

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian mengenai analisis sentimen opini masyarakat terhadap kenaikan harga BBM pada media sosial X (Twitter) menggunakan algoritma *Logistic Regression*, maka dapat disimpulkan sebagai berikut:

1. Berdasarkan hasil penelitian, opini masyarakat mengenai perubahan harga BBM pada media sosial X dapat dikelompokkan ke dalam tiga kategori sentimen, yaitu negatif, netral, dan positif. Hasil pelabelan data menunjukkan bahwa sentimen negatif memiliki jumlah data paling besar dibandingkan dengan sentimen netral dan positif. Dengan demikian, opini masyarakat yang terdapat dalam dataset penelitian ini secara dominan menunjukkan kecenderungan sentimen negatif.
2. Algoritma *Logistic Regression* dapat diterapkan untuk melakukan klasifikasi sentimen pada data opini masyarakat setelah melalui tahapan preprocessing dan ekstraksi fitur menggunakan TF-IDF. Representasi TF-IDF digunakan untuk mengubah data teks menjadi bentuk numerik yang dapat diproses oleh algoritma *Logistic Regression*. Dalam proses pemodelan, digunakan `class_weight="balanced"` untuk membantu menangani ketidakseimbangan jumlah data pada setiap kelas sentimen.
3. Penelitian ini berhasil mengumpulkan 2.262 cuitan melalui proses scraping menggunakan platform Apify. Setelah kumpulan data disaring, jumlah cuitan yang memenuhi kriteria penelitian berkurang menjadi 1.793. Data tersebut kemudian diproses melalui serangkaian langkah prapemrosesan,

termasuk cleaning, tokenizing, stopword removal, text normalization, dan stemming sehingga menghasilkan 1.791 cuitan yang siap dianalisis.

4. Metode TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan untuk mengubah data teks menjadi representasi numerik agar dapat diproses oleh algoritma klasifikasi. Proses ini menghasilkan 5.000 fitur, yang kemudian digunakan sebagai masukan untuk algoritma *Logistic Regression*.
5. Dataset dibagi menggunakan rasio 80:20, sehingga diperoleh 1.432 data training dan 359 data testing untuk proses pelatihan dan pengujian model.
6. Hasil klasifikasi yang menggunakan algoritma *Logistic Regression* menunjukkan tingkat akurasi sebesar 77,72%, dengan nilai rata-rata weighted average precision, recall, dan F1-score masing-masing sebesar 0,78. Hasil ini menunjukkan bahwa model tersebut berkinerja cukup baik dalam mengklasifikasikan sentimen publik terkait kenaikan harga bahan bakar.
7. Berdasarkan hasil analisis sentimen, mayoritas opini publik terkait kenaikan harga bahan bakar masuk ke dalam kategori sentimen negatif, dengan 1.358 data, sedangkan sentimen netral berjumlah 238 data dan sentimen positif berjumlah 195 data. Hasil ini menunjukkan bahwa sebagian besar pengguna platform media sosial X (Twitter) menanggapi kebijakan kenaikan harga bahan bakar secara negatif.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya, yaitu sebagai berikut.

1. Penelitian selanjutnya disarankan untuk menggunakan kumpulan data yang lebih besar dan memperpanjang periode pengumpulan data, sehingga hasil analisis sentimen dapat memberikan gambaran yang lebih komprehensif mengenai opini publik.
2. Penelitian selanjutnya disarankan untuk menerapkan teknik-teknik dalam menangani data yang tidak seimbang seperti oversampling, undersampling, atau metode lainnya agar kinerja model dalam mengklasifikasikan kategori sentimen dengan jumlah data yang lebih sedikit dapat ditingkatkan.
3. Penelitian selanjutnya dapat membandingkan algoritma Logistic Regression dengan algoritma klasifikasi lainnya, seperti *Support Vector Machines* (SVM), *Naïve Bayes*, *Random Forest*, dan metode berbasis *Deep Learning*, untuk mengidentifikasi algoritma yang memberikan kinerja terbaik dalam analisis sentimen terkait kenaikan harga bahan bakar.
4. Penelitian selanjutnya dapat mengeksplorasi penggunaan metode ekstraksi fitur selain TF-IDF, seperti Word2Vec, FastText, atau BERT, dengan tujuan meningkatkan kualitas representasi teks dan kinerja klasifikasi sentimen.
5. Hasil penelitian ini diharapkan dapat menjadi acuan bagi para peneliti di masa mendatang dalam mengembangkan penelitian yang berkaitan dengan analisis sentimen, text mining, *Natural Language Processing* (NLP), serta penerapan algoritma *Logistic Regression* dalam klasifikasi data teks.