

V. PENUTUP

5.1. Kesimpulan

Kesimpulan dari penelitian dengan judul “Implementasi SMOTE untuk Mengatasi *Class Imbalance* pada Naive Bayes *Classifier* dalam Analisis Sentimen Calon Presiden di Pemilihan Umum 2024” adalah sebagai berikut.

- a. Analisis sentimen Calon Presiden 2024 berhasil dilakukan dengan SMOTE pada Naive Bayes. Pembuatan model dilakukan menggunakan bahasa pemrograman Python pada Google Colaboratory dengan urutan pemodelan yaitu TF-IDF, SMOTE, dan Naive Bayes. Model yang menghasilkan akurasi tertinggi yaitu kelompok Uji 3 model Naive Bayes dan SMOTE (Uji 3 NB+SMOTE) dengan nilai akurasi 72,80% (Gambar 13).
- b. Implementasi SMOTE membawa pengaruh terhadap nilai akurasi yang didapatkan pada setiap pengujian. Terdapat peningkatan nilai akurasi sebesar 15-23% lebih tinggi di kelompok uji yang menggunakan implementasi SMOTE pada model Naive Bayes. Selain itu, perbedaan jumlah data yang digunakan juga memiliki pengaruh terhadap perbaikan hasil nilai akurasi pengujian sebesar 1-10% (Gambar 13).
- c. Penerapan model klasifikasi dalam aplikasi berbasis website dapat mempermudah pengguna menggunakan model untuk melakukan analisis sentimen. Aplikasi Analisis Sentimen yang berhasil dibuat mampu melakukan perhitungan, pengecekan validasi model, dan melakukan klasifikasi sentimen *tweets*. Pada perhitungan model, diketahui bahwa penggunaan SMOTE meningkatkan jumlah data kelas minoritas secara acak sehingga akurasi model dalam pengujian selalu berubah. Namun, aplikasi yang dibuat akan menggunakan model dengan akurasi terbesar (Gambar 23). Pengecekan model juga berjalan baik dengan model yang mampu melakukan klasifikasi sentimen positif dan negatif secara benar (Gambar 24-25).
- d. Hasil analisis sentimen menunjukkan bahwa Ganjar Pranowo selalu mendapatkan sentimen positif terbanyak dengan total 3102 data pada

Oktober 2023, 3481 data pada November 2023, dan 3474 data pada Desember 2023. Sementara itu, Prabowo Subianto selalu mendapatkan sentimen negatif terbanyak dengan total 1830 data pada Oktober 2023, 1575 pada November 2023, dan 1575 data pada Desember 2023 (Gambar 26).

5.2. Saran

Hal-hal yang perlu diperhatikan untuk penelitian mendatang adalah mengenai jumlah data yang digunakan. Semakin banyak data yang digunakan akan membuat pemodelan menjadi lebih baik sehingga dapat menghasilkan nilai akurasi yang lebih tinggi. Agar mendapatkan akurasi model yang cukup tinggi hingga mendekati 100%, estimasi jumlah data yang dapat digunakan dalam pengujian model kurang lebih 10000-20000 data *tweet*. Adapun untuk mengetahui bagaimana pengaruh SMOTE dengan lebih jelas, diperlukan jumlah data dengan ketidakseimbangan yang lebih tinggi pada setiap labelnya.

