

Analisis Sentimen

Gemini

10 Oktober 2025



Introduction



I Made **Krisna Jaya**

Latar Belakang

Di era digital, organisasi di berbagai sektor (politik, bisnis) menghadapi tantangan besar untuk memahami opini publik yang tersebar luas di media sosial. Mereka butuh jawaban cepat atas pertanyaan krusial, seperti:

- "Bagaimana respons publik terhadap kebijakan baru?"
- "Apakah kampanye marketing kami berhasil?"
- "Apa sentimen masyarakat terhadap kandidat politik kami?"

Namun, menganalisis jutaan komentar secara manual adalah hal yang mustahil, lambat, dan mahal.

Analisis sentimen hadir sebagai solusi teknologi untuk tantangan tersebut.

Dengan memanfaatkan Natural Language Processing (NLP), kita dapat:

1. Mengotomatisasi Pengolahan Data
2. Menghasilkan Wawasan Real-Time
3. Mendukung Keputusan Strategis

Tujuan dan Cakupan

Tujuan:

- Menerapkan beragam algoritma machine learning untuk klasifikasi sentimen.
- Mengevaluasi performa setiap model untuk menemukan yang paling akurat.
- Menarik wawasan dari hasil model terbaik sebagai solusi pemecahan masalah.

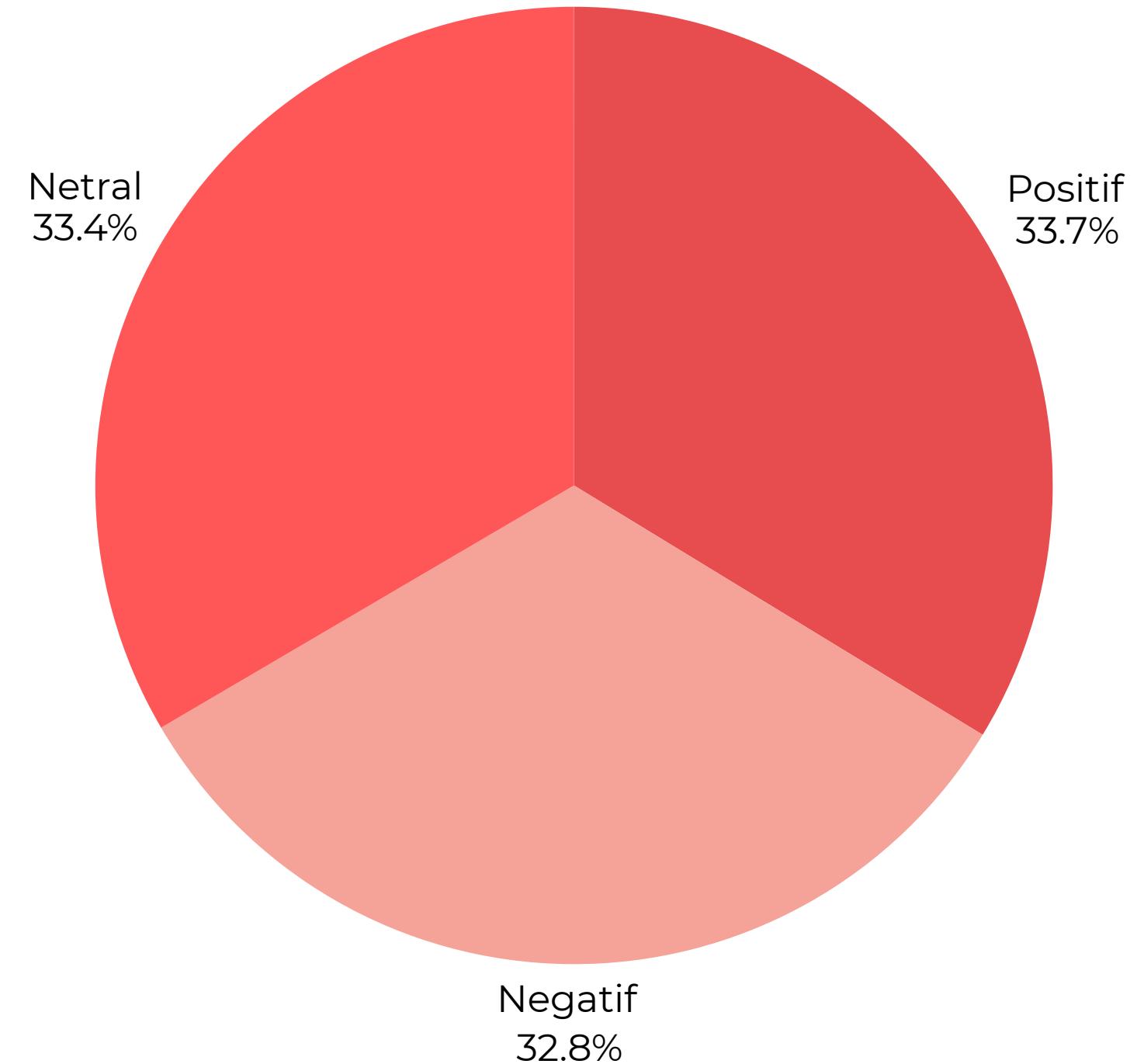
Cakupan:

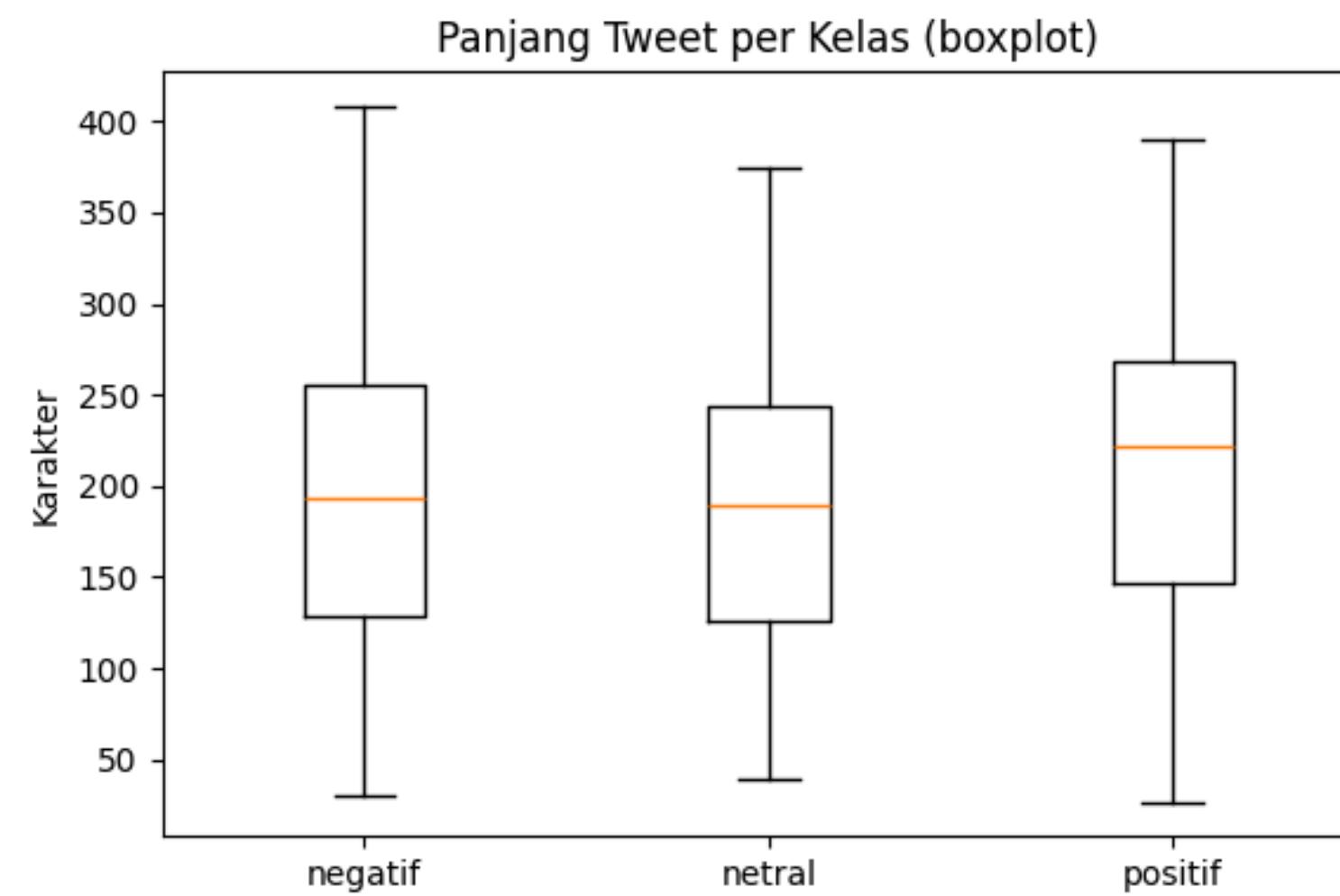
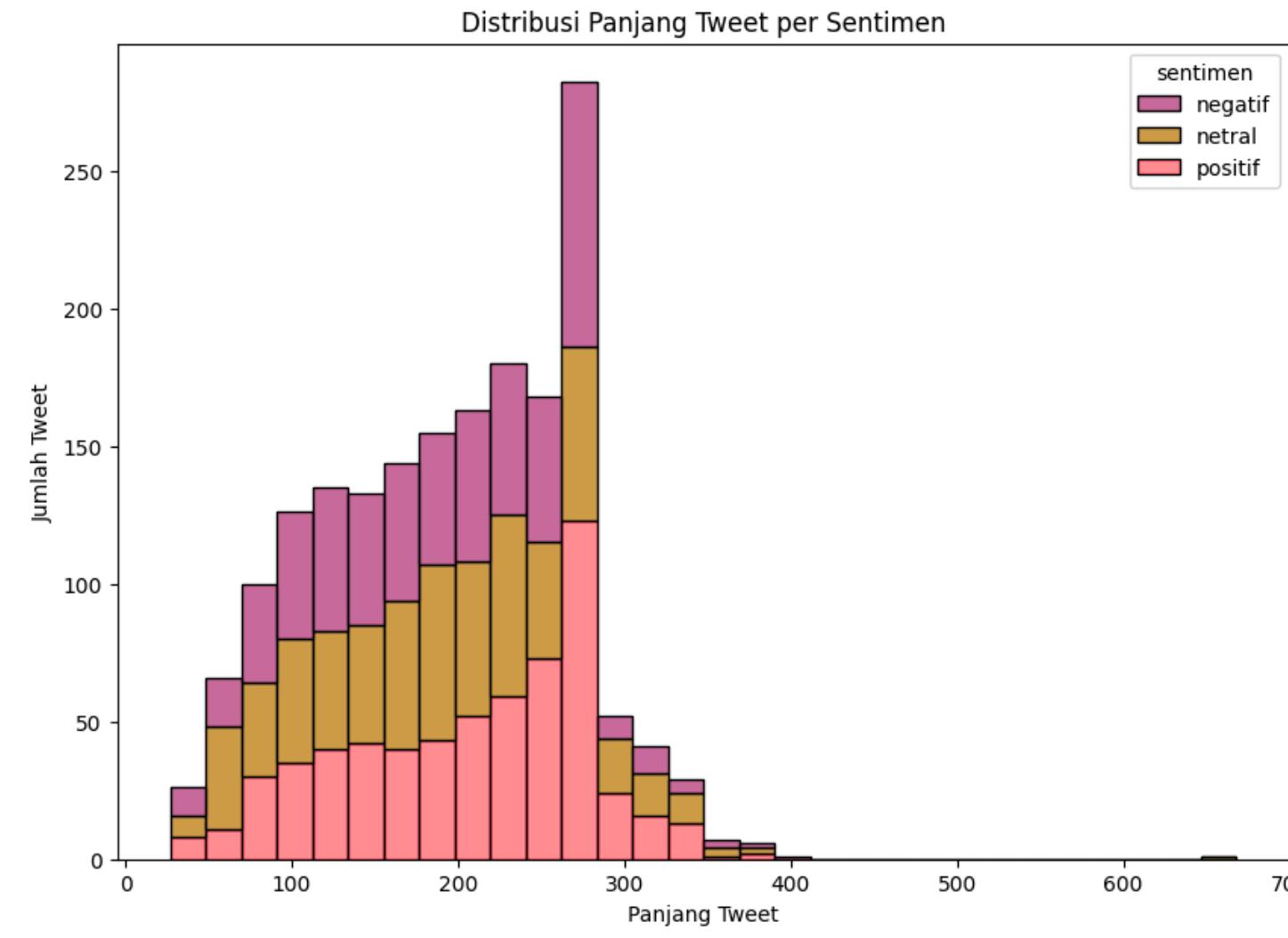
- Fokus proyek adalah mengklasifikasikan tweet ke dalam 3 kategori: Positif, Negatif, dan Netral, menggunakan dataset yang telah disediakan.
- Proyek ini memiliki batasan, yaitu tidak dirancang untuk mendeteksi sarkasme dan tidak menganalisis metadata tweet seperti jumlah likes atau retweet.

Eksplorasi Analisis Data

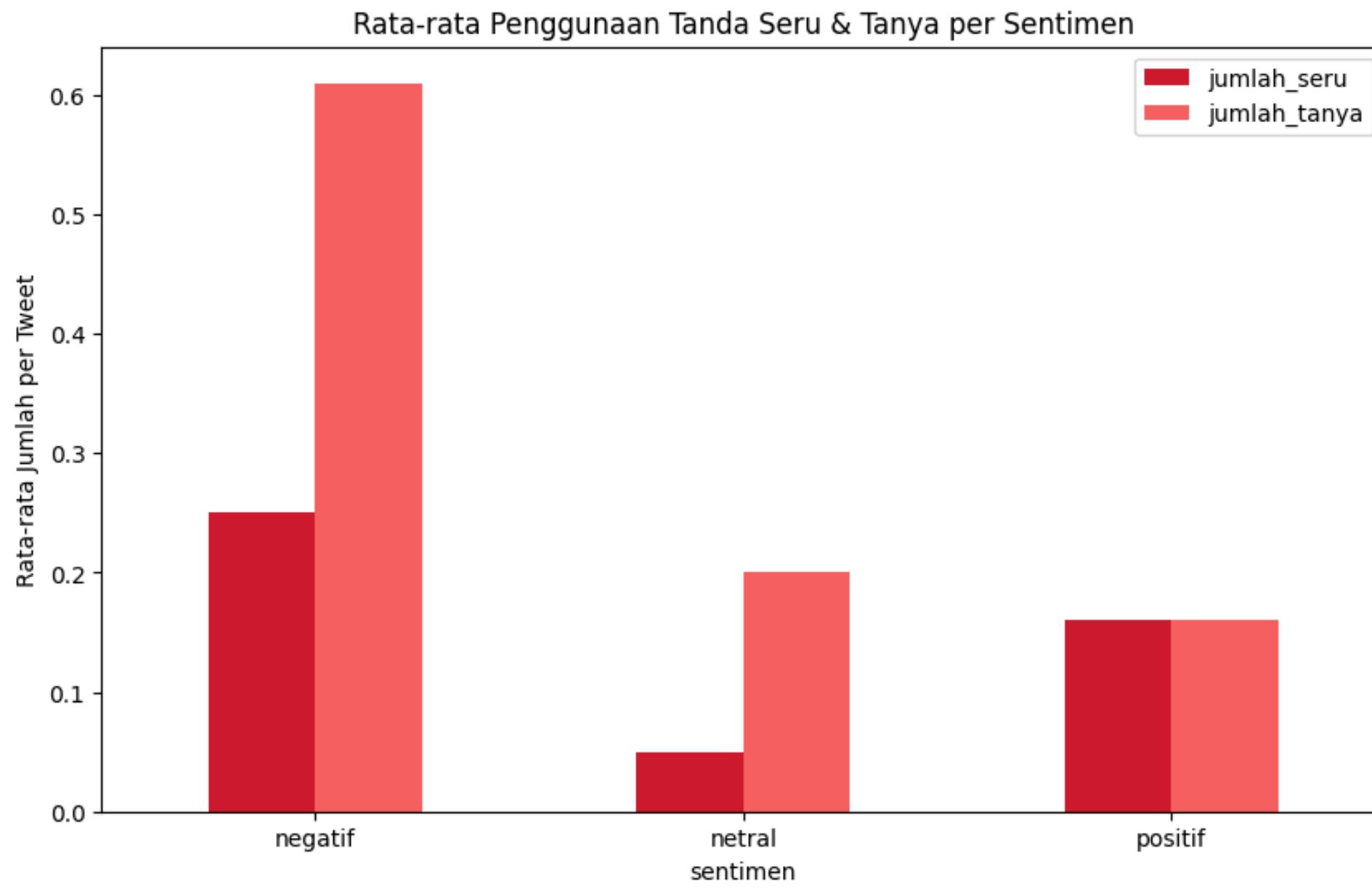
- Cuitan di twitter tentang debat Pilpres 2019.
- Jumlah keseluruhan cuitan: 1.815 cuitan
- Jumlah Tweet per Sentimen: positif (612 cuitan), netral (607 cuitan), dan negatif (596 cuitan).
- Datanya sentimennya cukup berimbang.

Contoh cuitan	
negatif	02 kelar hidupmu.. bisanya cuma harga mahal turunkan solusi 0% kubu Prabowo bahlull
netral	Terlalu berat nge bahas ekonomi dgn angka2.. Cukup bahas mobil lejen aj,
positif	Bagi Sumatera Selatan, Asian Games berdampak pd ekonomi langsung diprediksi mencapai 18,5 triliun. Indonesia maju, Jokowi hebat!

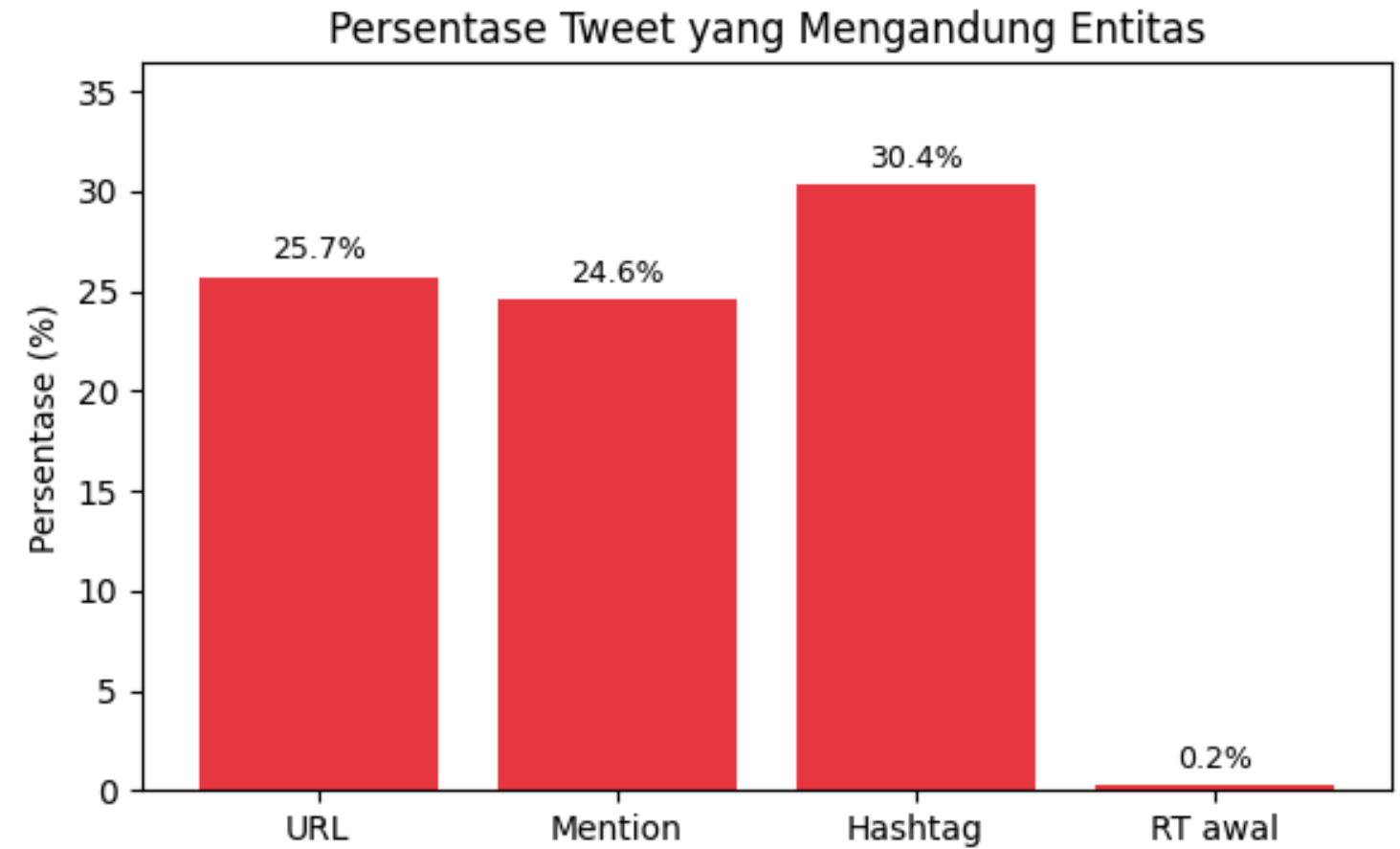




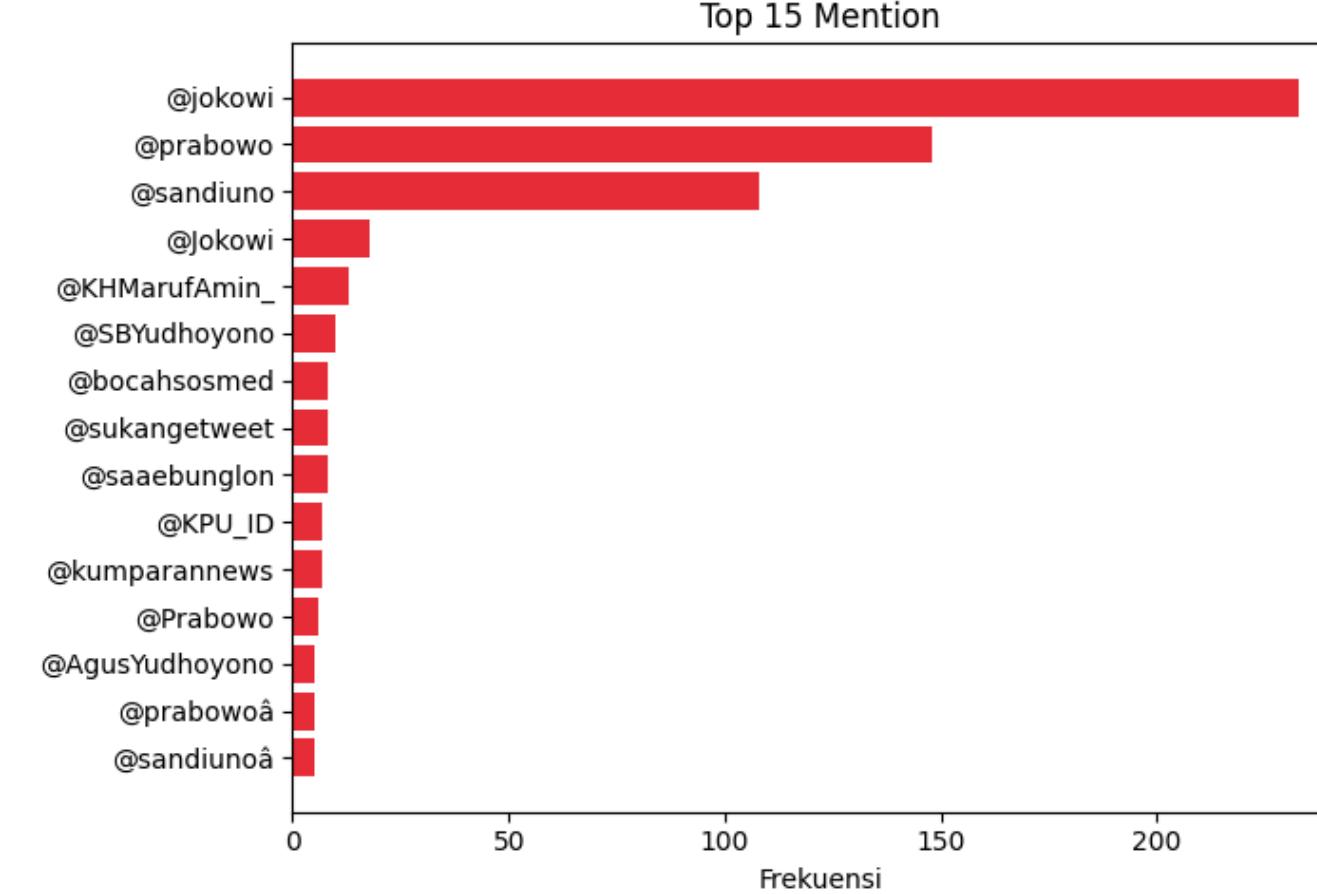
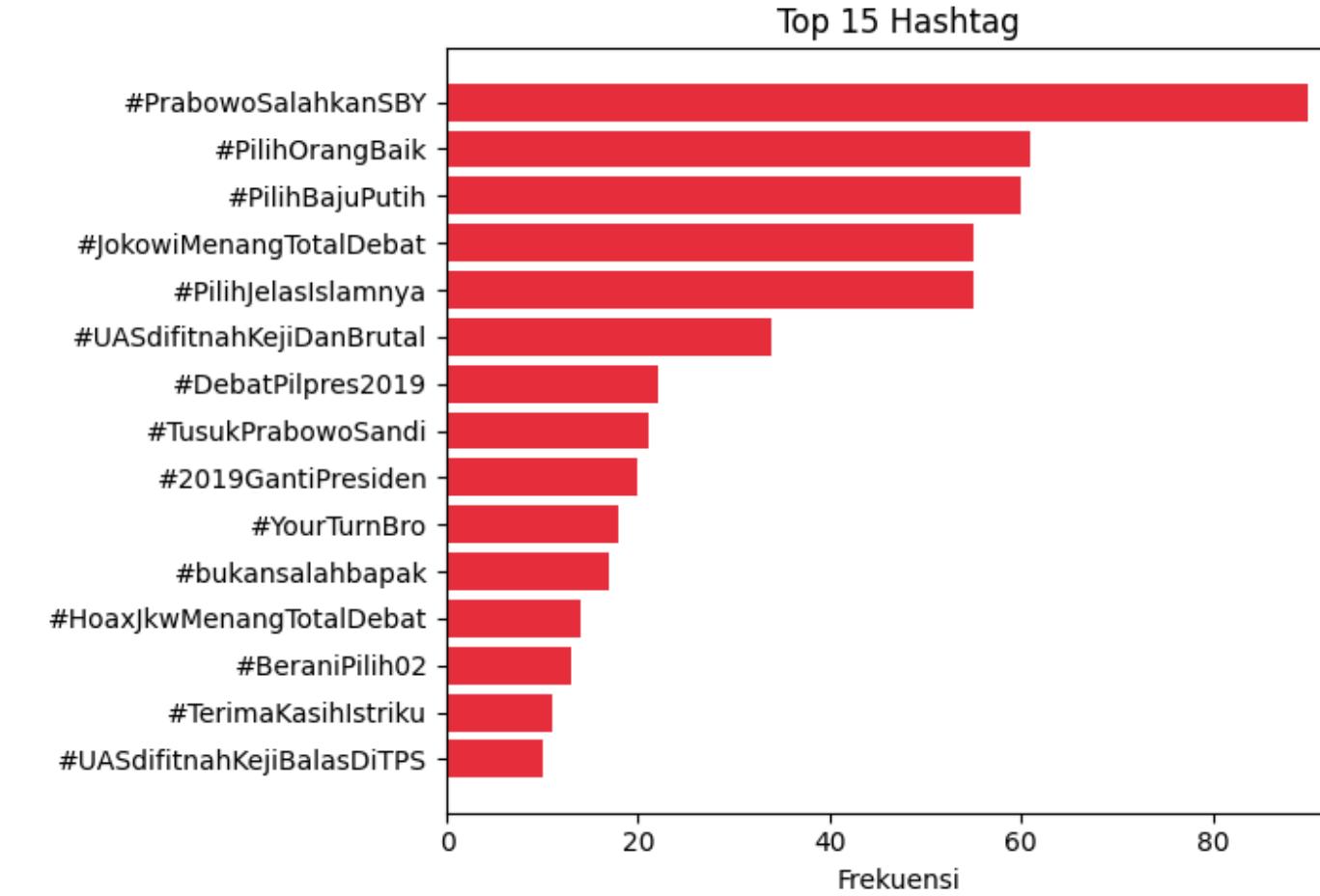
- Grafik di atas menunjukkan sebaran panjang tweet (dalam jumlah karakter) untuk tiga kelas sentimen: positif, netral, dan negatif.
- Sebagian besar tweet berada di rentang 100–300 karakter, dengan puncak sekitar 250 karakter.
- Distribusi ketiga kelas sentimen tampak relatif mirip, tetapi tweet positif cenderung sedikit lebih panjang dibandingkan tweet negatif.
- Hanya sedikit tweet yang memiliki panjang lebih dari 400 karakter, menunjukkan bahwa pengguna jarang memanfaatkan batas maksimal karakter di Twitter.
- **Penelitian Brady (2021)** menunjukkan bahwa Twitter memang didesain untuk “marah-marah” dan salah satu faktornya karena keterbatasan karakter.

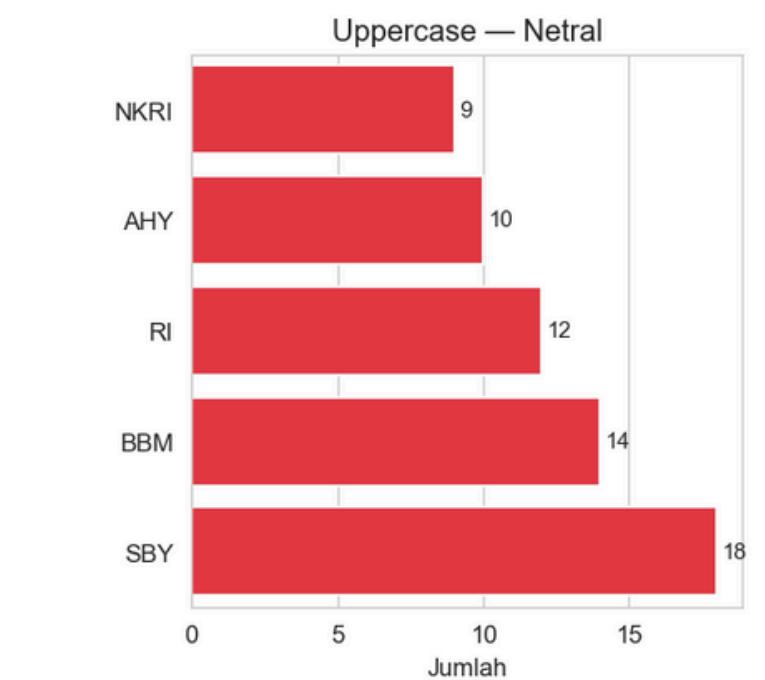
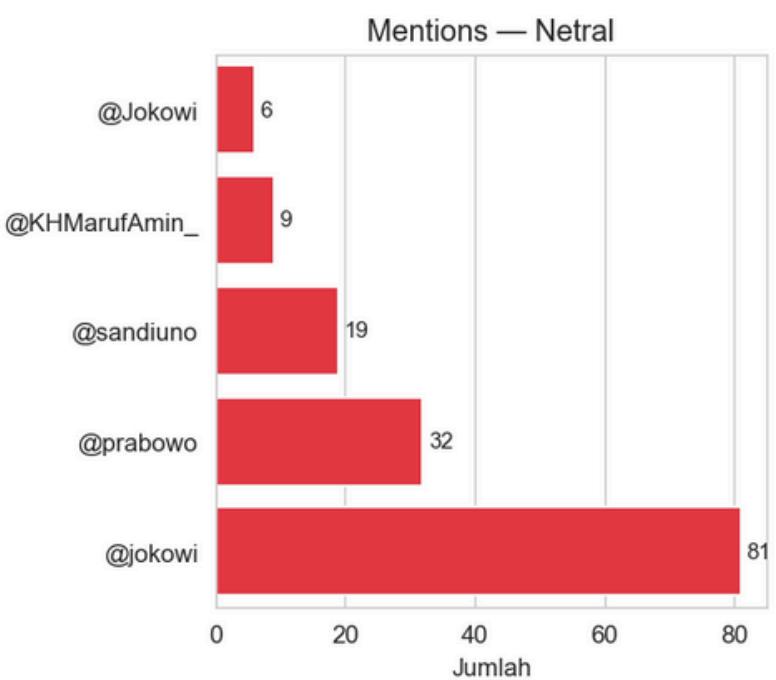
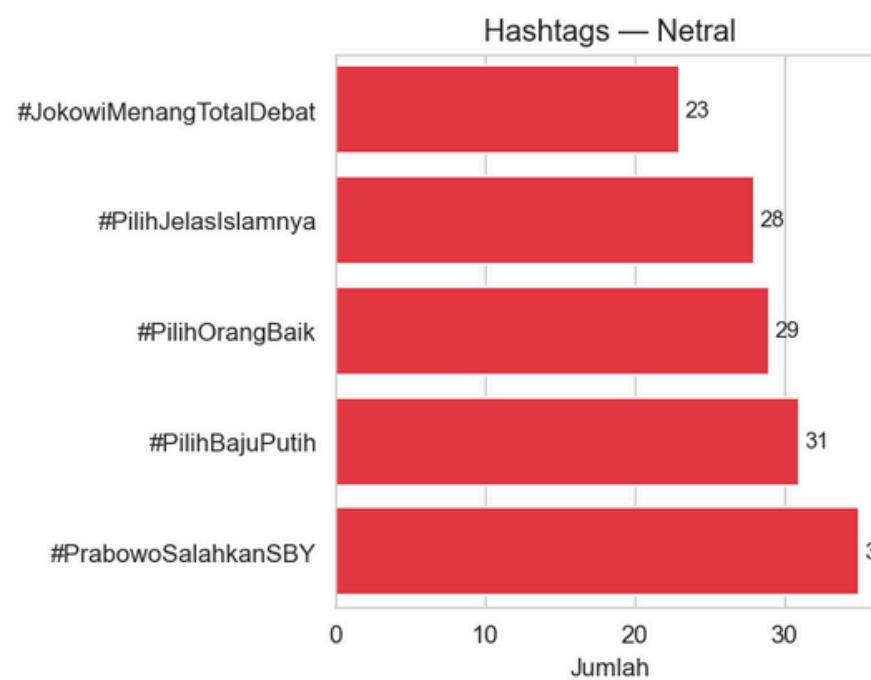
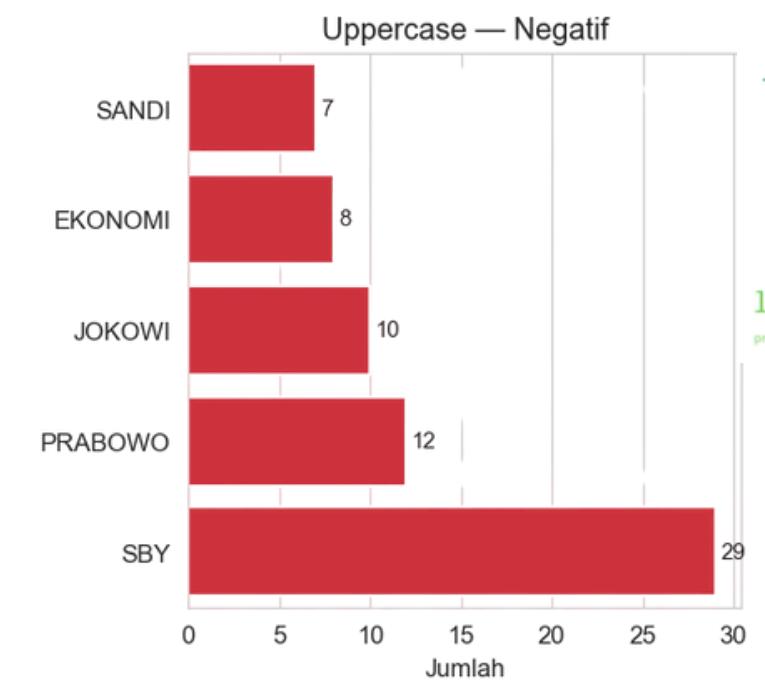
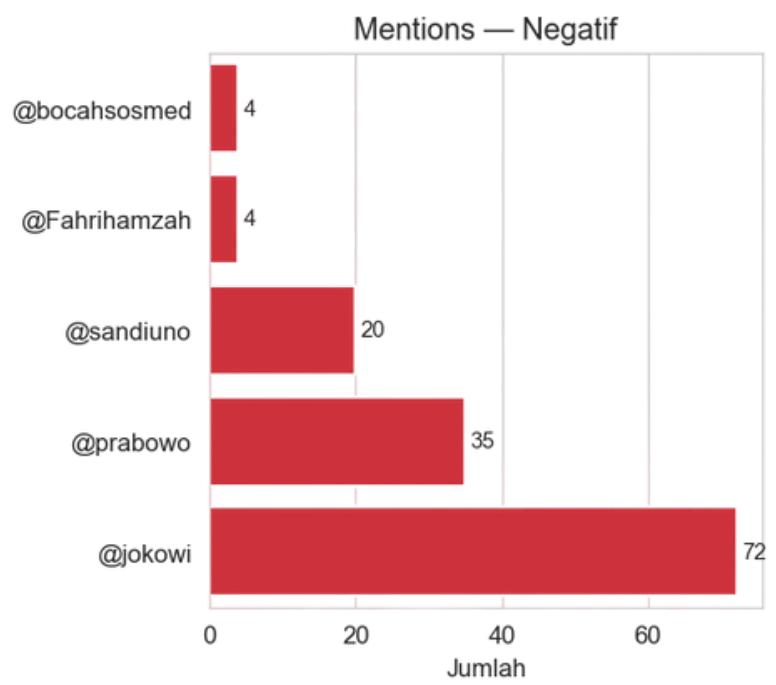
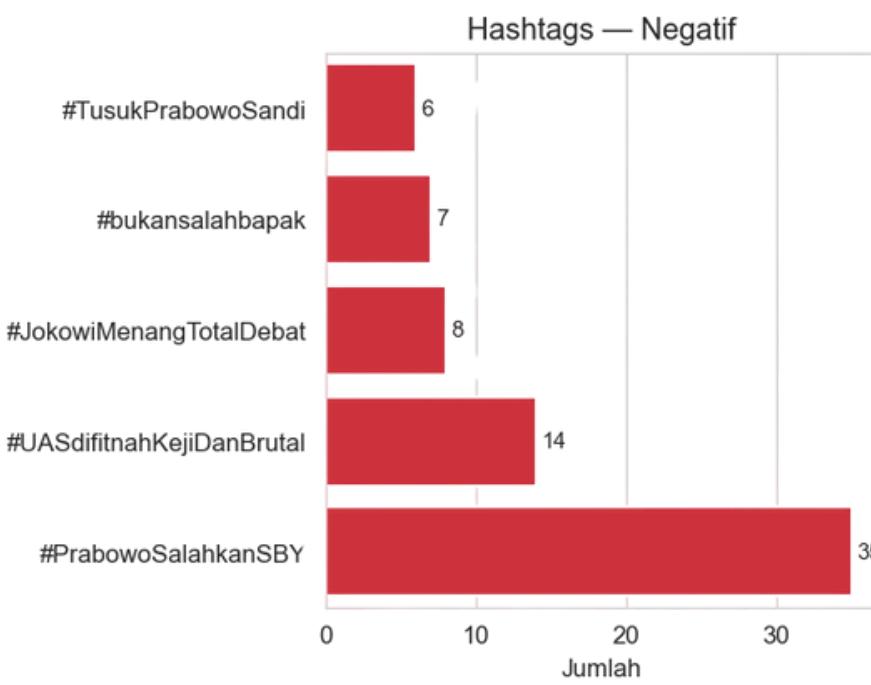
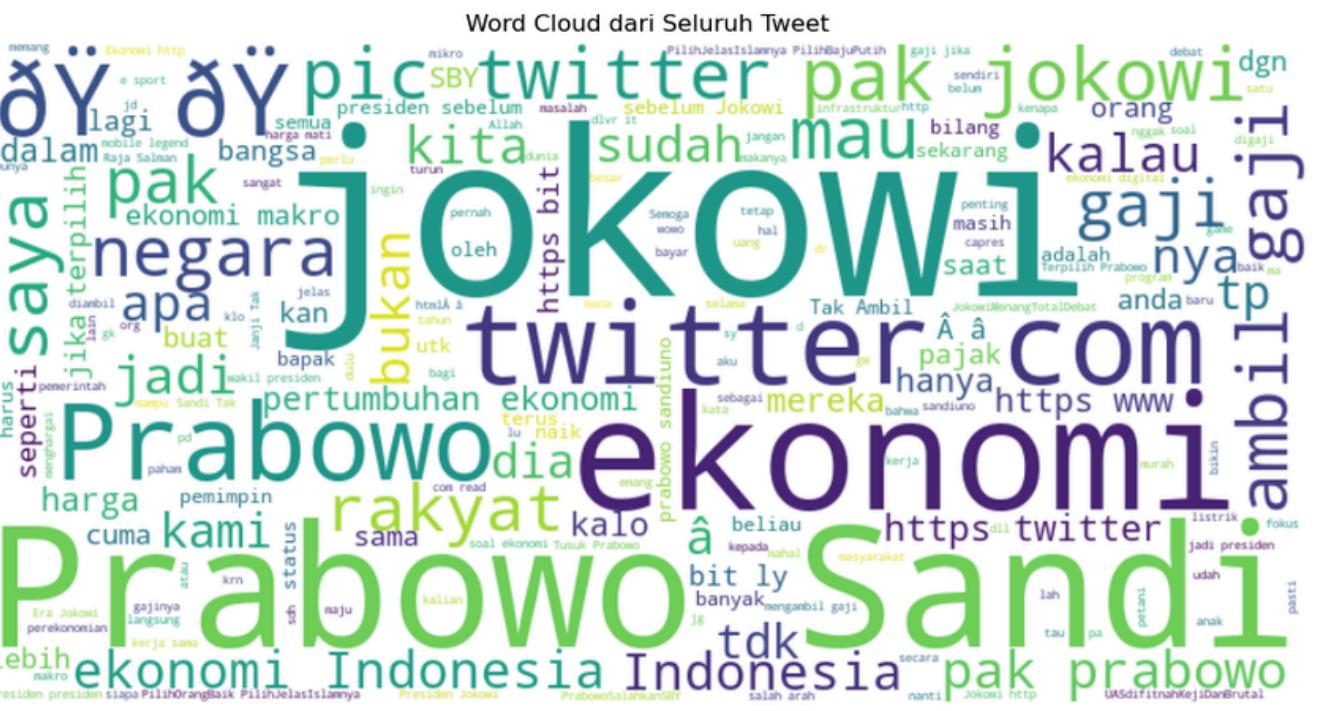
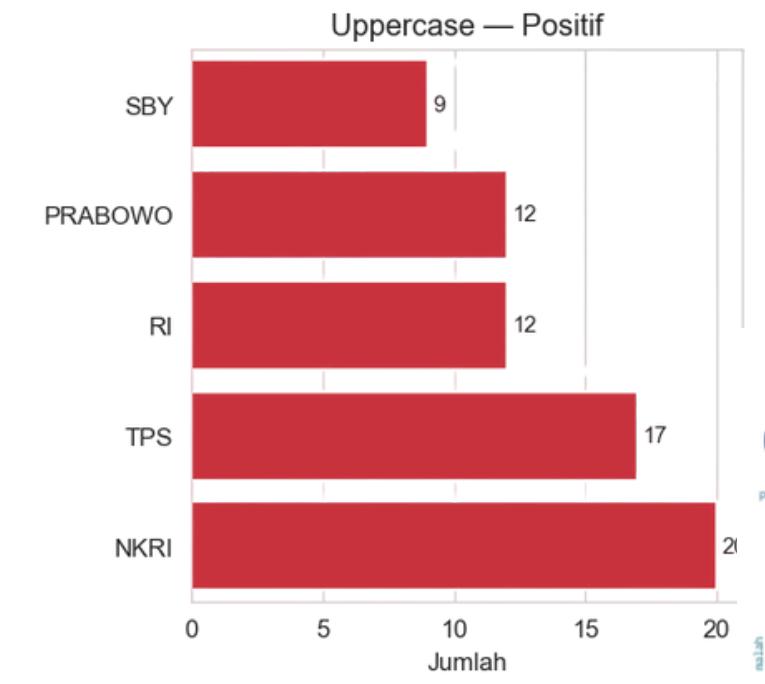
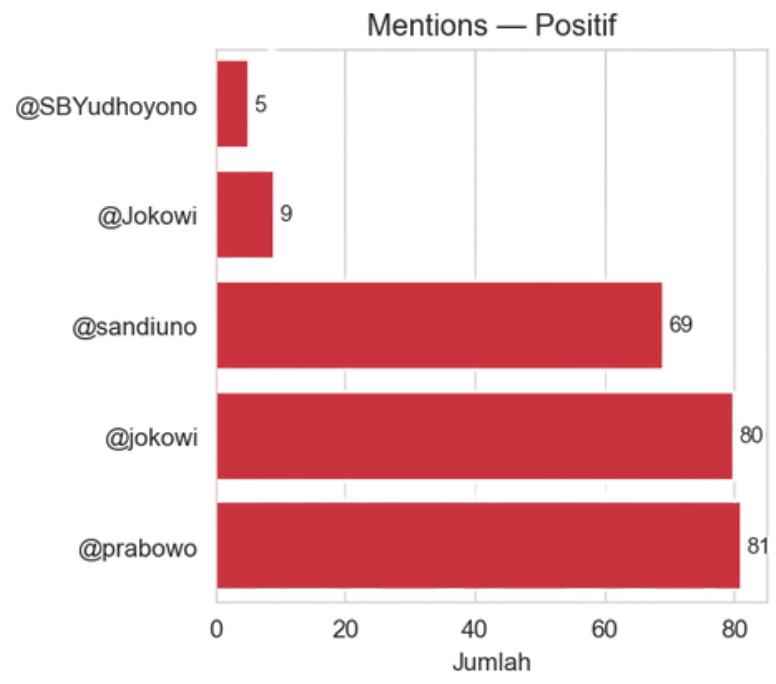
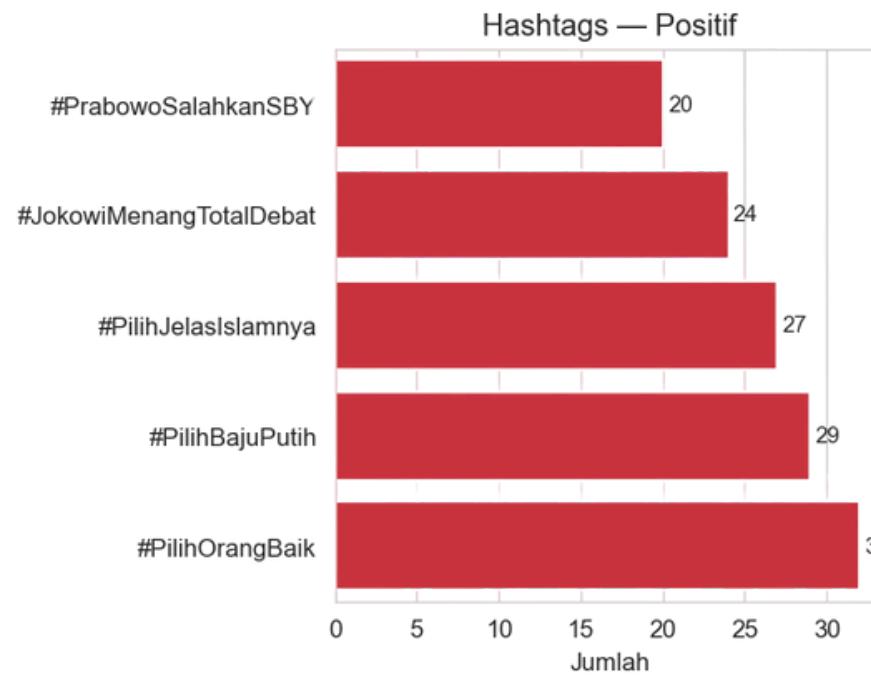


- Tweet ber-sentimen negatif paling banyak mengandung tanda tanya dan tanda seru, dengan rata-rata tanda tanya mencapai sekitar 0,6 per tweet.
- Tweet netral relatif jarang menggunakan tanda baca ekspresif.
- Tweet positif menggunakan tanda seru dan tanda tanya dalam jumlah yang hampir sama, tetapi lebih sedikit dibanding tweet negatif.
- Pola ini menunjukkan bahwa tweet negatif cenderung lebih ekspresif atau emosional, seringkali menggunakan tanda baca untuk menekankan emosi seperti marah, heran, atau frustrasi.
- Salah satu SLR paper berjudul **“Sarcasm in Sentiment Analysis of Indonesian Text: A Literature Review”** (**Christina, 2019**) menyatakan bahwa tanda seru dan tanda tanya dapat mengindikasikan sarkasme dalam bahasa Indonesia.



- Sekitar 30,4% tweet mengandung hashtag, menunjukkan bahwa banyak pengguna memanfaatkan tagar untuk memperluas jangkauan atau mengaitkan topik tertentu.
- URL (25,7%) dan mention (24,6%) juga sering muncul, menandakan adanya interaksi dan berbagi tautan dalam percakapan.
- Retweet di awal (RT) hampir tidak muncul (0,2%), artinya sebagian besar tweet merupakan tweet orisinal, bukan hasil retweet.





Data Preprocessing

Data Preprocessing:

- Menghapus *stopwords* yang tidak diperlukan. Hal itu dilakukan karena dikhawatirkan dapat mengubah sentimen.
- Menginstal module `indonesia.normalizer`
- Menormalisasi kata slang, singkatan, menjadi bentuk baku (sdg - sedang)
- Mengatasi Emoji dan Mojibake (Menghapusnya karena tidak relevan dan tidak berkorelasi)

Pembagian Dataset Training, Validation dan Testing:

- Menggunakan 2 opsi proporsi split data.
- Opsi 1: 80:20
- Opsi 2: 85:15
- Deep Learning dan Transformer based model final menggunakan KFold-Cross Validation dengan ratio train:validation 70:15.

Metode Vektorisasi

- Baseline model menggunakan TF-IDF Vectorizer (ambil kata dan bobotnya).
- Deep Learning dan Transformer based model menggunakan FastText Embedding.

Model Development

Models Used:

Baseline:

- Logistic Regression
- Random Forest
- Naive Bayes
- SVM

Deep Learning:

- LSTM
- Hybrid CNN-LSTM

Transformer Model:

- IndoBERT

Model Evaluation Metrics:

- Precision
- Recall
- F1-Score
- Accuracy

Hasil Baseline Modelling

Logistic Regression

	Precision	Recall	F1-score	Accuracy
negatif	0.64	0.70	0.67	0.67
netral	0.67	0.68	0.67	
positif	0.71	0.63	0.67	

Random Forest

	Precision	Recall	F1-score	Accuracy
negatif	0.63	0.62	0.63	0.62
netral	0.60	0.70	0.65	
positif	0.62	0.52	0.57	

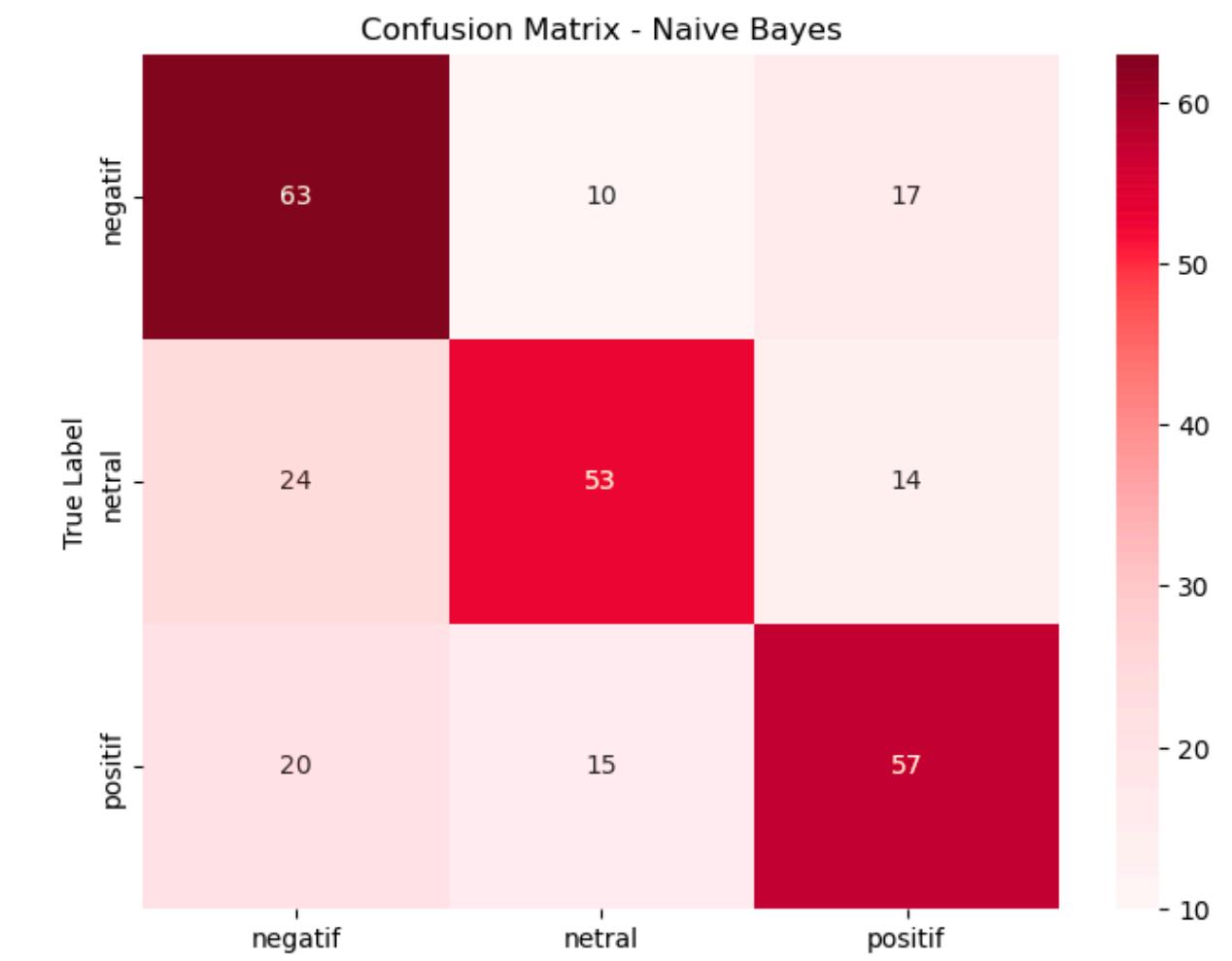
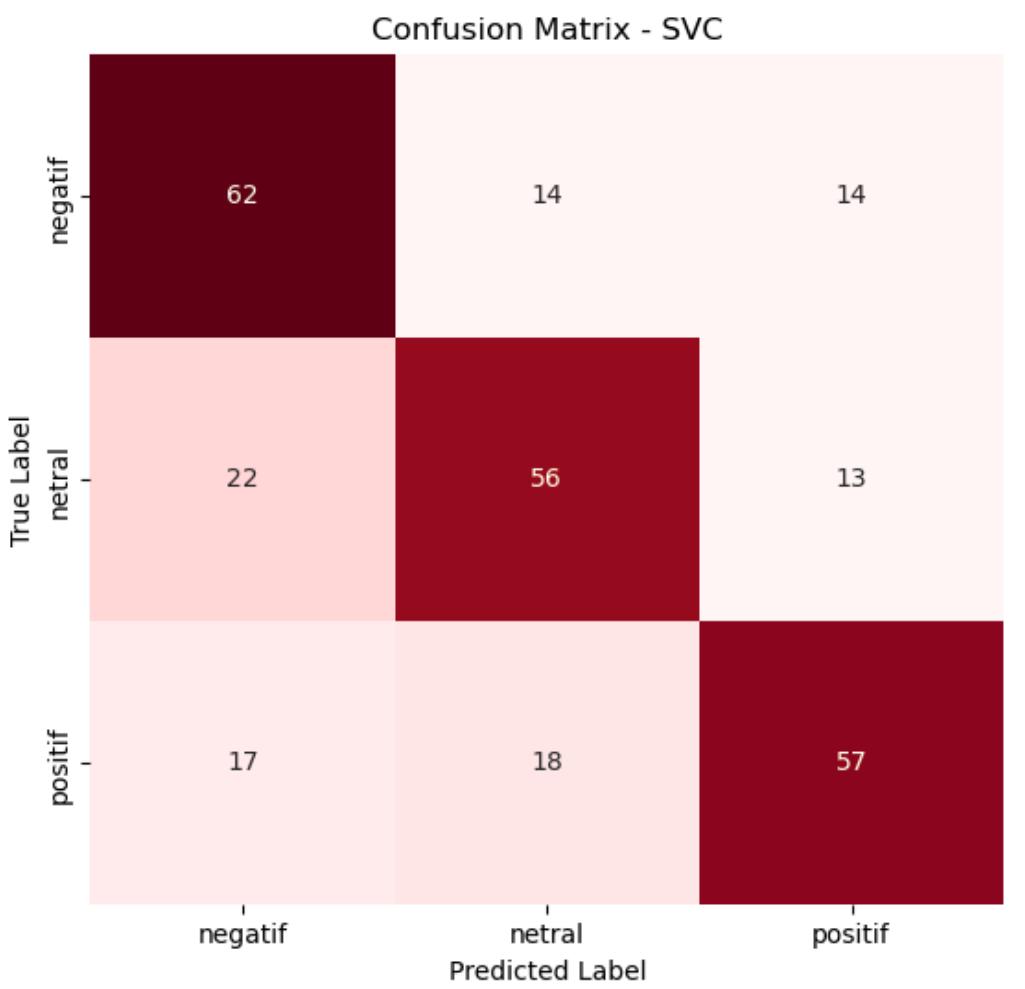
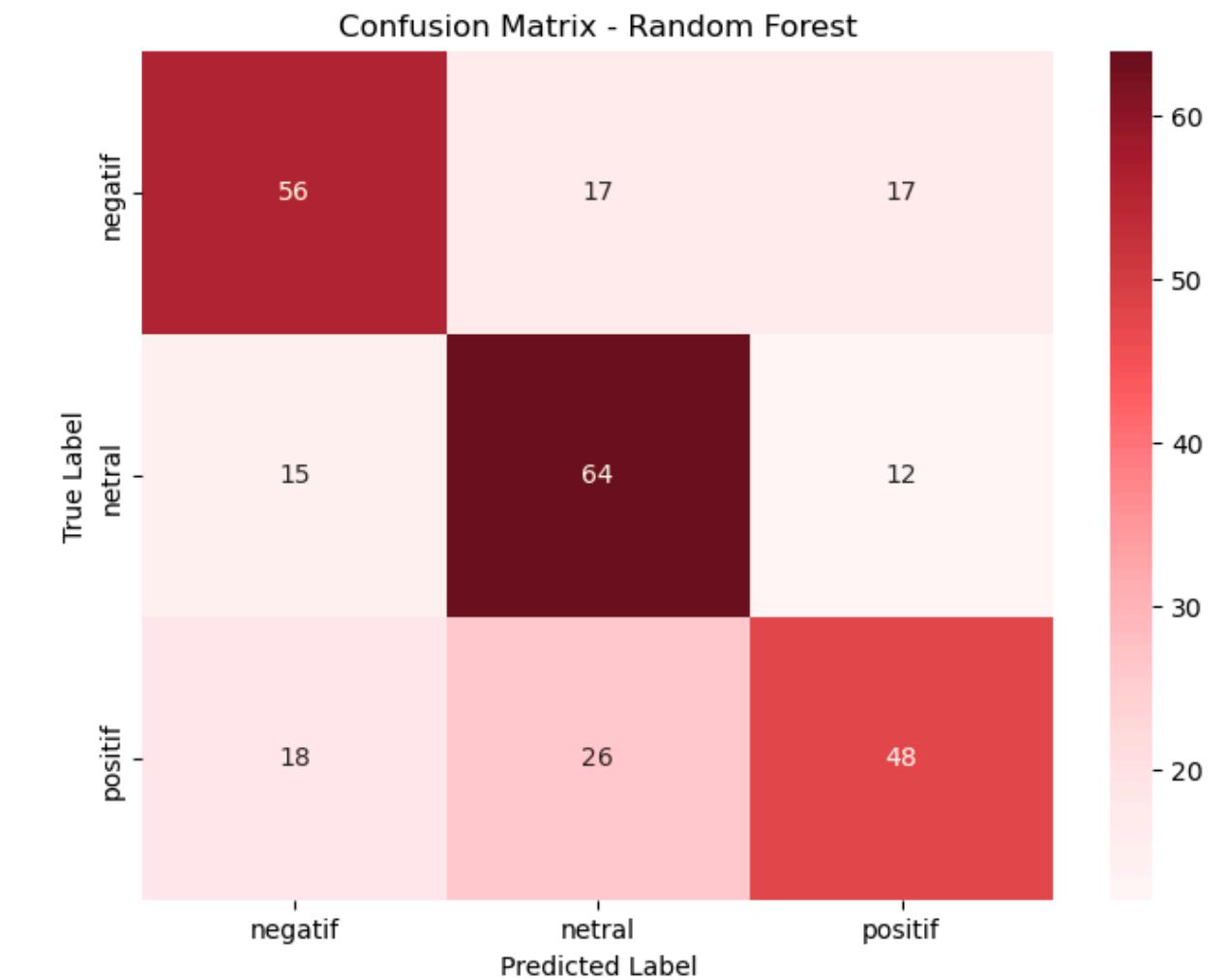
Naive Bayes

	Precision	Recall	F1-score	Accuracy
negatif	0.59	0.70	0.64	0.63
netral	0.68	0.58	0.63	
positif	0.65	0.62	0.63	

SVM

	Precision	Recall	F1-score	Accuracy
negatif	0.61	0.69	0.65	0.64
netral	0.64	0.62	0.63	
positif	0.68	0.62	0.65	

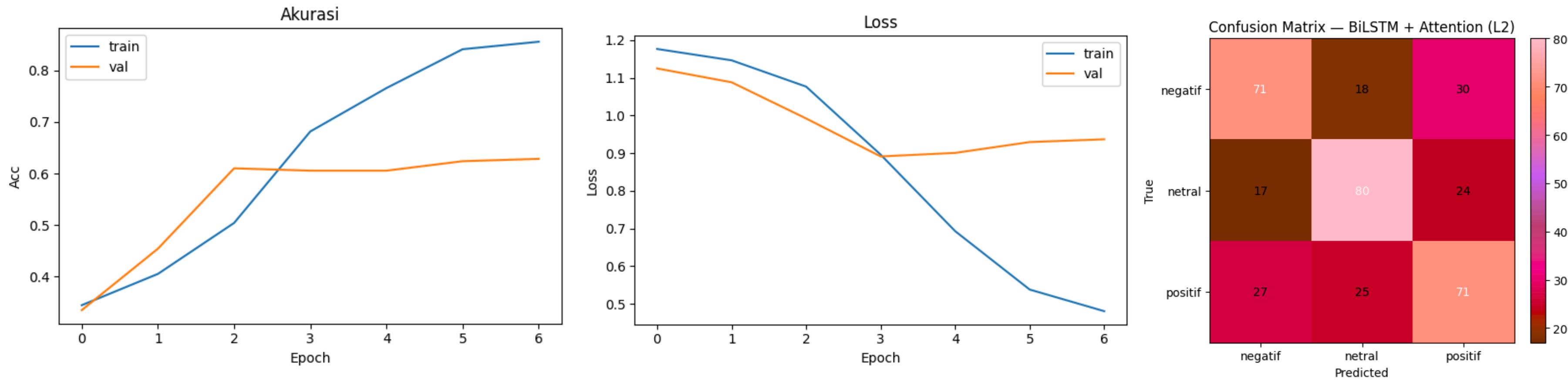
Matrik Evaluasi



Model	Akurasi Latih	Akurasi Uji	Selisih (Overfitting)
Logistic Regression	79,31%	67,03%	12,28%
SVM	80,29%	64,10%	16,19%
Naive Bayes	74,32%	63,37%	10,95%
Random Forest	96,95%	61,54%	35,41%

- Random Forest belajar terlalu banyak dari data latih (*overfitting*) sehingga performa pada data baru menurun drastis.
- Logistic Regression dan Naive Bayes paling stabil (selisih kecil antara train dan test).
- SVM sedikit lebih tinggi selisihnya, tapi masih dalam batas wajar.

Deep Learning LSTM



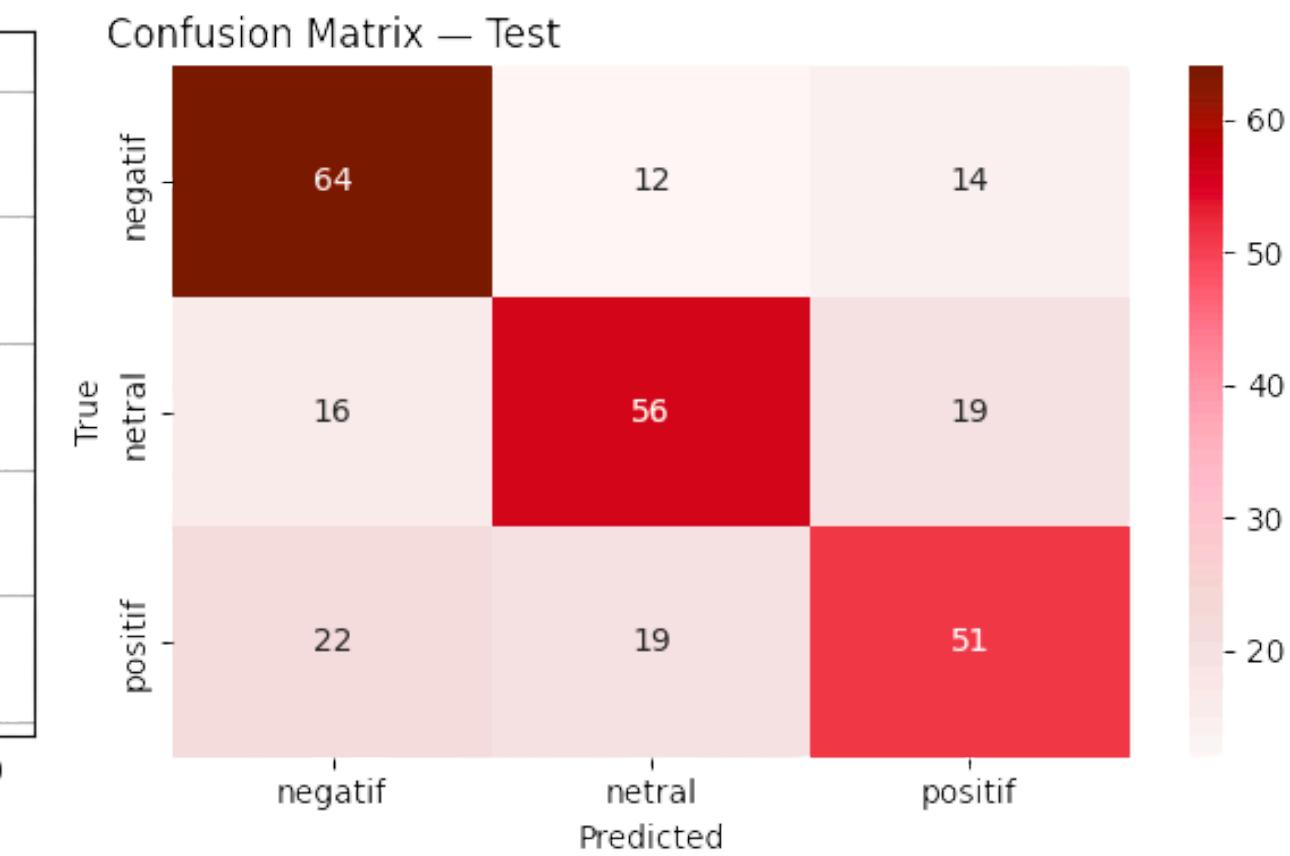
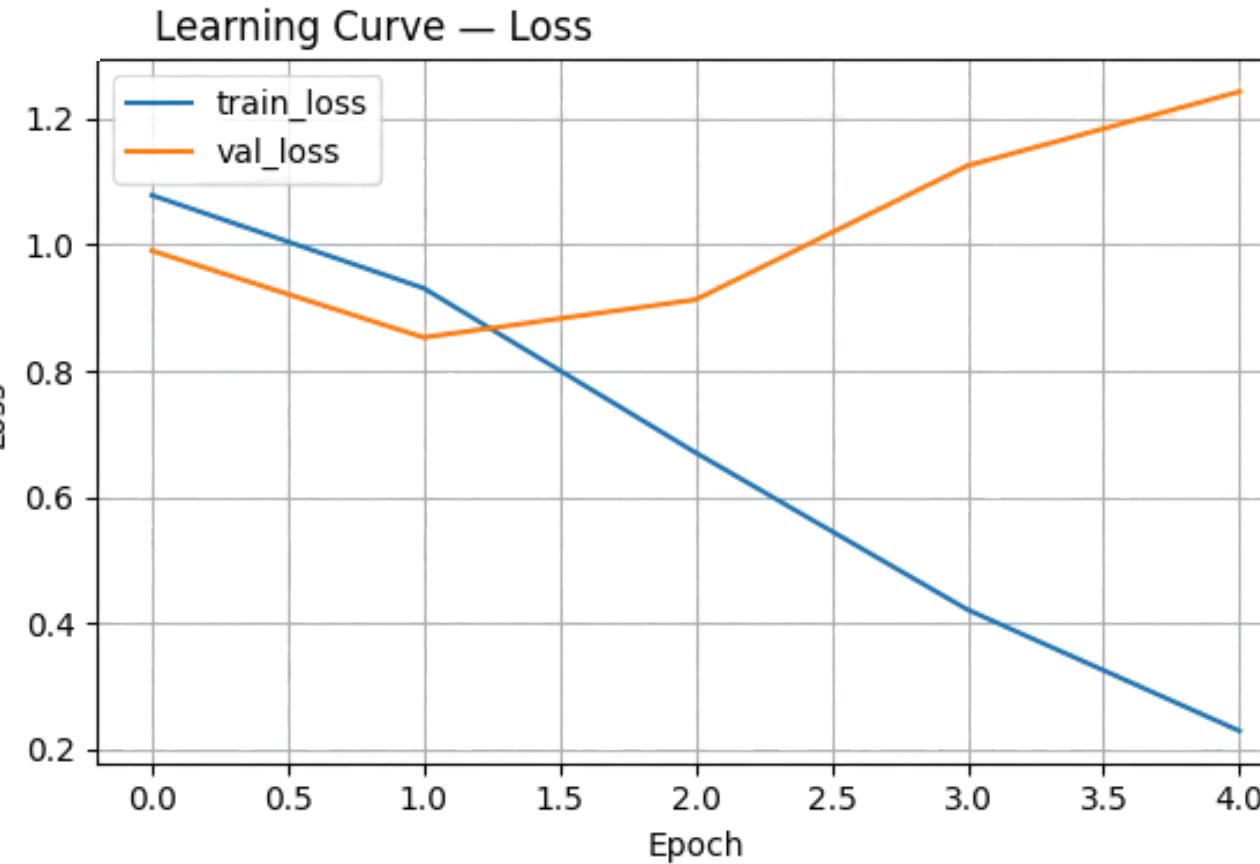
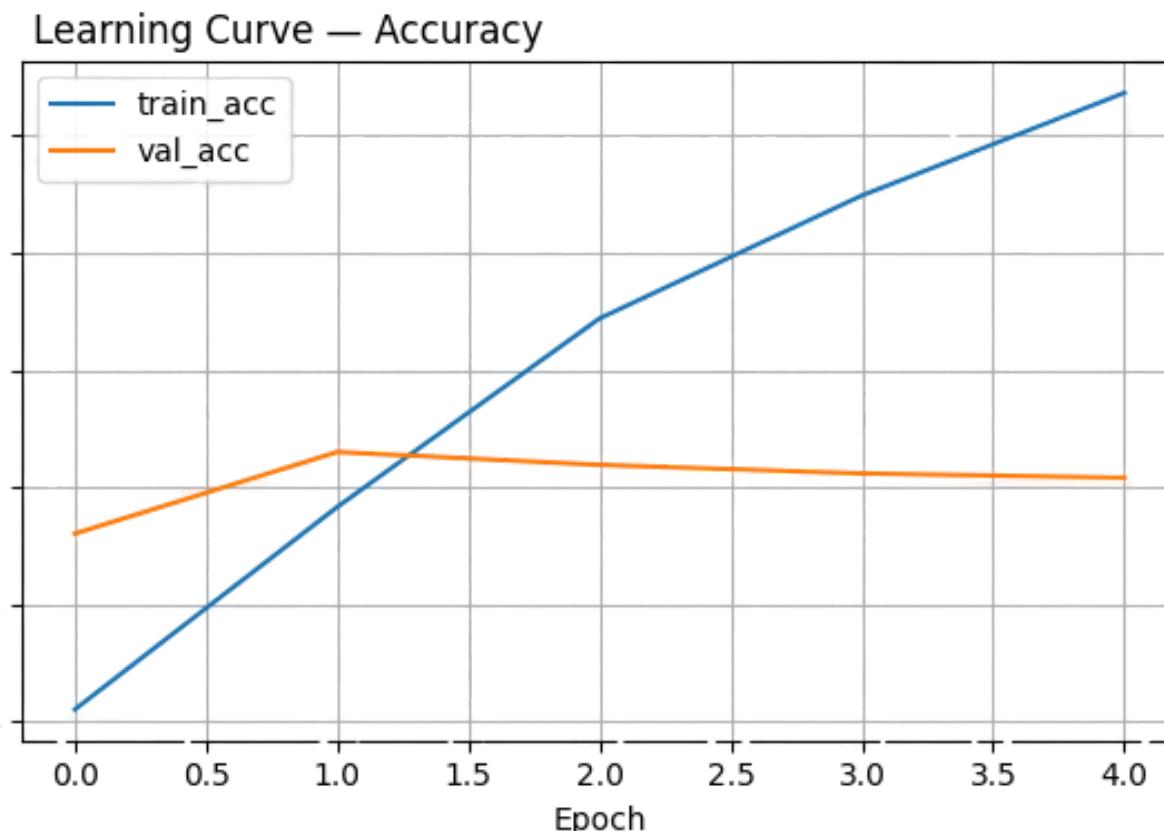
Classification Report (test):

	precision	recall	f1-score	support
negatif	0.6174	0.5966	0.6068	119
netral	0.6504	0.6612	0.6557	121
positif	0.5680	0.5772	0.5726	123
accuracy			0.6116	363
macro avg	0.6119	0.6117	0.6117	363
weighted avg	0.6117	0.6116	0.6115	363

Analisis Hasil:

- Akurasi train terus naik sampai 0.85, sedangkan akurasi val berhenti di 0.62–0.63.
- Loss train turun stabil dari 1.18 ke 0.50, loss val turun sampai epoch 3 lalu cenderung naik lagi.
- Interpretasi: model mulai overfitting setelah sekitar epoch 2–3. Minimum loss val tampak di sekitar epoch 2.
- Negatif → Positif dan Positif → Negatif masih cukup sering salah. Ini mengindikasikan batas keputusan yang belum tajam untuk polaritas kuat.
- Model masih kekurangan fitur yang menangkap intensitas sentimen dan konteks negasi/emotikon/ironi, sehingga polaritas berlawanan masih tertukar.

Deep Learning Hybrid CNN+LSTM



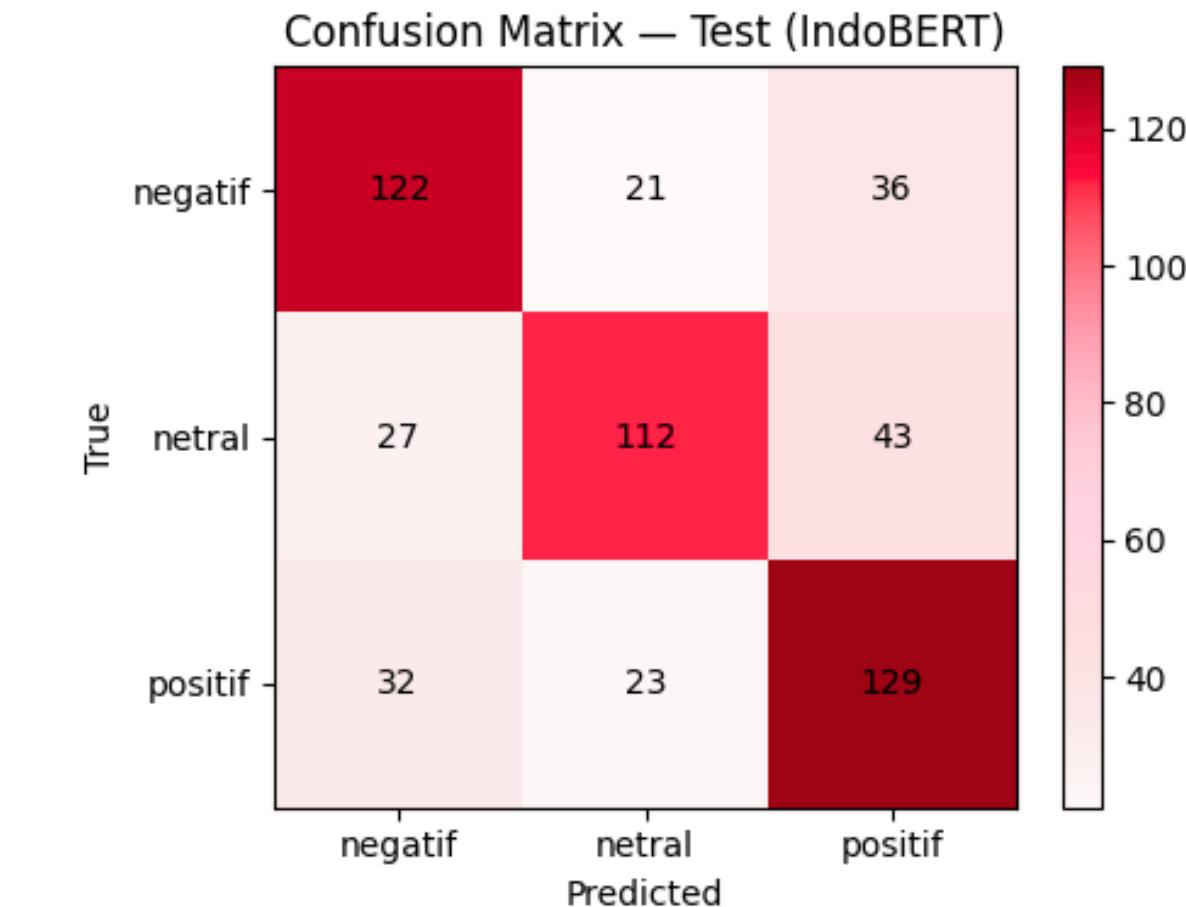
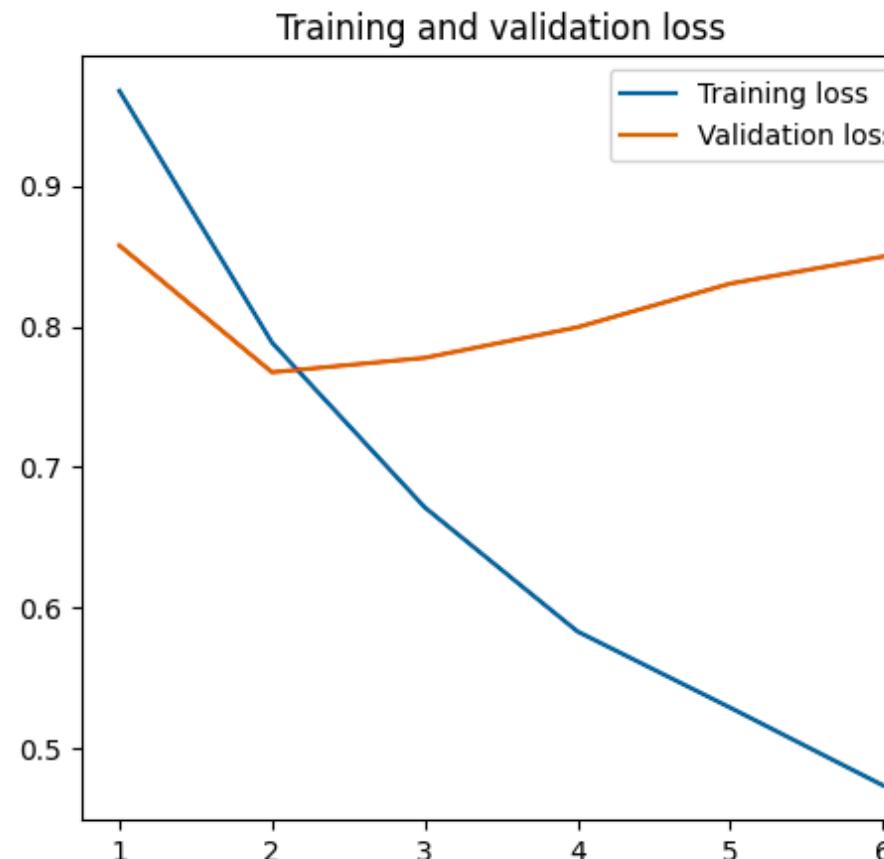
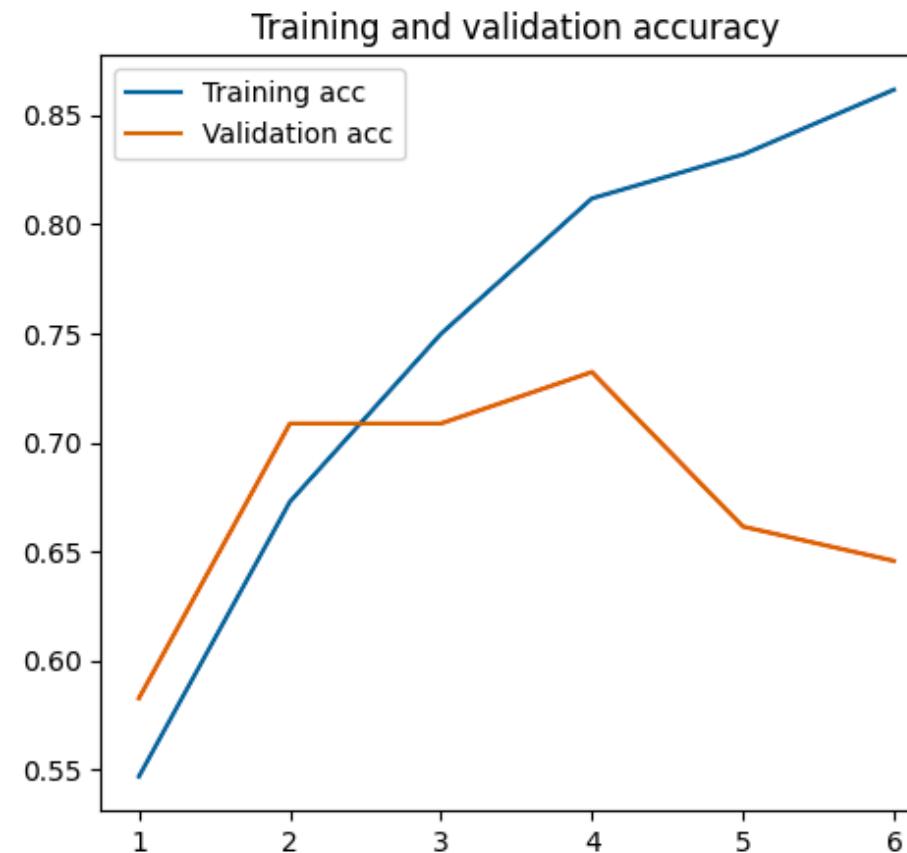
[Hybrid] Classification Report:

	precision	recall	f1-score	support
negatif	0.63	0.71	0.67	90
netral	0.64	0.62	0.63	91
positif	0.61	0.55	0.58	92
accuracy			0.63	273
macro avg	0.63	0.63	0.63	273
weighted avg	0.63	0.63	0.62	273

Analisis Hasil:

- Per kelas: negatif F1=0.67 (terbaik), netral F1=0.63, positif F1=0.58 (terlemah).
- Interpretasi: model relatif andal untuk “negatif”, moderat untuk “netral”, dan masih lemah membedakan “positif”.
- Overfitting kuat mulai sekitar epoch 1–2.
- Masih terdapat pola kesalahan Negatif → Positif dan Positif → Negatif.
- Terdapat pola drift pada kelas netral. Ketika sinyal polaritas kurang kuat atau bercampur dengan tagar/topik dukungan, prediksi bergeser ke netral atau positif.

Transformer Model IndoBERT



```
-- Test Report (IndoBERT) --
```

	precision	recall	f1-score	support
negatif	0.67	0.68	0.68	179
netral	0.72	0.62	0.66	182
positif	0.62	0.70	0.66	184
accuracy			0.67	545
macro avg	0.67	0.67	0.67	545
weighted avg	0.67	0.67	0.67	545

Analisis Hasil:

- Train acc naik $\approx 0.55 \rightarrow 0.86$; val acc naik sampai sekitar epoch 4 (~ 0.73) lalu turun ke ~ 0.65 .
- Train loss turun konsisten; val loss berbentuk U dan naik kembali setelah epoch 2–3.
- Interpretasi: overfitting mulai muncul paska epoch 3. Titik model terbaik kemungkinan sekitar epoch 2–4 (minimum val loss).
- Terjadi kebingungan polaritas kuat (negatif↔positif) serta bias ke “positif” untuk contoh netral. Model cukup baik membedakan positif (recall 0.70) namun dengan presisi yang lebih rendah pada kelas tersebut.

Simpulan

- Dataset memiliki distribusi sentimen yang berimbang (sekitar 33% untuk Positif, Netral, dan Negatif).
- Tweet negatif adalah yang paling ekspresif/emosional, terlihat dari rata-rata penggunaan tanda tanya (≈ 0.6) dan tanda seru yang paling tinggi.
- Tweet positif dominan dengan hashtag dukungan (#PilihOrangBaik, #PilihBajuPutih) , sementara tweet negatif memiliki nada kritik (#TusukPrabowoSandi).
- Tweet positif dan netral cenderung menyoroti isu ekonomi dan nama Jokowi/Prabowo, yang mengindikasikan bahwa sentimen terpusat pada isu politik dan ekonomi di Indonesia.
- Distribusi kelas relatif seimbang sehingga Macro-F1 layak dijadikan metrik utama selain akurasi.
- Baseline klasik menunjukkan pola umum:
 - Logistic Regression dan Naive Bayes paling stabil (gap train-test kecil).
 - Random Forest sangat overfit, akurasi latih tinggi tetapi uji turun tajam.
 - SVM kompetitif namun sensitif terhadap setelan fitur dan regulerisasi.
- Model sekuens (LSTM, Hybrid CNN+LSTM) meningkat pada sebagian pola lokal, tetapi cepat overfit.
- IndoBERT dan Logistic Regression memberi kinerja tertinggi di validasi namun masih menghadapi kebingungan polaritas kuat negatif↔positif dan drift netral→positif.
- Pola error paling konsisten: pertukaran negatif dan positif, sedangkan netral relatif stabil namun rentan terseret ke positif saat ada kata dukungan/topik populer.

