

BAB V

PENUTUP

5.1 Simpulan

Penelitian ini berhasil menerapkan model klasifikasi teks untuk memahami gambaran umum sentimen pengguna media sosial X terhadap calon presiden dan wakil presiden (Capres dan Cawapres) dalam pemilu 2024. Data dikumpulkan melalui crawling selama 28 November 2023 - 10 Februari 2024, dan label kelas diberikan menggunakan metode berbasis leksikon. Visualisasi kata dilakukan dengan wordcloud, pembobotan kata dilakukan dengan TF-IDF, dan kinerja model diukur menggunakan akurasi, presisi, recall, dan F1-score dari algoritma K-Nearest Neighbor (K-NN).

Berdasarkan hasil pembahasan klasifikasi teks data tweet menggunakan metode lexicon based, TF-IDF, K-Nearest Neighbor maka kesimpulan yang bisa ditarik pada penelitian ini sebagai berikut:

- a. Hasil pembersihan data dari tiga pasangan calon presiden dan wakil presiden menunjukkan variasi dalam sentimen yang terdeteksi. Untuk pasangan Anies dan Muhaimin, dari total 1006 tweet, tersisa 912 tweet setelah proses pembersihan. Data ini kemudian diberi label menggunakan metode berbasis leksikon, yang menghasilkan 318 tweet dengan sentimen positif, 391 tweet dengan sentimen negatif, dan 203 tweet dengan sentimen netral. Sementara itu, untuk pasangan Prabowo dan Gibran, dari 1006 tweet yang diperoleh, menjadi 962 tweet setelah pembersihan. Dengan metode pelabelan berbasis leksikon, ditemukan 239 tweet dengan sentimen positif, 459 tweet dengan sentimen negatif, dan 264 tweet dengan sentimen netral. Sedangkan untuk pasangan Ganjar dan Mahfud, dari total 1008 tweet, berhasil diproses menjadi 920

tweet. Dari data ini, ditemukan 544 tweet dengan sentimen positif, 214 tweet dengan sentimen negatif, dan 162 tweet dengan sentimen netral.

- b. Dalam analisis sentimen terhadap calon presiden dan wakil presiden menggunakan metode K-Nearest Neighbors (KNN) dengan menggunakan data 90:10, ditemukan variasi dalam akurasi berdasarkan nilai k yang digunakan. Untuk pasangan Anies dan Muhaimin, penggunaan $k=8$ menghasilkan akurasi sebesar 67,4%. Sebaliknya, pasangan Prabowo dan Gibran menunjukkan akurasi yang lebih rendah, yaitu 55,7% dengan $k=3$. Sementara itu, pasangan Ganjar dan Mahfud mencapai akurasi tertinggi sebesar 69,6% dengan nilai $k=10$. Setelah dilakukan penelitian dan membandingkannya dengan penelitian terdahulu, diketahui faktor-faktor yang mempengaruhi kinerja algoritma adalah banyaknya jumlah data, fitur pembobotan yang digunakan
- c. Berdasarkan hasil penelitian ini, penulis menyimpulkan bahwa penerapan algoritma K-Nearest Neighbor (K-NN) menggunakan metode split data 90:10 memberikan hasil yang lebih optimal dibandingkan pengujian lainnya, karena menghasilkan tingkat akurasi yang lebih tinggi.

5.2 Saran

Berdasarkan hasil penelitian, penulis memiliki beberapa saran yang bisa menjadi masukan dan bahan pertimbangan untuk penelitian selanjutnya sebagai berikut:

- a) Berdasarkan hasil analisis yang diperoleh, penelitian ini terbatas dan mengalami kendala salah satunya di tahap modify atau preprocessing data, dimana pengolahan data pada penelitian ini terbatas pada normalisasi kata yang terdapat pada kamus normalisasi Bahasa Indonesia saja. Sedangkan, di media sosial, dalam kasus ini twitter, banyak ditemukan bahasa tidak baku yang sering digunakan sehari-hari,

seperti “saya” menjadi “gue, eug, gw” dan banyak kata dengan huruf ganda, seperti “bangeeetttt, yaallaahhh” yang kemudian data seperti ini tidak ternormalisasi. Selain itu, beberapa data yang menggunakan bahasa daerah juga tidak mengalami normalisasi.

- b) Ketersediaan data sentimen dalam kamus leksikon bahasa Indonesia masih terbatas dan belum sepenuhnya lengkap, sehingga untuk penelitian berikutnya dapat mempertimbangkan penggunaan metode alternatif seperti TextBlob, dengan terlebih dahulu menerjemahkan dataset yang akan dianalisis.