimbangkan metode atau cara alternatif lainnya selain *SMOTE* untuk mengatasi ketidakseimbangan data, seperti *ADASYN (Adaptive Synthetic Sampling)* yang dapat menangani ketidakseimbangan secara adaptif dengan lebih fokus pada sampel minoritas yang sulit, atau seperti *contextual augmentation* berbasis *BERT* untuk memperkaya data pelatihan dan memperbaiki generalisasi model.
5. Penyesuaian yang lebih lanjut mengenai penggunaan *hyperparameter*, dapat Mengeksplorasi *tuning hyperparameter* lebih lanjut, seperti ukuran *batch*, *learning rate*, atau *jumlah epoch*, untuk memaksimalkan performa model.