

Analisis Sentimen

Arjuna Rivaldo

AI UNTUK[®]
SEMUA

About me



Arjuna Rivaldo Satria Prayoga

Di project ini saya berperan dalam proses analisis sentimen menggunakan data dari Twitter. Tujuannya untuk memahami kecenderungan opini publik — apakah positif, negatif, atau netral — terhadap topik politik tertentu. Melalui project ini, saya belajar bagaimana NLP bekerja dalam memproses bahasa alami secara komputasional

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.

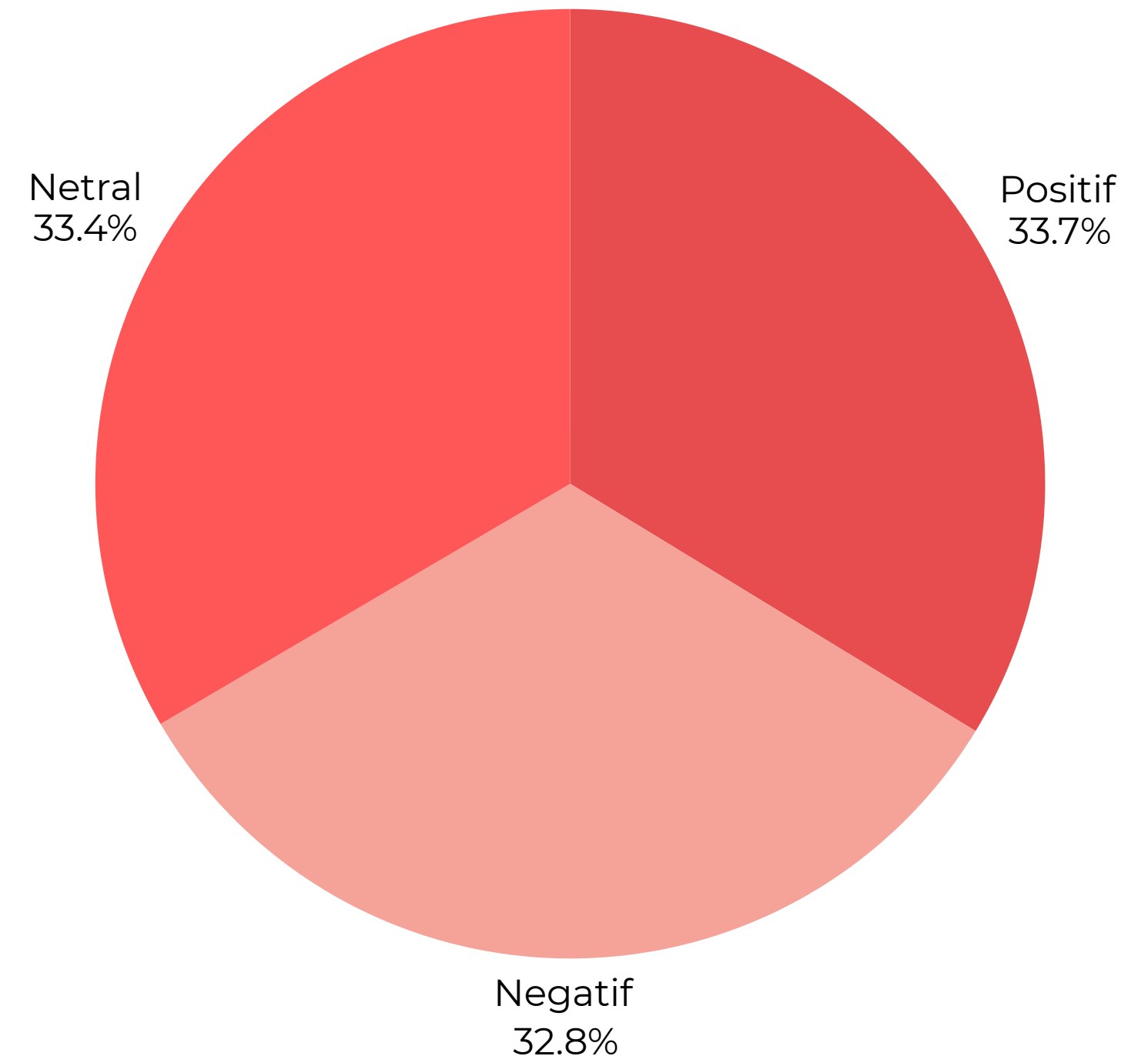
Cakupan:

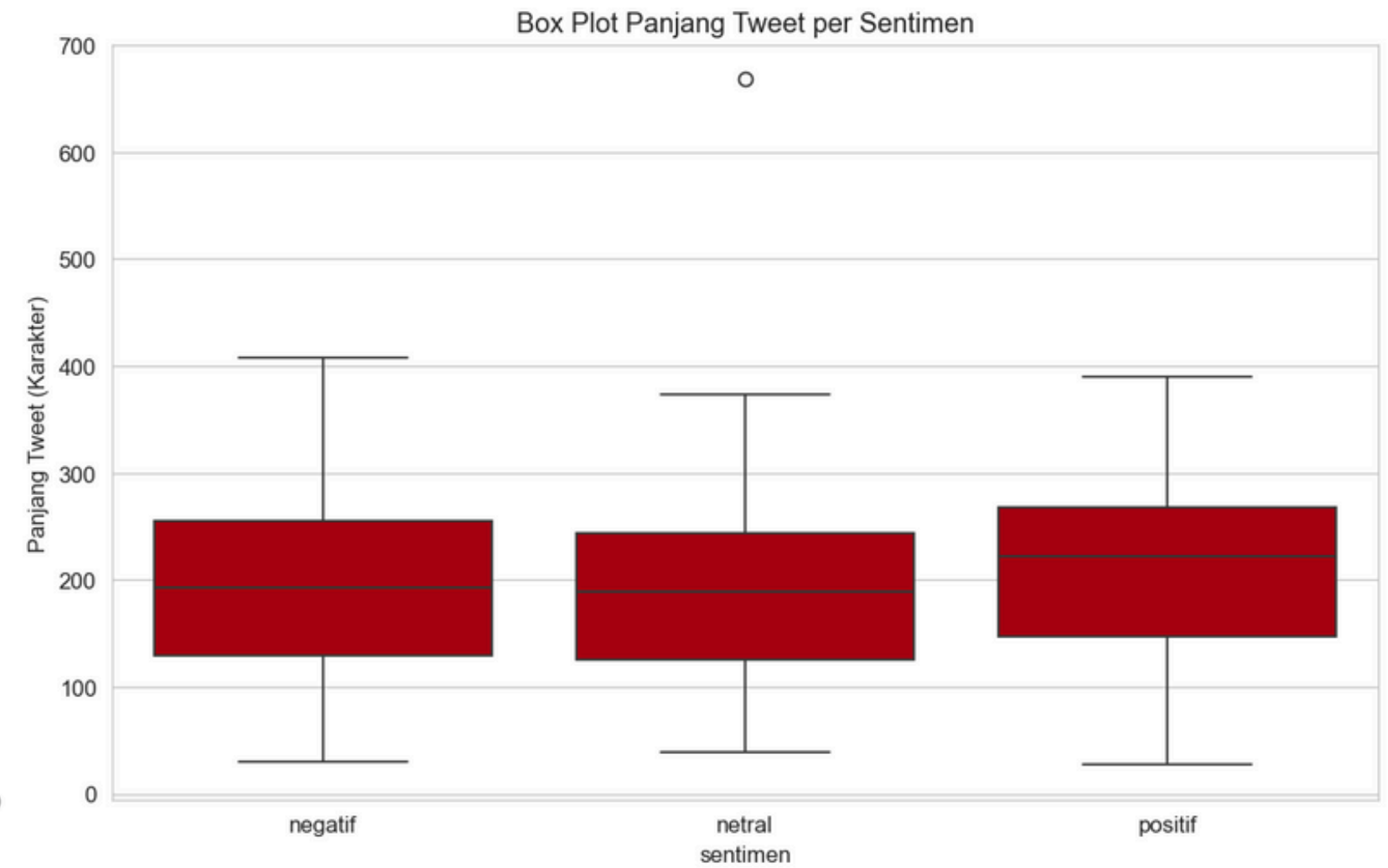
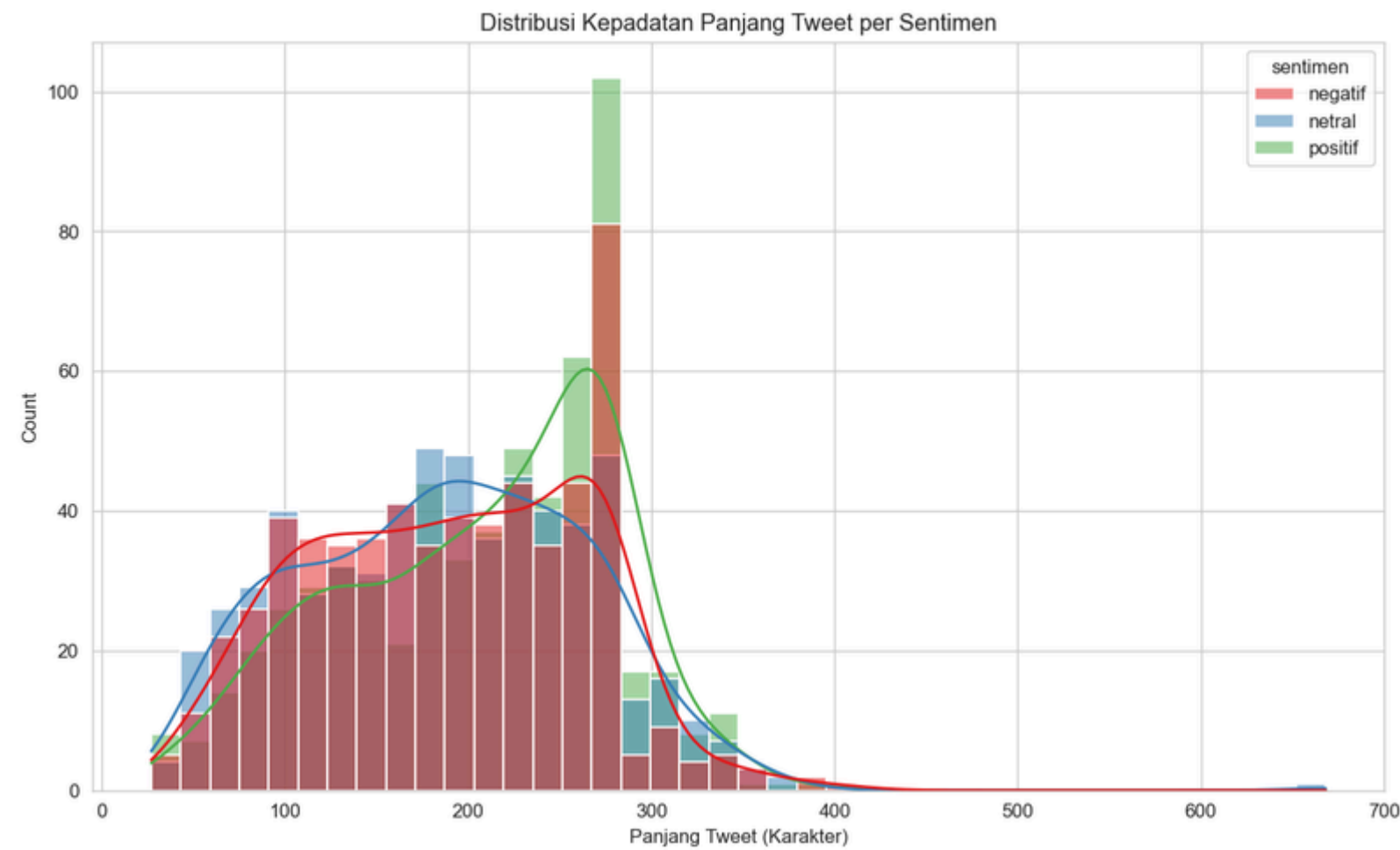
- Fokus proyek adalah mengklasifikasikan tweet ke dalam 3 kategori: Positif, Negatif, dan Netral, menggunakan dataset yang telah disediakan.
- Proyek ini memiliki batasan, yaitu tidak dirancang untuk mendeteksi sarkasme dan tidak menganalisis metadata tweet seperti jumlah likes atau retweet.

Eksplorasi Analisis Data

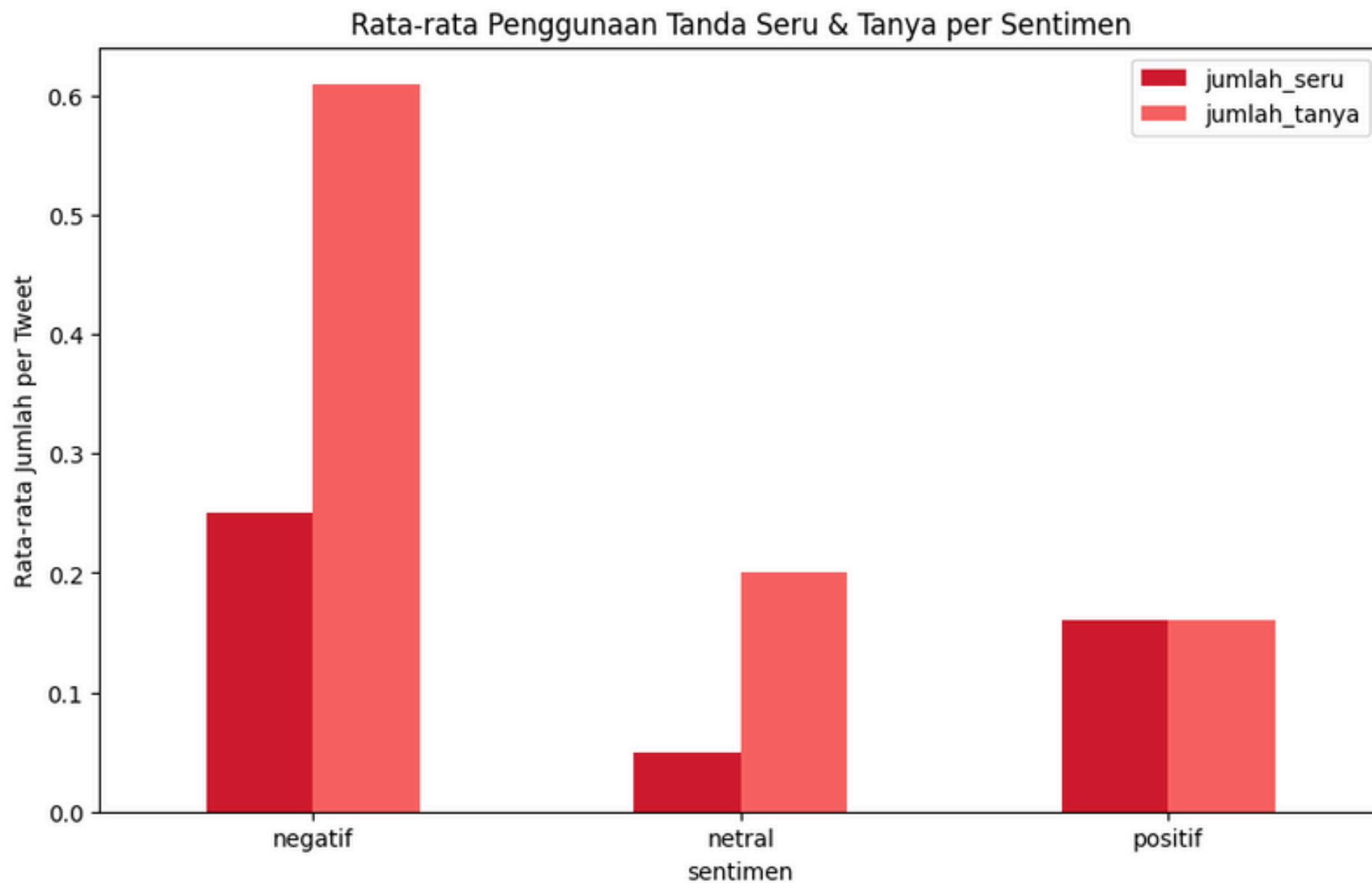
- Cuitan di twitter tentang debat Pilpres 2019.
- Jumlah keseluruhan cuitan: 1.815 cuitan
- Jumlah Tweet per Sentimen: positif (612 cuitan), netral (607 cuitan), dan negatif (596 cuitan).
- Datanya sentimennya cukup berimbang.

	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!

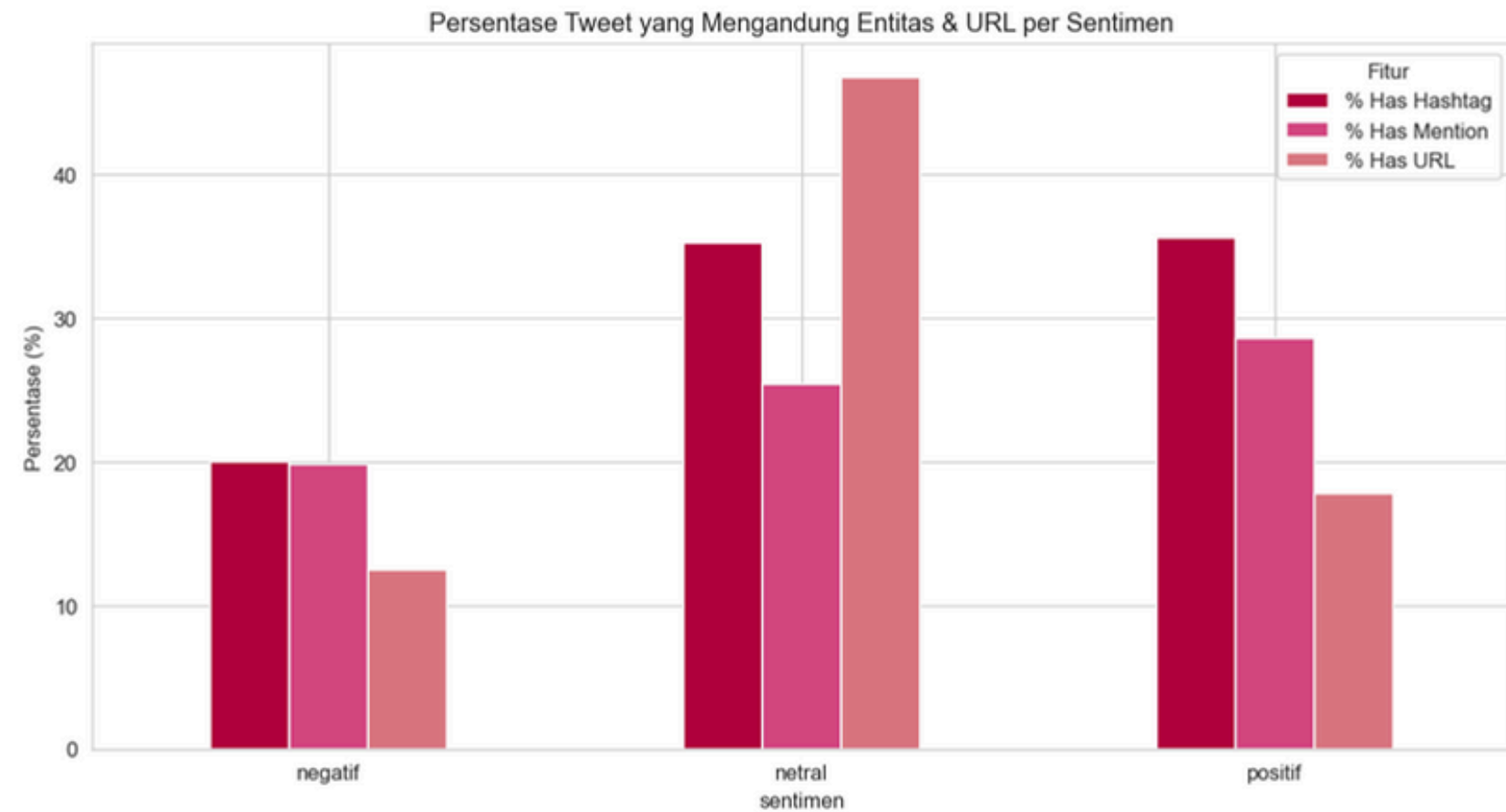




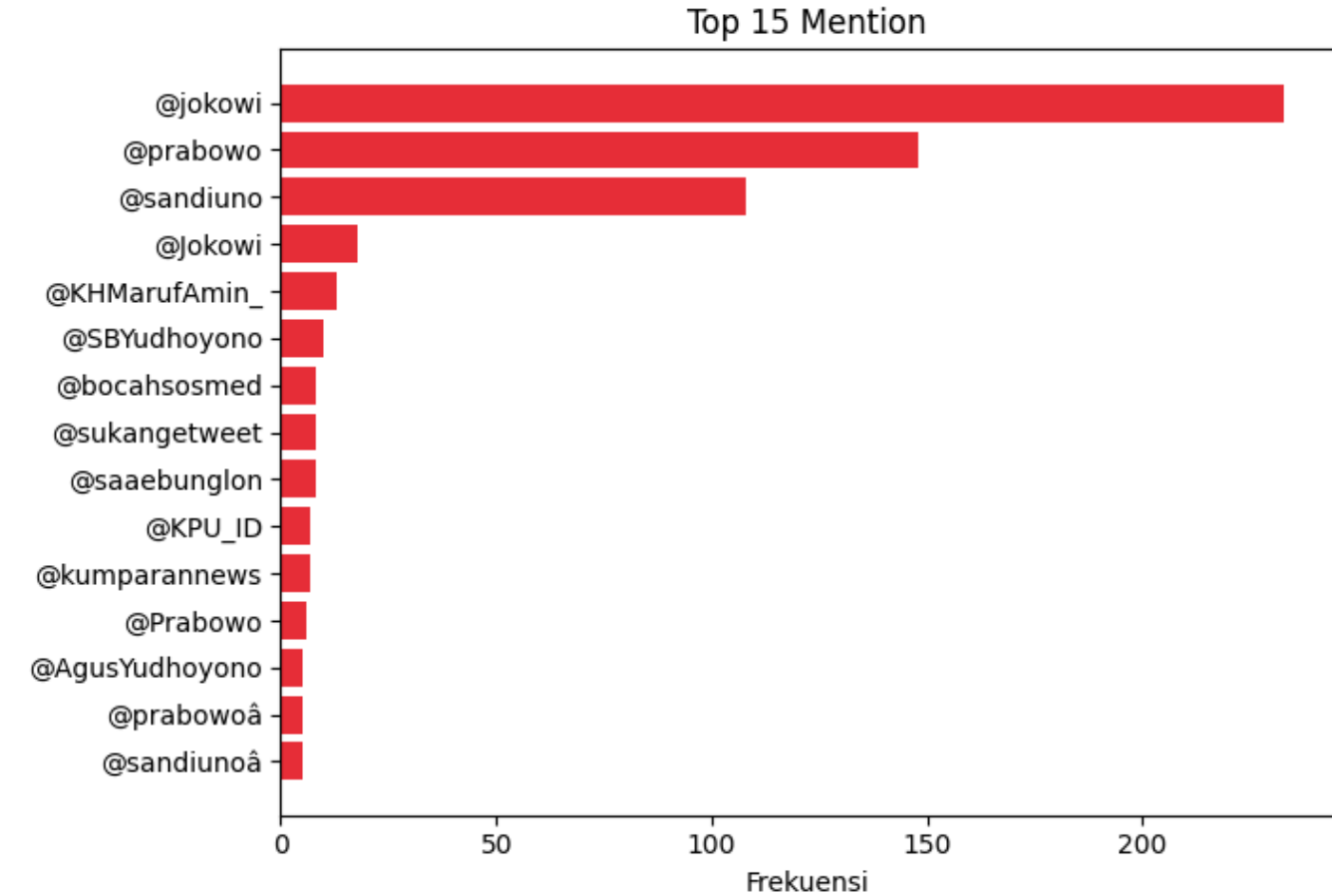
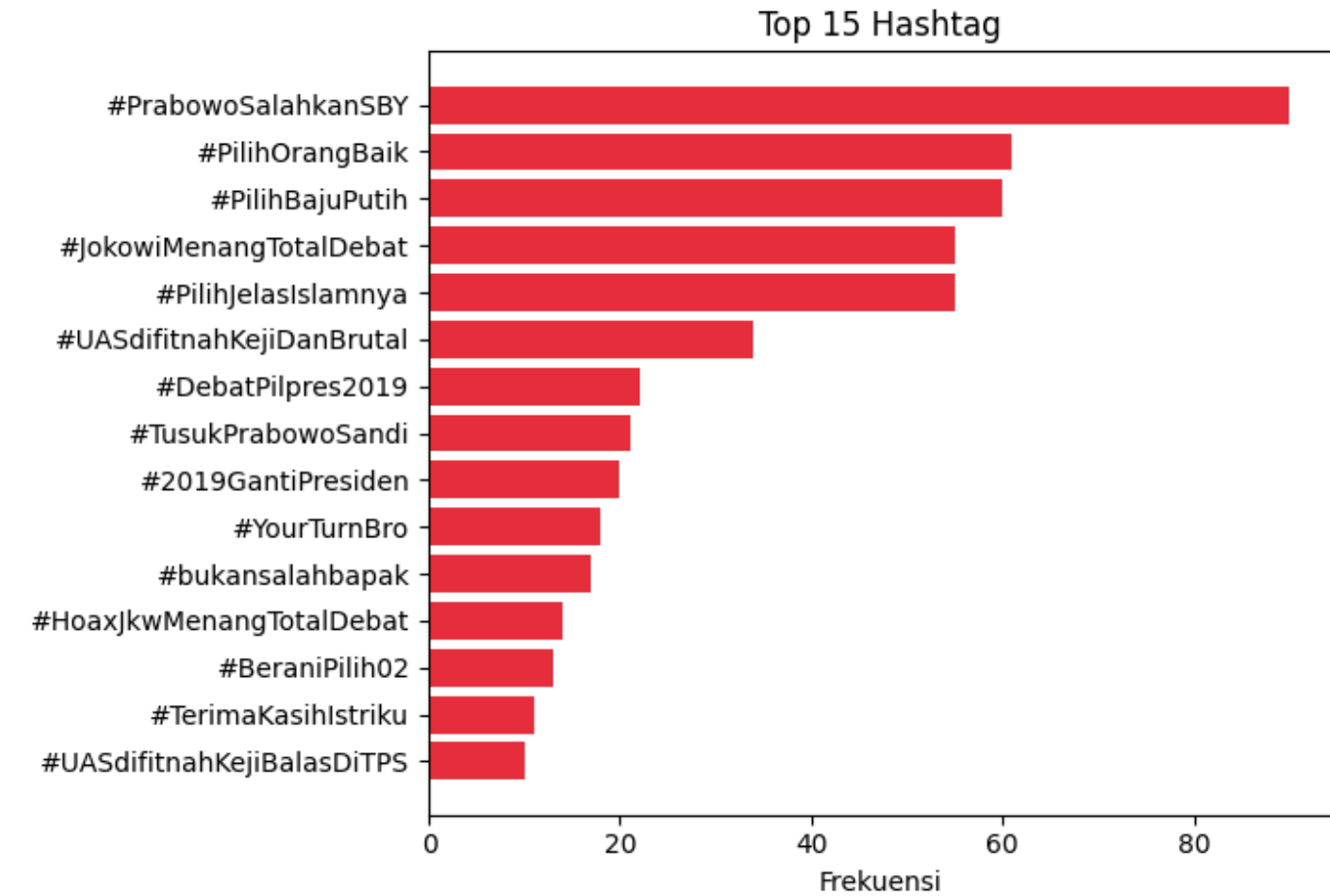
- Grafik di atas menunjukkan sebaran panjang tweet (dalam jumlah karakter) untuk tiga kelas sentimen: positif, netral, dan negatif.
- Sebagian besar tweet berada di rentang 100–300 karakter, dengan puncak sekitar 250 karakter.
- Distribusi ketiga kelas sentimen tampak relatif mirip, tetapi tweet positif 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.

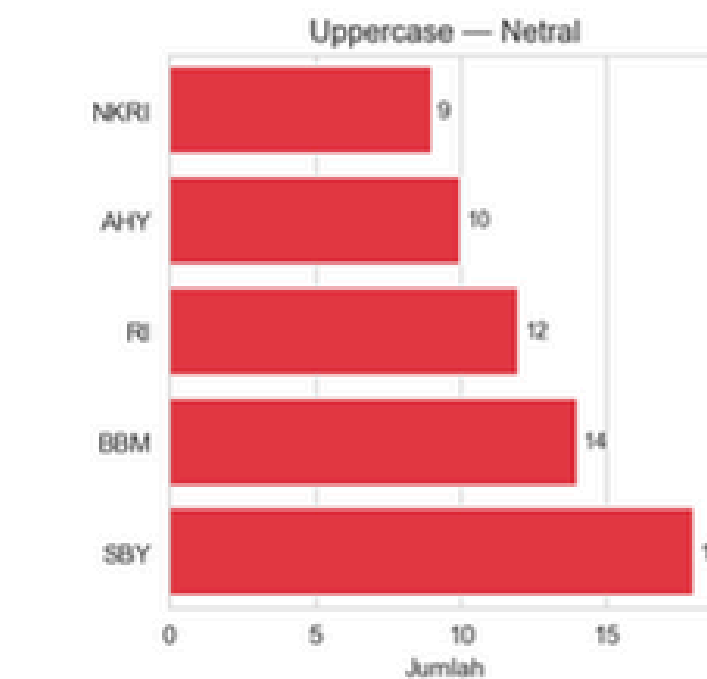
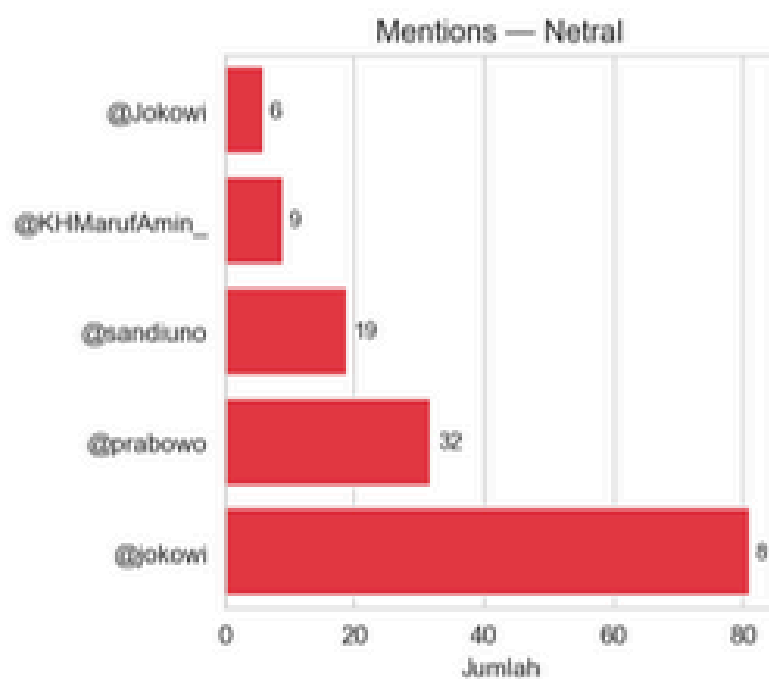
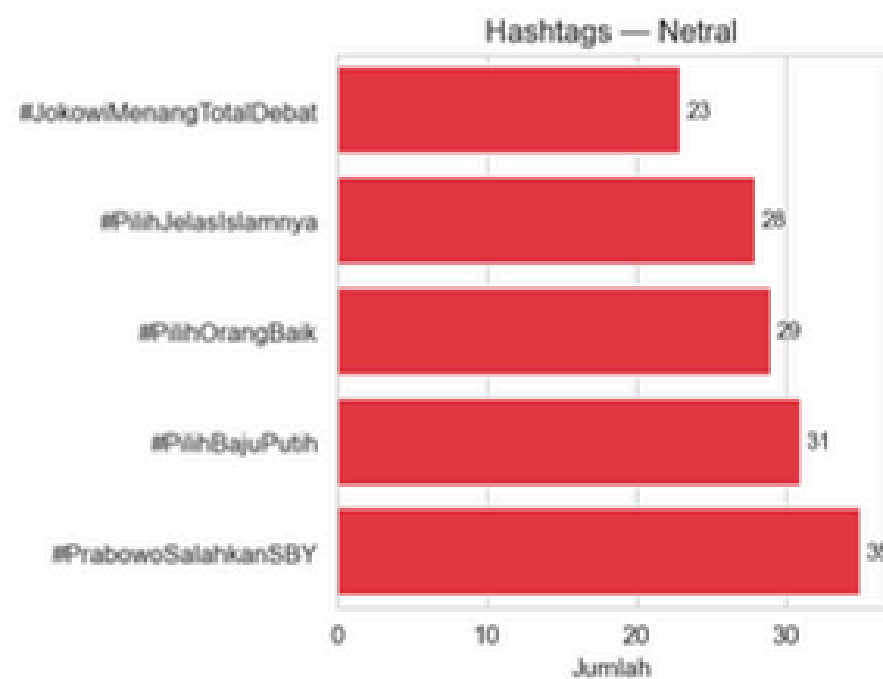
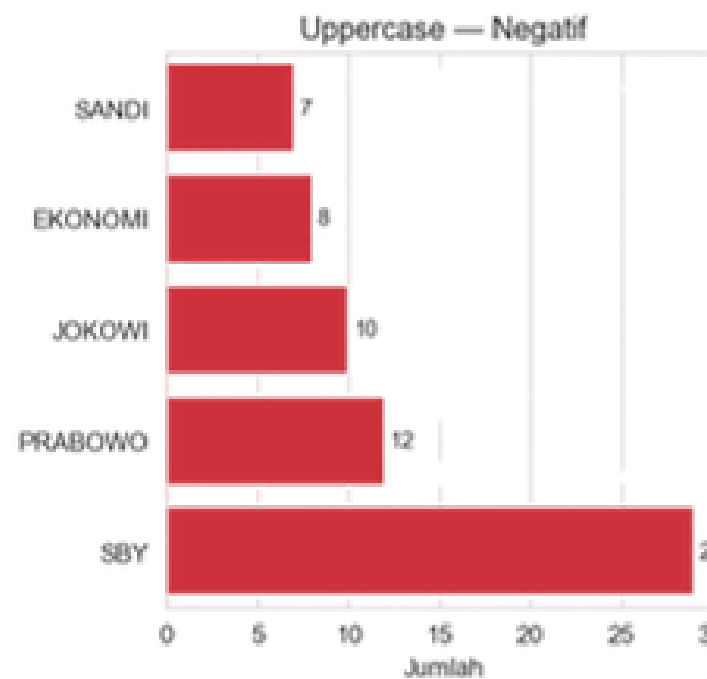
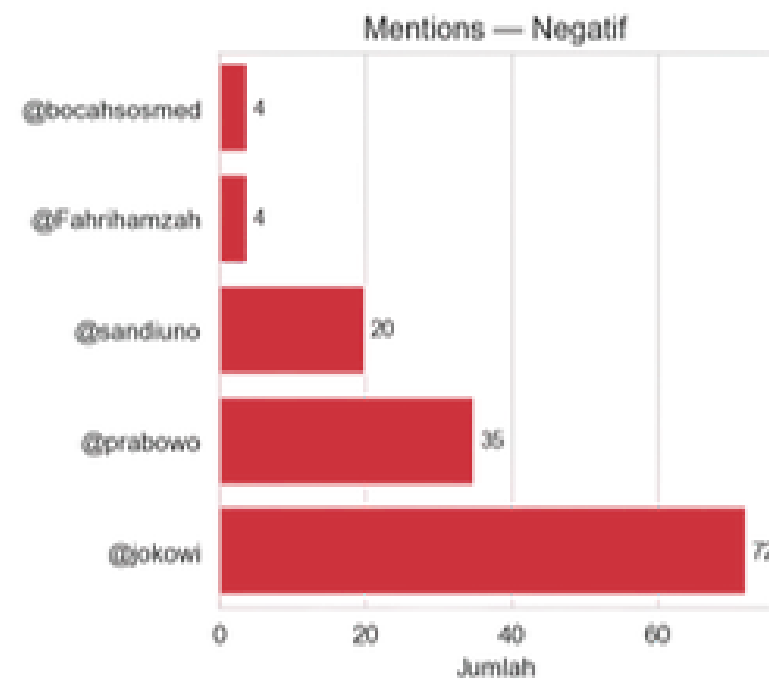
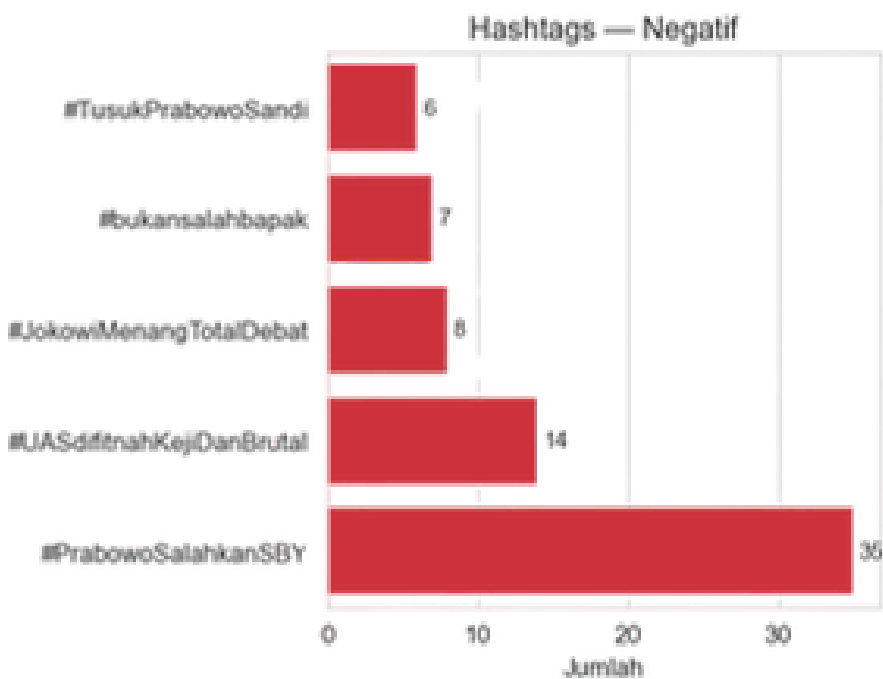
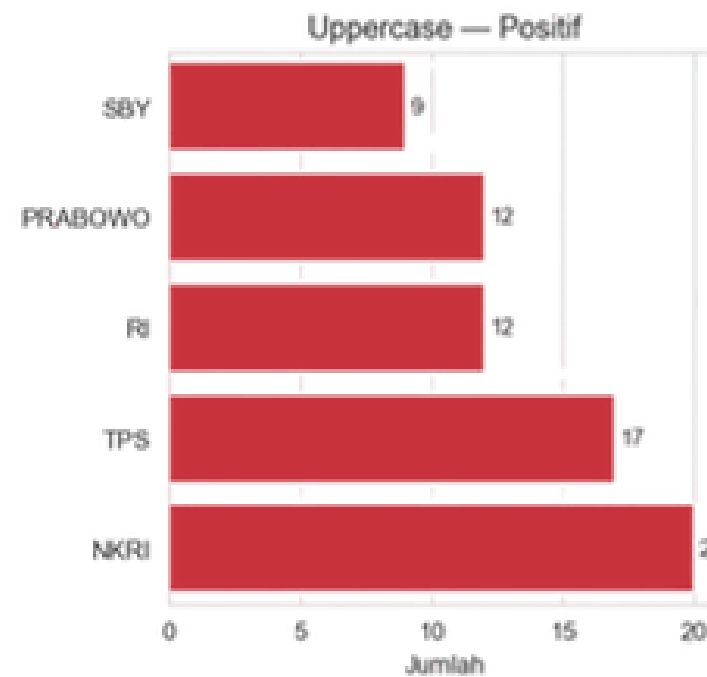
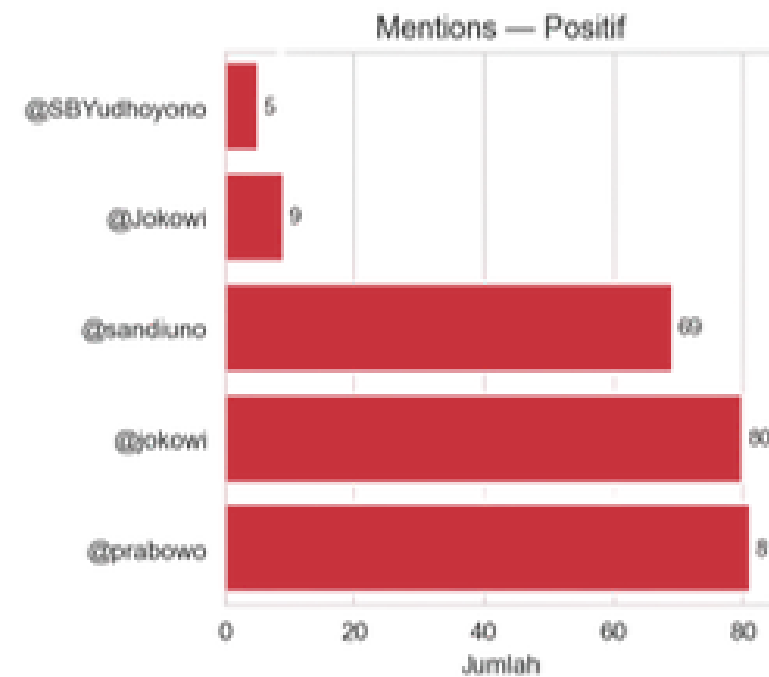
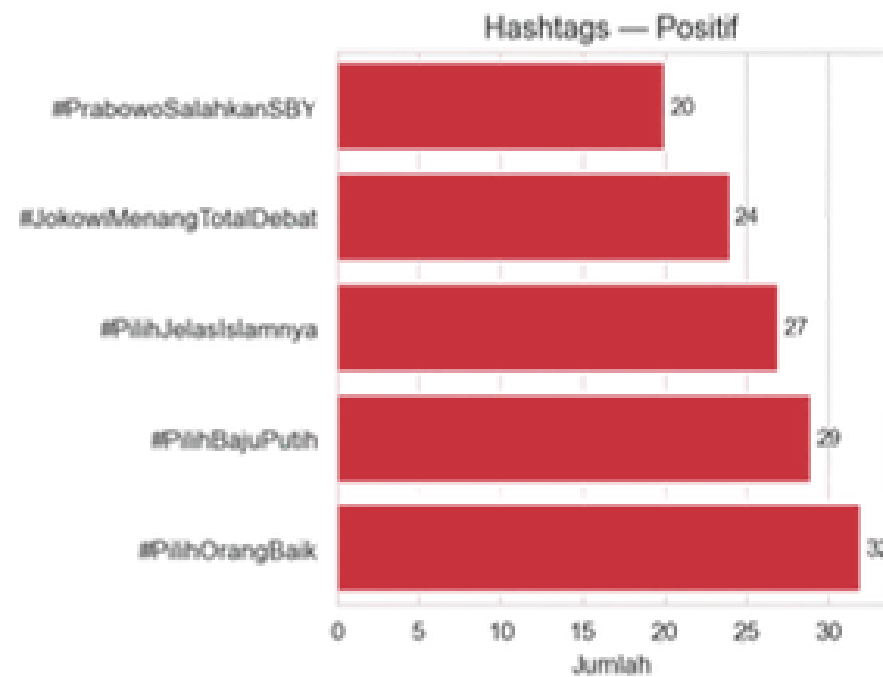


- Tweet ber-sentimen negatif paling banyak mengandung tanda tanya dan tanda seru, dengan rata-rata tanda tanya mencapai sekitar 0,6 per tweet.
- Tweet netral relatif jarang menggunakan tanda baca ekspresif.
- Tweet positif menggunakan tanda seru dan tanda tanya dalam jumlah yang hampir sama, tetapi lebih 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

Transformer Model:

- GloVe
- 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	0.61
netral	0.59	0.68	0.63	
positif	0.59	0.52	0.55	

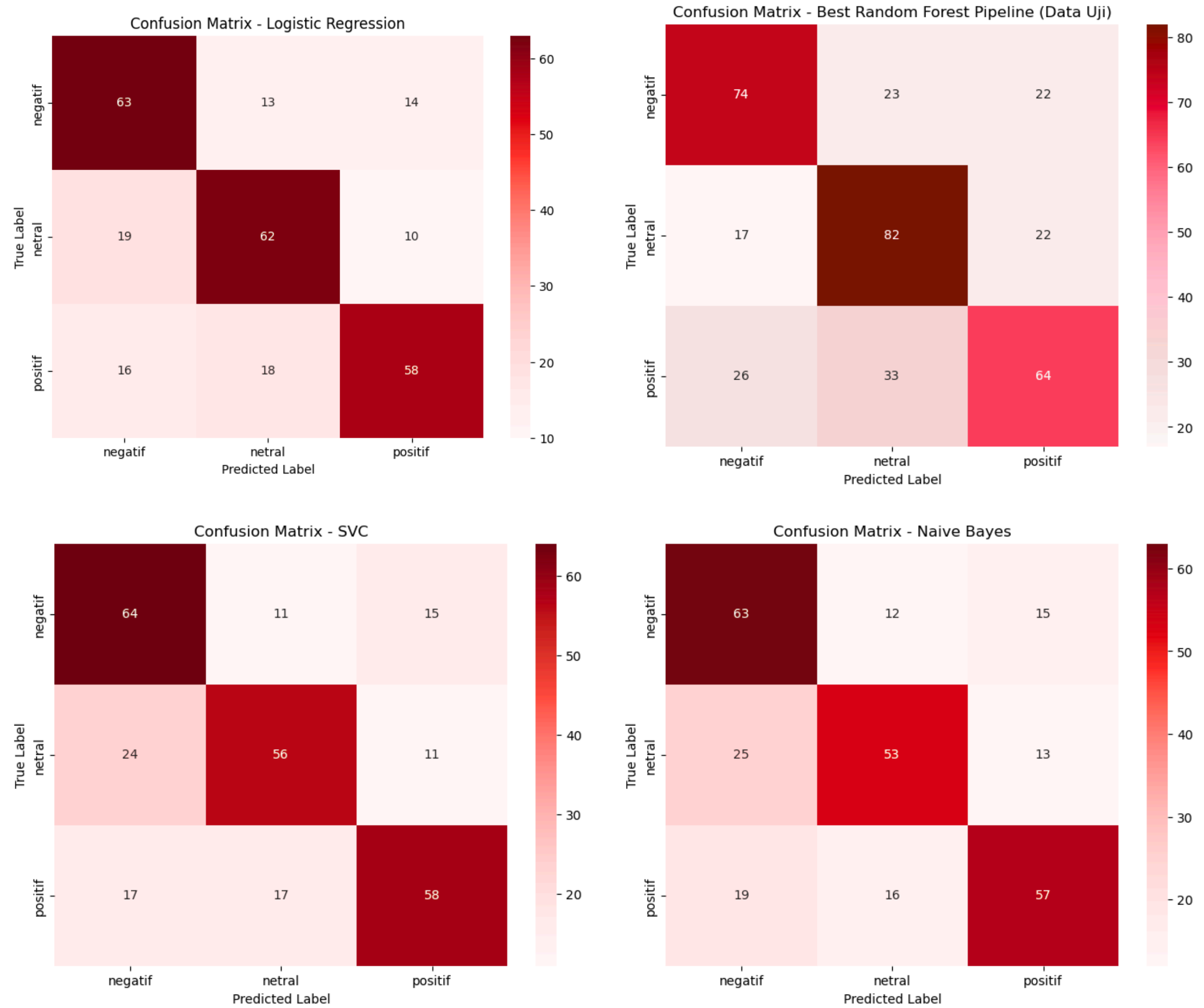
Naive Bayes

	Precision	Recall	F1-score	Accuracy
negatif	0.59	0.70	0.64	0.63
netral	0.65	0.58	0.62	
positif	0.67	0.62	0.64	

SVM

	Precision	Recall	F1-score	Accuracy
negatif	0.61	0.71	0.66	0.65
netral	0.67	0.62	0.64	
positif	0.69	0.63	0.66	

Matrik Evaluasi

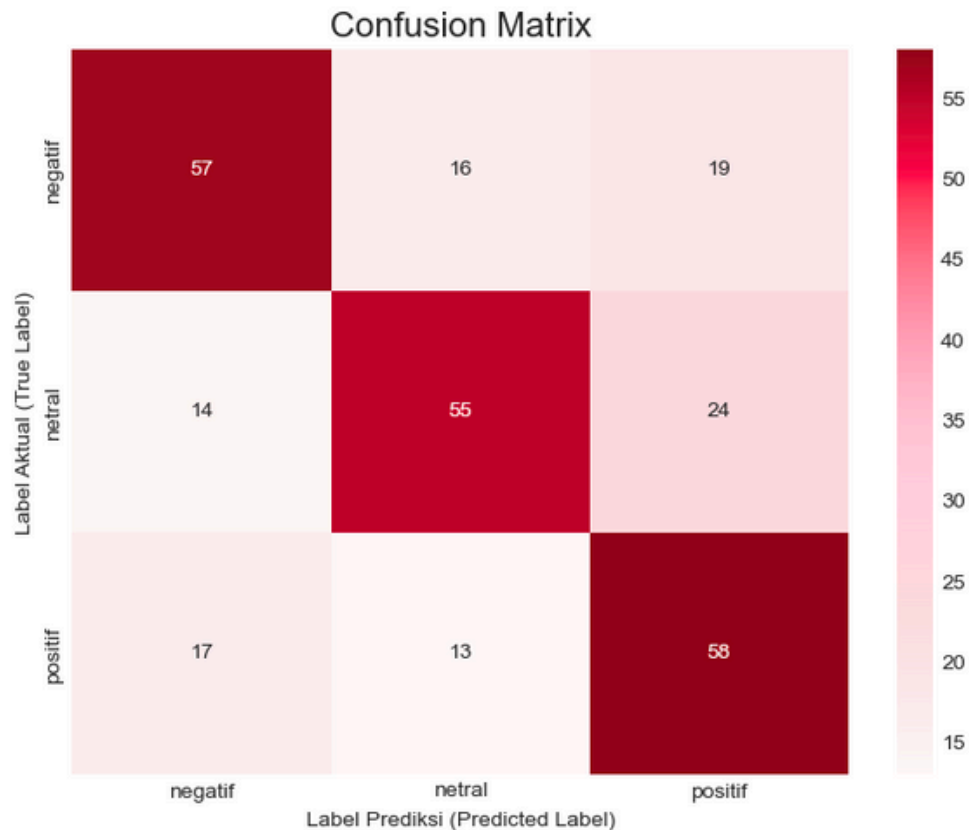
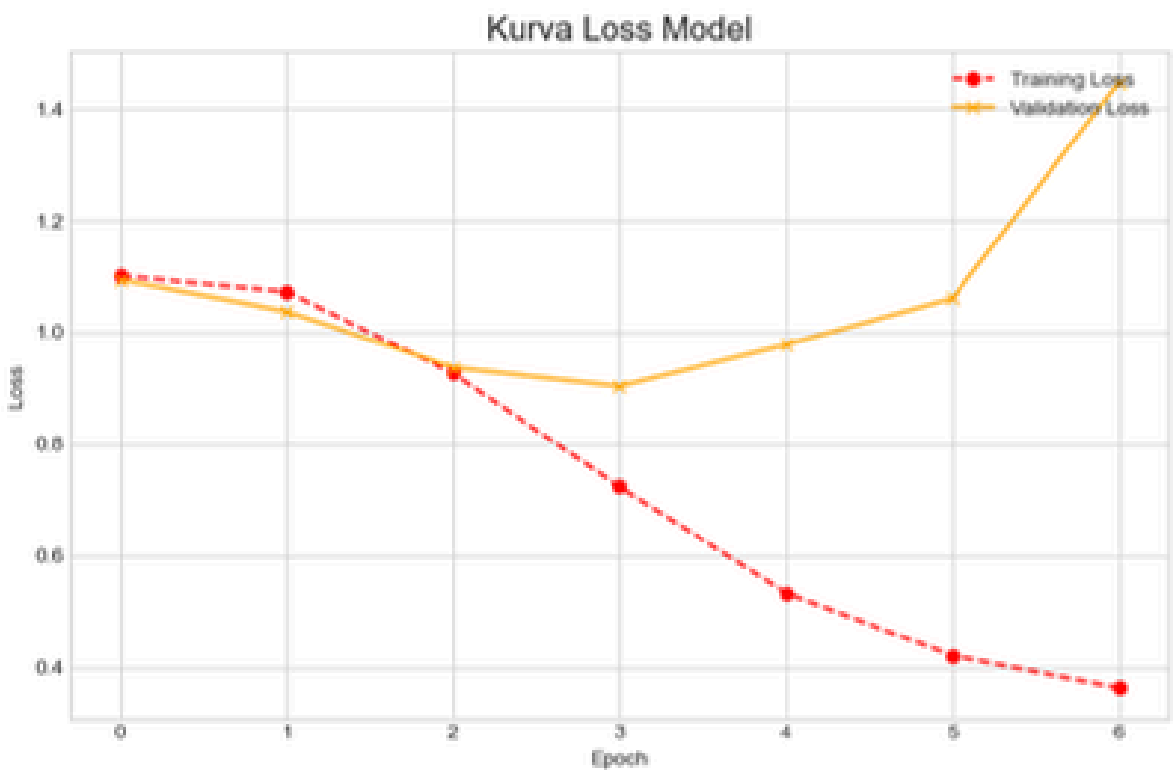
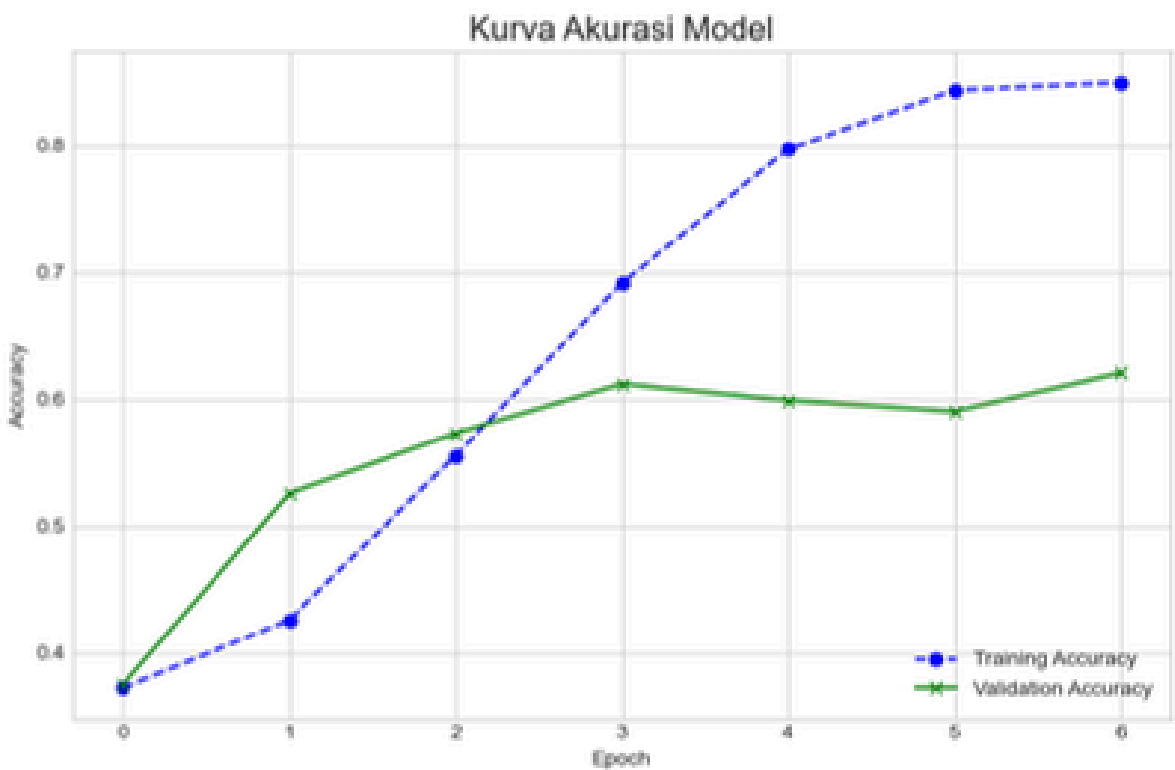


Model	Akurasi Latih	Akurasi Uji	Selisih (Overfitting)
Logistic Regression	79,31%	67,03%	12,28%
SVM	73,80%	63,37%	10,43%
Naive Bayes	81,19%	65,20%	15,99%
Random Forest	82,85%	60,61%	22,24%

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

Deep Learning BiLSTM

Analisis Kurva Pelatihan Model



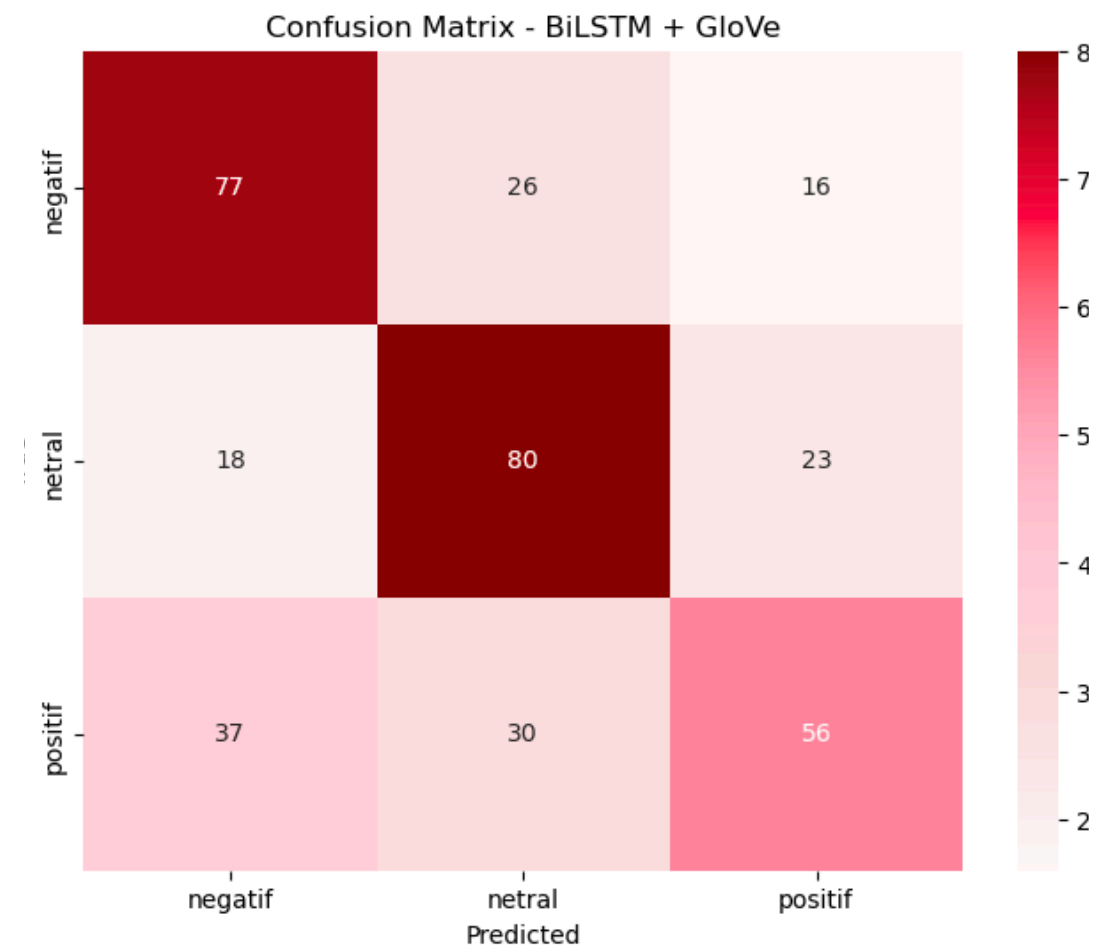
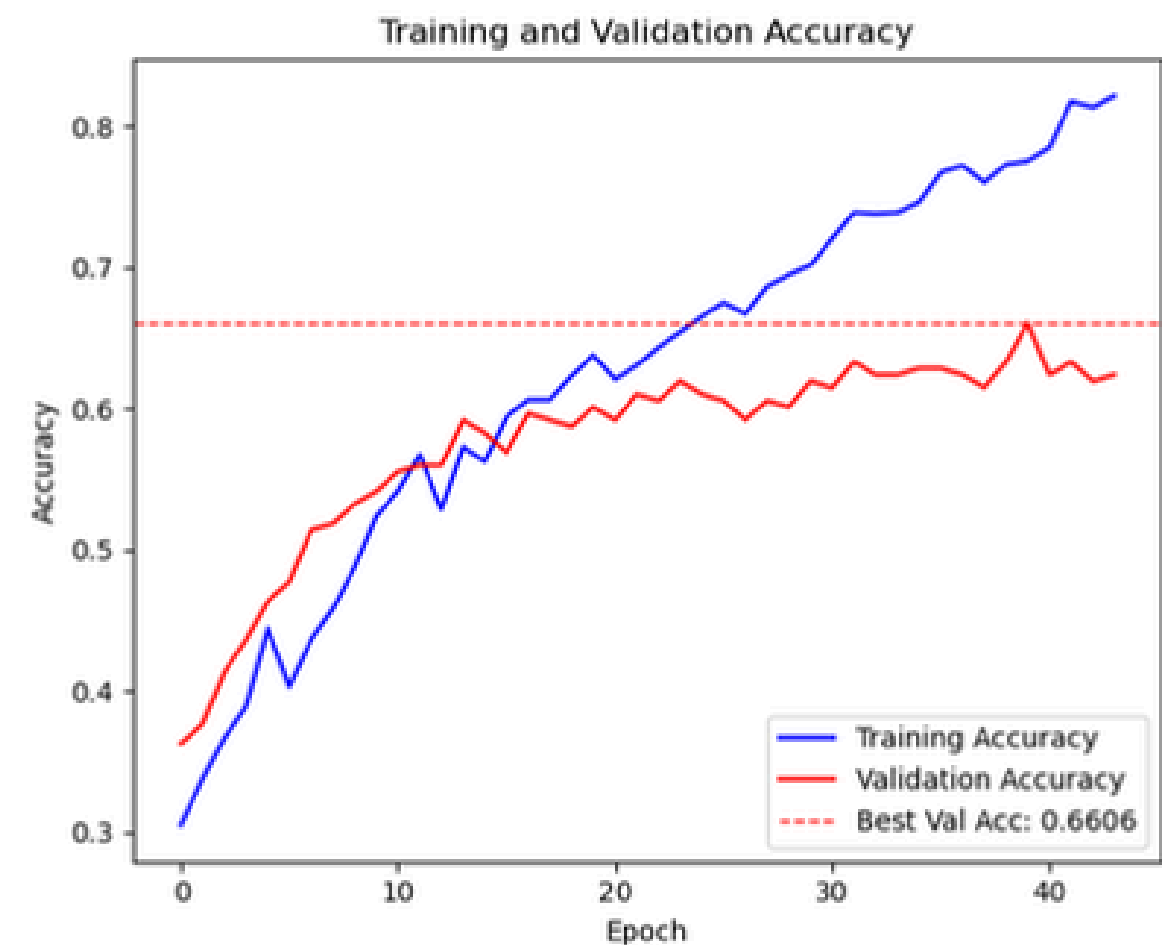
Akurasi Training : 0.8496
Akurasi Testing : 0.6227

Classification Report:				
	precision	recall	f1-score	support
negatif	0.65	0.62	0.63	92
netral	0.65	0.59	0.62	93
positif	0.57	0.66	0.61	88
accuracy			0.62	273
macro avg	0.63	0.62	0.62	273
weighted avg	0.63	0.62	0.62	273

Analisis Hasil:

- Model ini menunjukkan overfitting cukup parah dan sangat cepat. Performa puncaknya tercapai pada epoch 3, di mana setelah itu kurva loss validation mulai berbalik naik, menandakan model kehilangan kemampuan generalisasi. Terdapat kesenjangan performa yang signifikan antara akurasi training (84%) dan validation (62%).
- Kelemahan utamanya adalah ketidakmampuan mengenali sentimen netral, dengan kecenderungan tinggi untuk salah mengklasifikasikannya sebagai positif. Ini mengindikasikan model terlalu reaktif terhadap kata-kata berkonotasi positif dan gagal memahami konteks kalimat yang ambigu, meskipun menggunakan arsitektur bidirectional.

Deep Learning BiLSTM - GloVe



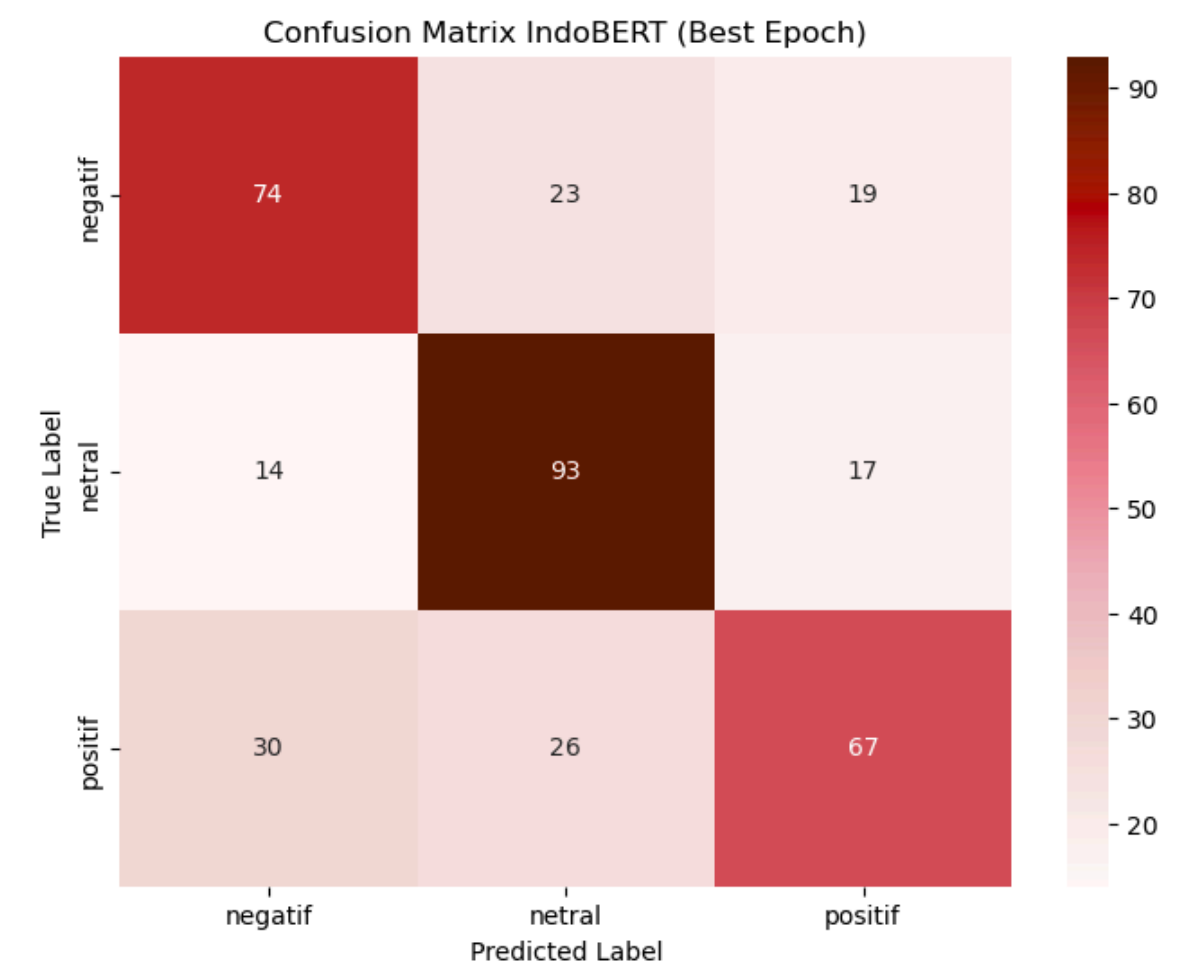
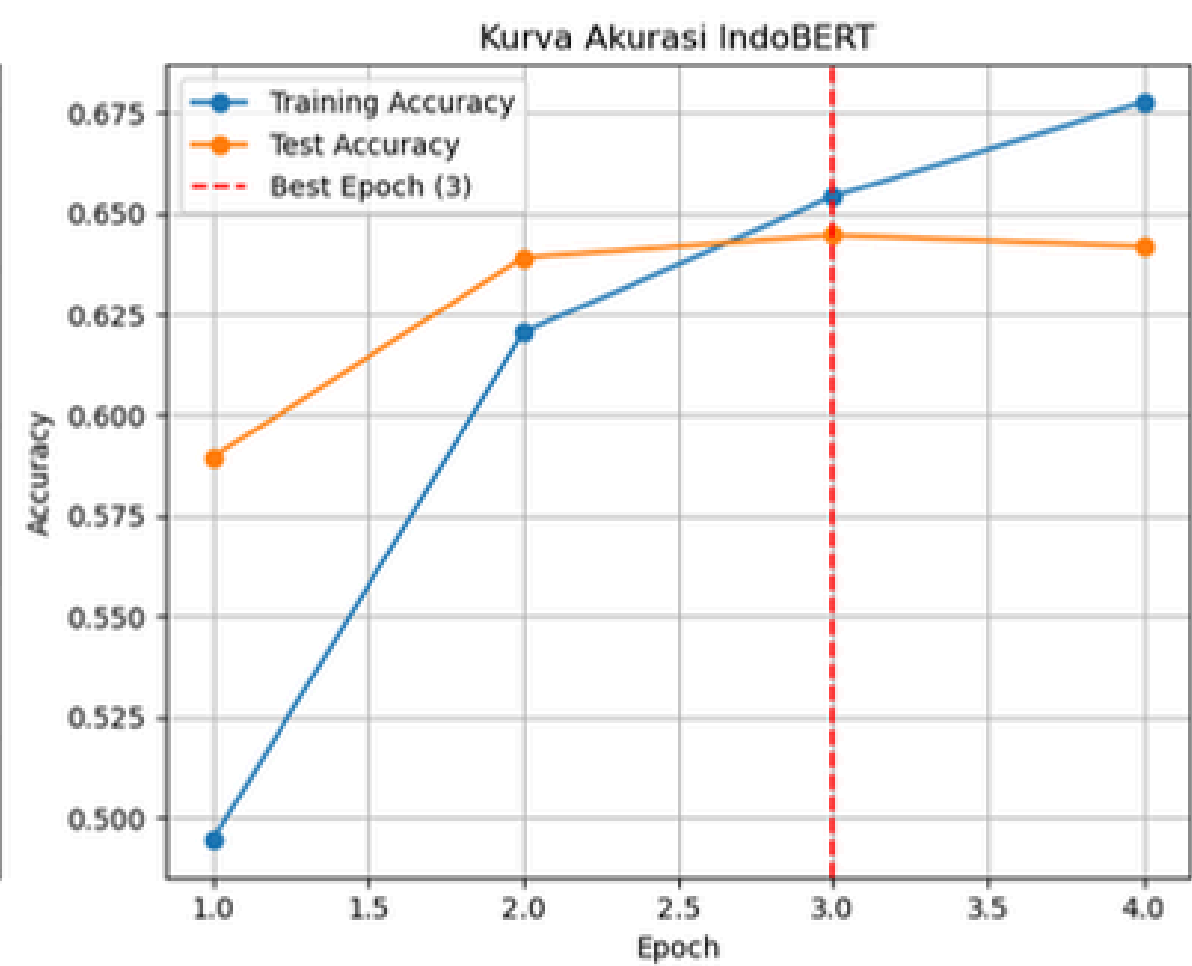
Akurasi Training: 0.8154
Akurasi Test: 0.6612

Classification Report:				
	precision	recall	f1-score	support
negatif	0.66	0.68	0.67	116
netral	0.69	0.75	0.72	124
positif	0.62	0.55	0.59	123
accuracy			0.66	363
macro avg	0.66	0.66	0.66	363
weighted avg	0.66	0.66	0.66	363

Analisis Hasil:

- Akurasi & Loss: Akurasi training terus naik ke 0.81, sementara akurasi validation stagnan di 0.66. Loss validation juga berhenti membaik setelah epoch 15, menunjukkan overfitting yang jelas.
- Interpretasi Overfitting: Model terlalu "menghafal" data latih. Performa terbaik pada data baru (validasi) sebenarnya sudah tercapai di pertengahan proses training, namun model terus dilatih pada data yang sama.
- Analisis Kesalahan: Kebingungan utama terjadi pada kelas positif. Model sering salah memprediksi sentimen yang seharusnya positif menjadi negatif (37 kasus) dan netral (30 kasus).
- Kesimpulan: Batas keputusan (decision boundary) model untuk sentimen positif sangat lemah. Model gagal mengidentifikasi pola unik untuk kelas ini, sehingga mudah tertukar dengan kelas lain.

Deep Learning BiLSTM - IndoBERT



Akurasi Training: 0.6543
Akurasi Test: 0.6446

	precision	recall	f1-score	support
negatif	0.63	0.64	0.63	116
netral	0.65	0.75	0.70	124
positif	0.65	0.54	0.59	123
accuracy			0.64	363
macro avg	0.64	0.64	0.64	363
weighted avg	0.64	0.64	0.64	363

Analisis Hasil:

- Penggunaan IndoBERT secara drastis mengurangi overfitting. Tidak seperti model sebelumnya, kurva training dan test bergerak sangat berdekatan, menunjukkan model ini lebih mampu melakukan generalisasi. Akurasi akhirnya (~64%) sebanding dengan model pertama, namun dicapai dengan cara yang jauh lebih stabil.
- Meskipun lebih robust, kelemahan utamanya kembali pada kesulitan mengidentifikasi sentimen positif (recall 0.54), yang sering tertukar dengan negatif dan netral. Dibandingkan model BiLSTM murni yang lemah pada kelas netral, model ini menunjukkan masalah yang berbeda. IndoBERT berhasil mengatasi overfitting, namun belum sepenuhnya memecahkan masalah pemahaman sentimen positif.

Model (Baseline)	Akurasi Latih	Akurasi Uji	Selisih (Overfitting)
Logistic Regression	79,31%	67,03%	12,28%
SVM	73,80%	63,37%	10,43%
Naive Bayes	81,19%	65,20%	15,99%
Random Forest	82,85%	60,61%	22,24%
Model (Deep Learning)	Akurasi Latih	Akurasi Uji	Selisih (Overfitting)
BiLSTM	84,96%	62,27%	22,69%
BiLSTM + GloVe	81,54%	66,12%	15,42%
BiLSTM + IndoBERT	65,43%	64,46%	0,97%

Model Baseline *Logistic Regression* memberikan akurasi uji terbaik dengan tingkat overfitting yang masih wajar, menunjukkan bahwa model sederhana sudah cukup efektif untuk dataset berukuran kecil (1815 data).

Model Deep Learning lebih rentan overfit karena keterbatasan data, meskipun arsitektur seperti *BiLSTM* dengan *IndoBERT* mampu menekan overfitting dengan baik. Namun, akurasi tetap sulit meningkat, kemungkinan akibat adanya banyak sarkasme dalam data yang menyulitkan model memahami konteks sentimen.

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 regulisasi.
- Model sekuens (BiLSTM, LSTM + GloVe) meningkat pada sebagian pola lokal, tetapi cepat overfit, berbeda dengan model dengan transformer IndoBERT yang cenderung stabil dan goodfit.
- IndoBERT dan Logistic Regression memberi kinerja tertinggi di validasi namun masih menghadapi kebingungan polaritas kuat negatif \leftrightarrow positif dan drift netral \rightarrow positif.
- Pola error paling konsisten: pertukaran negatif dan positif, sedangkan netral relatif stabil namun rentan terseret ke positif saat ada kata dukungan/topik populer.

Kesimpulan akhir, menjadikan Logistic Regression sebagai model yang cukup handal di project kali ini, karena hasilnya yang cukup baik namun tidak memerlukan banyak sumber daya.

BiLSTM dengan IndoBERT memiliki performa yang paling stabil, walaupun hasil akurasi test nya tidak sebaik Logistic Regression, namun dia memiliki kelebihan menekan overfit dengan baik.

