

Nama : Alim Musyaffa

NIM: 24917020

Konsentrasi : Sains Data

1. Task

Task yang dilakukan adalah analisis sentimen terhadap data tweet menggunakan model BERT. Tujuan dari task ini adalah untuk memprediksi sentimen dari setiap tweet, yang dapat termasuk kategori positif, netral, atau negatif.

2. Dataset

Dataset yang digunakan dalam tugas ini adalah dataset tweet yang berisi beberapa kolom, antara lain:

- **Date:** Tanggal dan waktu tweet diposting
- **Sentimen:** Label sentimen yang diberikan pada tweet (positif, netral, negatif)
- **Username:** Pengguna yang memposting tweet
- **Kategori:** Kategori lain terkait tweet

Berikut merupakan contoh dari beberapa baris dari data yang saya pakai :

Date	Sentimen	Username	Kategori
Sun Aug 20 23:58:34	Komisi Pemilihan Umum, Sabtu (19/8/2023), memperbarui daftar calon sementara Pemilu Legislatif 2024 sel	hariankompas	netral
Sun Aug 20 23:57:12	Targetkan Enam Kursi di Pemilu 2024, Ketua DPC PDIP Sarolangun Sebut Semua Sudah Bergerak	https://t.co/metrojambi	netral
Sun Aug 20 23:56:45	Presiden 2024 hanya Ganjar PranowoTerbaik !! yang lainnya Bermasalah semua	#GanjarHebat #GanjarPresic ganjarist_satu	positif
Sun Aug 20 23:56:08	kontestasi berharap cawapres" versus "kandidat cawapres" Bagaimana resepsi gen Z?	https://t.co/Nlh8cZC irwan_dwi_a	netral
Sun Aug 20 23:55:29	*Minggu Kasih Kapolres Tabanan, Ajak Jemaat Sukseskan Pemilu 2024* Tabanan, 20 Agustus 2023 Bertempat HKerambitan		netral
Sun Aug 20 23:47:12	Kejagung Tunda Kasus Korupsi Terkait Capres hingga Pemilu 2024 Selesai. Sementara FORMULA E, JIS, dll A pat_sagan47286		netral

3. Preprocessing

Tahapan preprocessing yang dilakukan adalah:

3.1 Membersihkan teks:

- Menghapus URL, mention, hashtag, dan karakter khusus yang tidak relevan dengan analisis sentimen
- Mengubah semua teks menjadi huruf kecil untuk konsistensi

3.2 Encoding label kategori:

Kategori sentimen diubah menjadi angka dengan menggunakan LabelEncoder dari scikit-learn.

3.3 Splitting data:

Dataset dibagi menjadi data training, data validasi, dan data testing dengan proporsi 70%, 15%, dan 15% menggunakan fungsi `train_test_split`.

3.4 Tokenisasi:

Teks tweet di-tokenize menggunakan `AutoTokenizer` dari model `IndoBERT`. Tokenisasi ini mengubah teks menjadi format yang dapat diproses oleh model BERT, termasuk membuat `input_ids` dan `attention_mask`.

4. LLM yang Digunakan

Model yang digunakan pada proyek ini adalah `IndoBERT` model transformer pra-latih yang dioptimalkan khusus untuk bahasa Indonesia. Model ini telah terbukti lebih efektif dalam menangani tugas-tugas pemrosesan bahasa alami berbahasa Indonesia karena pada Penelitian oleh Koto (Koto et al., 2020) menunjukkan bahwa `IndoBERT` memberikan hasil yang lebih baik dalam berbagai tugas NLP bahasa Indonesia dibandingkan dengan model bahasa Inggris. Model ini telah diuji pada berbagai tugas seperti analisis sentimen, pengenalan entitas bernama, dan lainnya, dengan hasil yang baik.

5. Fine-Tuning Setup

Fine-tuning model dilakukan dengan menggunakan `Trainer` dari `Hugging Face Transformers`.

Berikut adalah beberapa pengaturan untuk fine-tuning:

- Epochs: 3 epoch.
- Batch size: 16 untuk training dan 64 untuk validasi
- Warmup steps: 500 langkah
- Weight decay: 0.01 untuk regularisasi
- Logging: Diaktifkan setiap 10 langkah dengan hasil disimpan pada folder `./logs`

Berikut hasilnya :

Epoch	Training Loss	Validation Loss	Accuracy
1	0.912400	0.923972	0.612284
2	0.813500	0.798728	0.658349
3	0.655900	0.921311	0.608445

Epoch 2 menunjukkan performa terbaik dengan accuracy tertinggi (65.83%) dan training loss yang paling rendah. Ini menunjukkan bahwa model sudah cukup baik dalam mempelajari data pelatihan dan menghasilkan prediksi yang lebih akurat pada data validasi.

Epoch 3 mengalami sedikit penurunan accuracy pada data validasi dan peningkatan validation loss, yang mengindikasikan adanya kemungkinan overfitting pada model, meskipun training loss terus menurun.

6. Hasil

Metrik Evaluasi

Berdasarkan hasil evaluasi model pada data test, berikut adalah metrik yang dihasilkan:

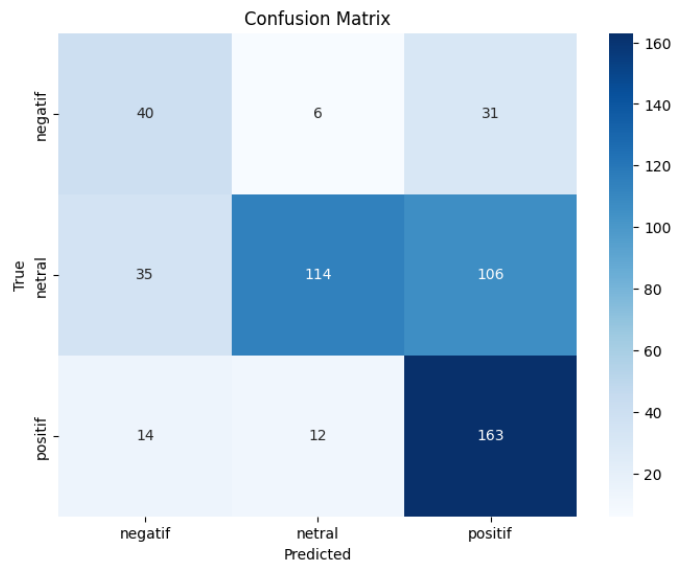
	precision	recall	f1-score	support
negatif	0.45	0.52	0.48	77
netral	0.86	0.45	0.59	255
positif	0.54	0.86	0.67	189
accuracy			0.61	521
macro avg	0.62	0.61	0.58	521
weighted avg	0.69	0.61	0.60	521

Hasil evaluasi model menunjukkan bahwa model memiliki akurasi sebesar 61% secara keseluruhan. Untuk precision, model lebih baik dalam mengidentifikasi sentimen netral dengan nilai 0.86, namun kurang akurat dalam mengidentifikasi sentimen negatif dan positif, dengan precision masing-masing 0.45 dan 0.54. Dalam hal recall, model menunjukkan kinerja terbaik pada sentimen positif dengan nilai 0.86, namun memiliki recall yang lebih rendah untuk netral (0.45) dan negatif (0.52). Nilai f1-score untuk sentimen positif adalah yang tertinggi (0.67), sedangkan sentimen negatif dan netral masing-masing memiliki f1-score 0.48 dan 0.59.

Secara keseluruhan, model menunjukkan performa yang lebih baik pada sentimen positif dan netral, tetapi masih memiliki kesulitan dalam memprediksi dengan akurat sentimen negatif. Weighted average menunjukkan nilai precision 0.69, recall 0.61, dan f1-score 0.60, menggambarkan bahwa model memberikan hasil yang lebih baik pada kategori yang lebih sering muncul dalam dataset.

Confusion Matrix

Berikut adalah Confusion Matrix yang menunjukkan perbandingan antara prediksi dan label sebenarnya dari data test:



Analisis Hasil Confusion Matrix:

Berdasarkan Confusion Matrix yang diberikan, model menunjukkan performa yang cukup baik dalam memprediksi sentimen positif, dengan 163 tweet positif yang diprediksi dengan benar. Namun, model memiliki kesulitan dalam memprediksi sentimen negatif, di mana hanya 40 tweet negatif yang diprediksi dengan benar, sementara ada 31 tweet positif dan 6 tweet netral yang salah diprediksi sebagai negatif. Untuk sentimen netral, model berhasil memprediksi 114 tweet netral dengan benar, tetapi masih ada 35 tweet negatif dan 106 tweet positif yang salah diprediksi sebagai netral. Secara keseluruhan, model lebih efektif dalam memprediksi tweet dengan sentimen positif, tetapi kesulitan dalam membedakan tweet negatif dan netral.