

DAFTAR KODE SUMBER

4.1	Persiapan Data dan Variabel.	65
4.2	Pencarian <i>Tweet</i> Berdasarkan Kata Kunci.	66
4.3	Menggunakan Google Colab untuk mengunggah file dataset yang akan digunakan dalam analisis atau pelatihan model.	67
4.4	Membaca file CSV yang diunggah ke Google Colab dan mengubahnya menjadi DataFrame, kemudian menampilkan 5 baris pertama untuk memeriksa isi dataset	68
4.5	Memilih kolom 'full_text', 'username', dan 'created_at' dari dataset untuk analisis lebih lanjut.	68
4.6	Menghapus duplikasi dalam kolom 'full_text' untuk memastikan setiap teks hanya muncul sekali dalam dataset.	68
4.7	Menghapus baris yang mengandung nilai kosong (NaN) dari dataset untuk memastikan data yang bersih dan lengkap.	69
4.8	Menerapkan fungsi pembersihan teks untuk menghapus mentions, hashtags, <i>retweet</i> , link, angka, karakter non-alfabetik, dan kata satu huruf pada kolom 'full_text' dalam dataset.	69
4.9	Menyaring baris dengan lebih dari dua kata pada kolom 'full_text' dan menyimpan hasilnya ke file CSV.	70
4.10	Melakukan normalisasi teks pada kolom full_text dengan mengganti kata-kata slang atau informal menjadi bentuk baku menggunakan dictionary <i>norm</i>	71
4.11	Menghitung jumlah kata asing dalam teks dan memfilter baris yang mengandung lebih dari 10 kata asing.	72
4.12	Menghapus <i>stopwords</i> dari kolom full_text pada DataFrame.	73
4.13	Memisahkan setiap kalimat dalam kolom full_text menjadi daftar kata (token).	74
4.14	Melakukan <i>stemming</i> pada data yang telah ditokenisasi dan menyimpan hasilnya dalam file CSV di folder "Preprocessing".	75
4.15	Membagi dataset menjadi data latih, validasi, dan uji dengan memastikan hasil yang dapat direproduksi menggunakan seed acak dengan 90% data digunakan untuk pelatihan dan validasi, dan 10% untuk pengujian.	80

4.16	Menyiapkan label aktual dari kolom 'label' pada dataset uji (df_test) untuk evaluasi model, kemudian menampilkannya.	80
4.17	Membuat folder untuk menyimpan file CSV yang berisi data latih, validasi, dan uji, serta menghapus file lama jika sudah ada.	80
4.18	Mendefinisikan jalur file CSV untuk data latih, validasi, dan uji, kemudian memuat dataset dari file-file tersebut menggunakan load_dataset.	81
4.19	Inisialisasi tokenizer dan model IndoBERT untuk tugas klasifikasi dengan tiga kelas.	82
4.20	Inisialisasi tokenizer dan model mT5 untuk tugas klasifikasi dengan tiga kelas.	83
4.21	Kelas dataset untuk tokenisasi teks dan penyusunan data siap latih. .	86
4.22	Analisis panjang token pada dataset pelatihan menggunakan tokenizer. .	87
4.23	Mengonversi dataset teks menjadi <i>DataLoader</i> untuk kemudahan pemrosesan dalam batch saat pelatihan dan evaluasi.	88
4.24	Mengecek ketersediaan GPU dan memilih perangkat yang sesuai untuk pelatihan model.	88
4.25	Proses pelatihan dengan mode evaluasi, menghitung <i>loss</i> dan akurasi tanpa pembaruan bobot selama inferensi.	89
4.26	Menghitung akurasi dan <i>loss</i> pada data validasi tanpa pembaruan bobot selama evaluasi model.	90
4.27	Visualisasi hasil pelatihan dan validasi dalam bentuk tabel.	91
4.28	Visualisasi <i>loss</i> pelatihan dan validasi sepanjang epoch.	92
4.29	Visualisasi akurasi pelatihan dan validasi sepanjang epoch.	93
4.30	Langkah evaluasi model dengan menggunakan confusion matrix untuk menggambarkan seberapa baik model dalam memprediksi label yang benar.	94
4.31	<i>Confusion matrix</i> yang ditampilkan dalam heatmap ini menunjukkan distribusi prediksi model dibandingkan dengan label aktual. .	95
4.32	Laporan klasifikasi menunjukkan metrik evaluasi model untuk setiap kelas, termasuk <i>precision</i> , <i>recall</i> , dan <i>F1-score</i> , yang menggambarkan kinerja model dalam mengklasifikasikan data dengan lebih rinci.	96
4.33	Prediksi Sentimen Kalimat Menggunakan Model Transformer. . . .	97
4.34	Menyimpan model yang telah dilatih.	97

4.35	Pengaturan parameter pelatihan seperti jumlah epoch, learning rate, dan ukuran batch, serta mengonfigurasi optimizer AdamW untuk pelatihan model.	98
4.36	Proses pelatihan model dengan menghitung loss, akurasi, dan memperbarui parameter menggunakan optimizer.	98
4.37	Evaluasi model pada data validasi dengan menghitung <i>loss</i> dan akurasi.	100