BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan pada penelitian ini, diperoleh beberapa kesimpulan bahwa proses analisis sentimen terhadap isu dugaan korupsi Chromebook di Kemendikbudristek telah berhasil diterapkan menggunakan data komentar yang berasal dari platform Facebook, Instagram, dan YouTube dengan jumlah keseluruhan sebanyak 6.991 data. Data tersebut kemudian melalui beberapa proses pengolahan data teks, diawali dengan tahap pra-pemrosesan dan dilanjutkan dengan proses penentuan label sentimen. Dari rangkaian tahapan tersebut, data berhasil dikelompokkan ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral.

Memperlihatkan bahwa kelas sentimen positif memiliki jumlah komentar terbanyak dibandingkan dua kategori lainnya, yaitu sentimen negatif dan netral. Kondisi tersebut mengindikasikan bahwa komentar yang dikumpulkan memiliki variasi respons dari pengguna media sosial terhadap topik yang dianalisis. Berdasarkan metode pelabelan yang diterapkan, sebagian besar data komentar termasuk dalam kategori sentimen positif.

Berdasarkan hasil pengujian dan perbandingan algoritma klasifikasi, metode *Support Vector Machine* (SVM) menunjukkan performa yang lebih unggul dibandingkan *Multinomial Naïve Bayes* (MNB). Model SVM memperoleh nilai akurasi sebesar 82% dengan hasil evaluasi *precision*, *recall*, dan *F1-score* yang lebih stabil pada setiap kategori sentimen. Sementara itu, metode MNB memperoleh akurasi sebesar 69% dan mengalami kendala dalam mengenali kelas sentimen netral. Oleh karena itu, SVM dinilai lebih mampu dan konsisten dalam melakukan klasifikasi sentimen komentar media sosial terkait dugaan kasus korupsi Chromebook dibandingkan MNB.

Analisis kinerja model Support Vector Machine (SVM) dan Multinomial Naïve Bayes (MNB) dilakukan melalui beberapa pengukuran evaluasi klasifikasi, yaitu *accuracy*, *precision*, *recall*, *F1-score*, *macro average*, dan *weighted average*. Penggunaan beberapa metrik evaluasi tersebut dipilih karena setiap metrik memiliki fungsi yang berbeda dalam menggambarkan kemampuan model. *Accuracy* digunakan untuk mengetahui tingkat kesesuaian antara hasil prediksi model dengan label sebenarnya pada seluruh data pengujian. Sementara itu, *precision* digunakan untuk melihat tingkat ketepatan model ketika memprediksi suatu kategori tertentu, sedangkan *recall* menunjukkan kemampuan model dalam mengidentifikasi seluruh data yang memang termasuk ke dalam kategori tersebut. Nilai *F1-score* digunakan sebagai ukuran gabungan yang menunjukkan keseimbangan antara *precision* dan *recall*. Selain pengukuran pada setiap kelas, evaluasi juga memperhatikan nilai rata-rata performa melalui *macro average* dan *weighted average*. *Macro average* menghitung rata-rata hasil evaluasi dengan memberikan pengaruh yang sama pada setiap kelas, sedangkan *weighted average* memperhitungkan banyaknya data pada masing-masing kelas dalam proses perhitungan. Dengan menerapkan beberapa indikator evaluasi tersebut, performa kedua model dapat dianalisis secara lebih objektif sehingga kemampuan masing-masing algoritma dalam melakukan klasifikasi sentimen dapat diketahui dengan lebih baik.

5.2 Saran

Berdasarkan hasil penelitian yang telah diperoleh, terdapat beberapa hal yang dapat dijadikan sebagai bahan pertimbangan untuk pengembangan penelitian selanjutnya. Penelitian berikutnya dapat mempertimbangkan penggunaan dataset komentar dengan jumlah yang lebih banyak serta mencakup lebih luas sumber media sosial. Perluasan sumber data tersebut diharapkan dapat menghasilkan analisis yang lebih beragam, meningkatkan representasi opini pengguna, serta memberikan gambaran yang lebih mendekati kondisi sebenarnya.

Proses pelabelan sentimen sebaiknya tidak hanya menggunakan metode lexicon-based, tetapi juga divalidasi secara manual oleh beberapa penilai untuk

meningkatkan kualitas label dan mengurangi kemungkinan kesalahan klasifikasi. Mengingat komentar di media sosial sering mengandung sarkasme, sindiran, serta penggunaan bahasa yang tidak baku, validasi manual dapat membantu menghasilkan dataset yang lebih akurat dan berkualitas.

Selain itu, penelitian selanjutnya dapat membandingkan kinerja Support Vector Machine (SVM) dan Multinomial Naïve Bayes (MNB) dengan metode pembelajaran mesin maupun deep learning yang lebih modern. Beberapa metode yang dapat digunakan antara lain Random Forest, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), serta IndoBERT. Penggunaan berbagai metode tersebut memungkinkan peneliti untuk melakukan perbandingan performa klasifikasi sentimen secara lebih luas dari berbagai pendekatan, baik machine learning maupun deep learning.

Diharapkan penelitian ini memberikan kontribusi informasi yang bermanfaat bagi pihak terkait khususnya dalam memahami perkembangan opini publik di media sosial serta mengidentifikasi kecenderungan sentimen pengguna terhadap suatu isu tertentu, sehingga dapat menjadi salah satu sumber informasi dalam proses evaluasi maupun pengambilan keputusan.