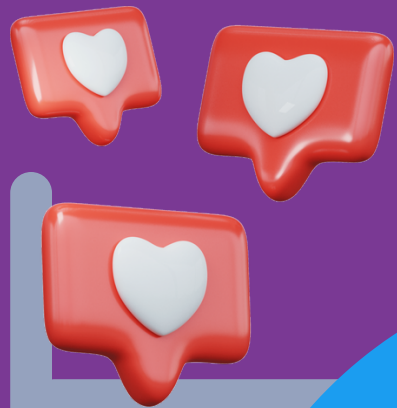


ANALISIS SENTIMEN PADA TWITTER DENGAN RECURRENT NEURAL NETWORK (RNN) **DAN** LONG SHORT-TERM MEMORY (LSTM)

KELOMPOK 3

2 Juli 2023



LATAR BELAKANG

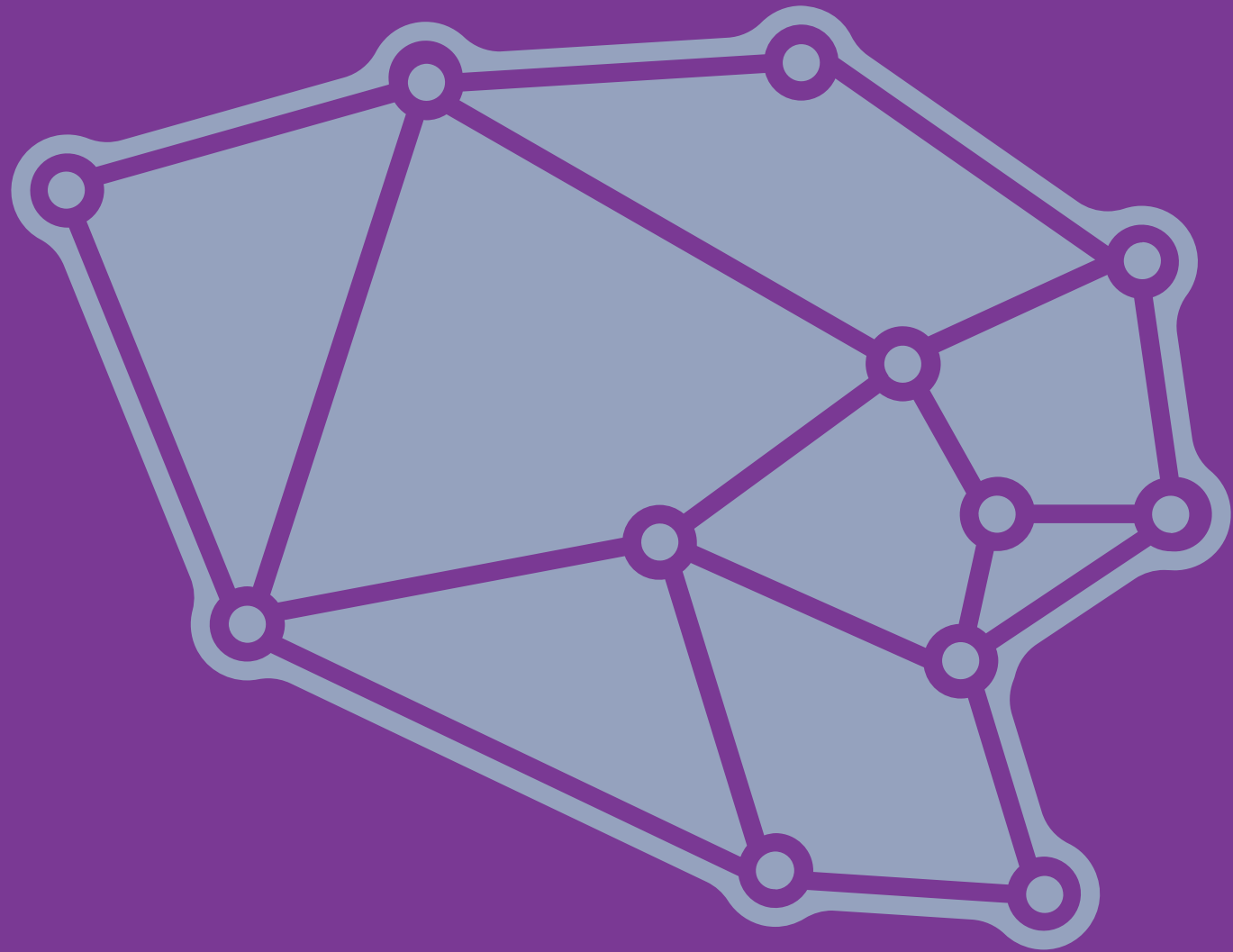
- Twitter merupakan salah satu wadah interaksi daring terbesar pengguna media sosial di Indonesia.
- Angka pengguna aktif Twitter di Indonesia mencapai 24 juta atau 36,2% dari total 66,3 juta pengguna aktif Twitter di Asia Tenggara (DataReportal.com, Februari 2023).
- Perkembangan interaksi daring dalam ekosistem Twitter turut membentuk budaya dan etika digital pengguna didalamnya.



PERMASALAHAN

- Analisis sentimen diperlukan untuk melihat bagaimana kecenderungan sentimen pengguna dalam ekosistem Twitter.
- Automatisasi analisis sentimen dengan deep learning dalam ekosistem Twitter diperlukan untuk memudahkan pengumpulan data *baseline*, sehingga dapat digunakan dalam penentuan strategi kebijakan *stakeholder* Twitter kedepannya.





Tujuan



Membangun model analisis sentimen dengan menggunakan Recurrent Neural Network (RNN) dan Long Short-Term Memory (LSTM)





ANALISIS SENTIMEN

merupakan teknik yang digunakan untuk mendeteksi sentimen yang mendasari sebuah kalimat. Proses ini umumnya mengklasifikasi teks ke dalam kategori positif, negatif, dan netral. Analisis sentimen dapat dilakukan dengan menggunakan deep learning, khususnya dengan model recurrent neural network (RNN) atau long short-term memory (LSTM).

PENGUJIAN MODEL RNN DAN LSTM

PROSES KERJA





PROSES AWAL

Dataset

12.000 data text

Sumber:

- Ibrohim, M. O., & Budi, I. (2019). Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter. 46–57. <https://doi.org/10.18653/v1/w19-3506>.
- Winata, G. I., Aji, A. F., Cahyawijaya, S., Mahendra, R., Koto, F., Romadhony, A., Kurniawan, K., Moeljadi, D., Prasojo, R. E., Fung, P., Baldwin, T., Lau, J. H., Sennrich, R., & Ruder, S. (2022). NusaX: Multilingual Parallel Sentiment Dataset for 10 Indonesian Local Languages. <http://arxiv.org/abs/2205.15960>.

Dataset

Supervised text

- Positif (57%)
- Negatif (32%)
- Neutral (12%)

Normalisasi Teks

Lowercase,
nonalphanumeric,
karakter yang tidak
diperlukan, emoji,
kalimat tidak
baku/alay,
stopwords, stemming

Ekstraksi Fitur

- 100.000 ekstraksi kata
- 20% subset data uji
- 80% data training
- Nilai randomisasi 1



TRAINING MODEL RNN

Parameter & Regularization

- Dimensi vektor: 64
- Unit matriks: 64
- Setting model: Sequential, SimpleRNN
- Dropout unit: 20%
- Dropout: 20
- Recurrent dropout: 20%
- Output/Dense: 3 (positive, negative, neutral), aktivasi softmax.
- Fungsi loss: `sparse_categorical_crossentropy`
- Optimizer: adam
- Learning rate: 0,001
- Epoch 10, dengan early stopping
- batch size: 10

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 78, 100)	10000000
simple_rnn (SimpleRNN)	(None, 64)	10560
dense (Dense)	(None, 3)	195

=====
Total params: 10010755 (38.19 MB)
Trainable params: 10010755 (38.19 MB)
Non-trainable params: 0 (0.00 Byte)



TRAINING MODEL RNN (2)

	precision	recall	f1-score	support
0	0.83	0.73	0.78	770
1	0.77	0.71	0.74	251
2	0.86	0.94	0.90	1379
accuracy			0.85	2400
macro avg	0.82	0.79	0.80	2400
weighted avg	0.84	0.85	0.84	2400

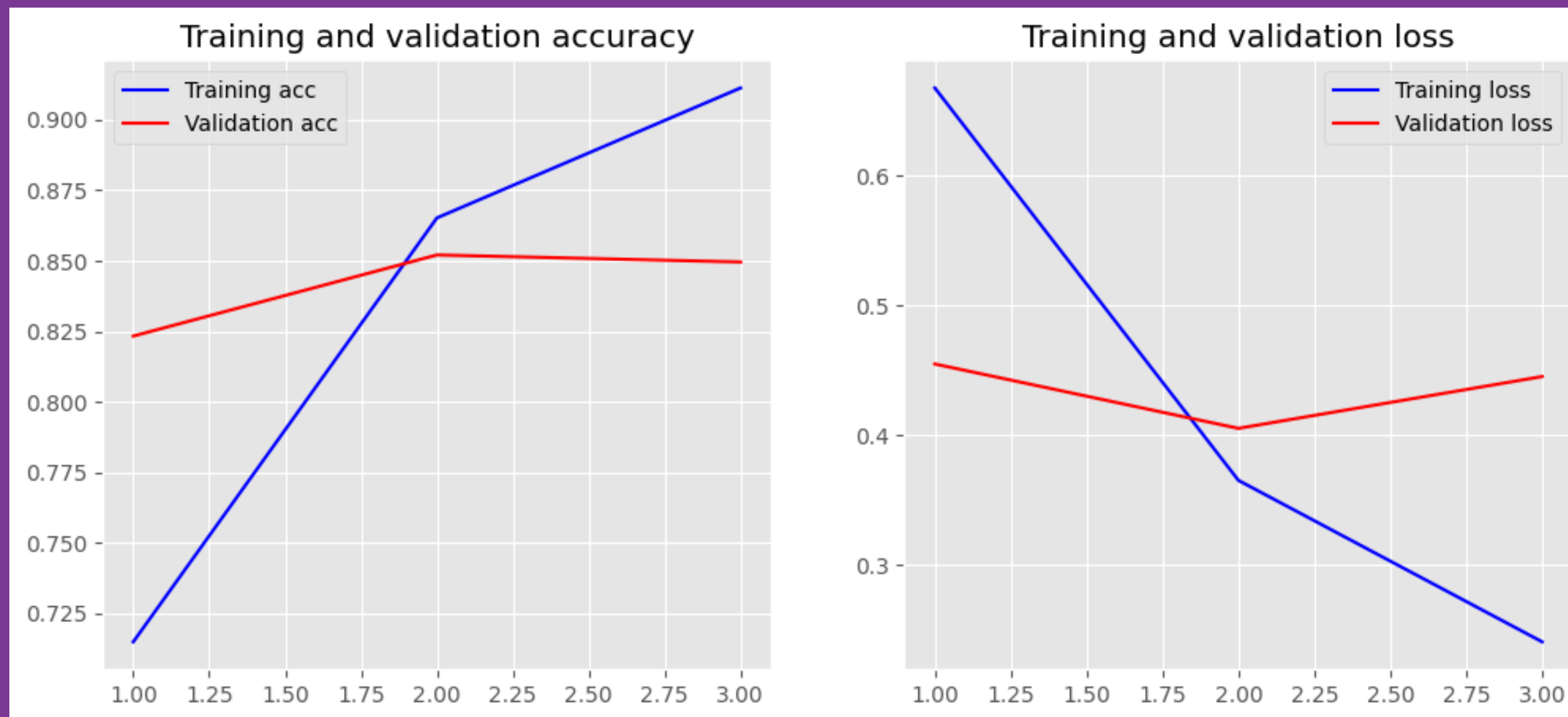
Nilai akurasi rata-rata
validitas silang:

84%



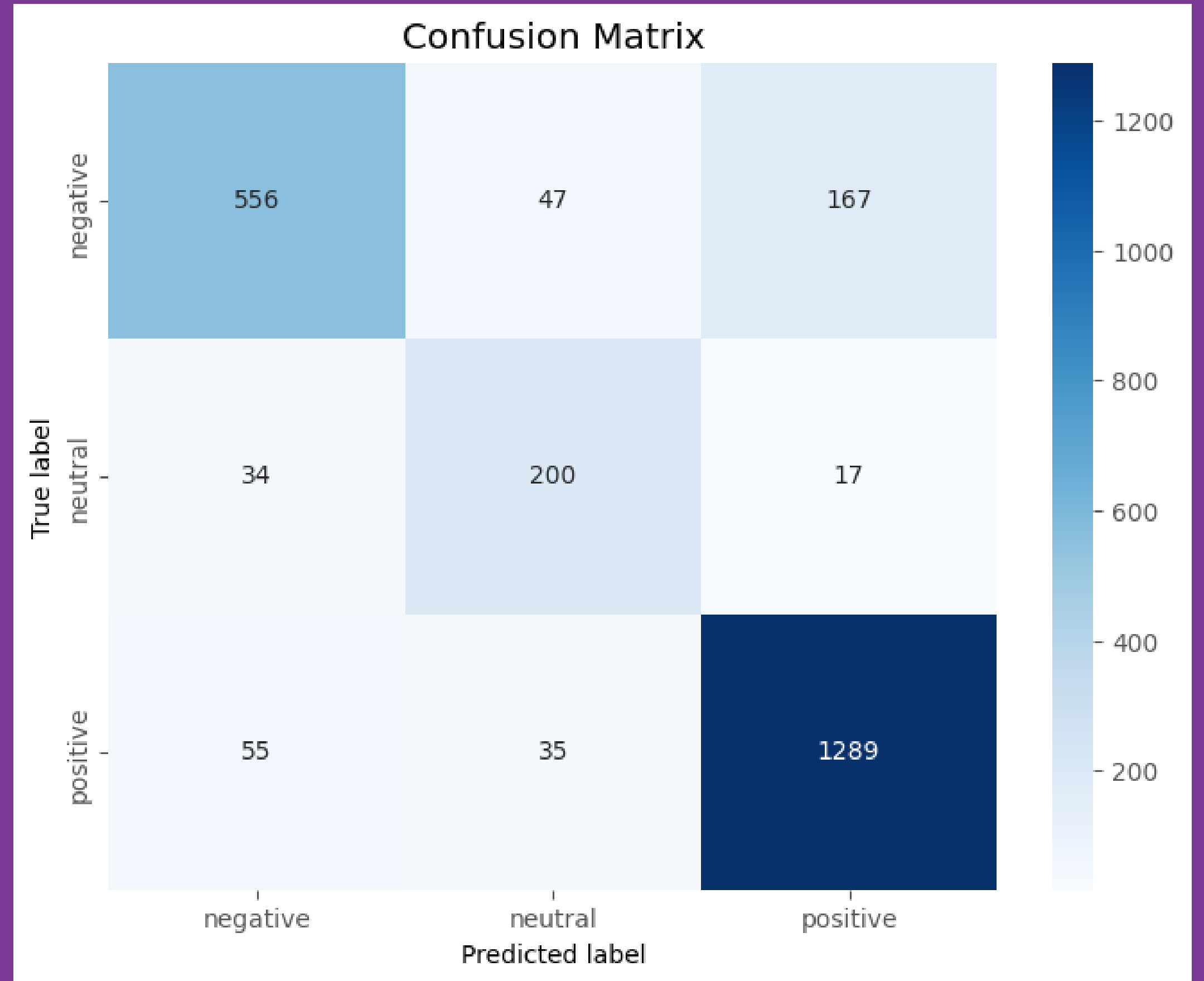
TRAINING MODEL

RNN (3)





CONFUSION MATRIX





TRAINING MODEL LSTM

Parameter & Regularization

- Dimensi vektor: 64
- Unit matriks: 64
- Setting model: Sequential, LSTM
- Dropout unit: 20%
- Dropout: 20
- Recurrent dropout: 20%
- Output/Dense: 3 (positive, negative, neutral), aktivasi softmax.
- Fungsi loss: `sparse_categorical_crossentropy`
- Optimizer: adam
- Learning rate: 0,001
- Epoch 10, dengan early stopping
- batch size: 10

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 78, 100)	10000000
lstm (LSTM)	(None, 64)	42240
dense (Dense)	(None, 3)	195

=====
 Total params: 10,042,435
 Trainable params: 10,042,435
 Non-trainable params: 0



TRAINING MODEL LSTM (2)

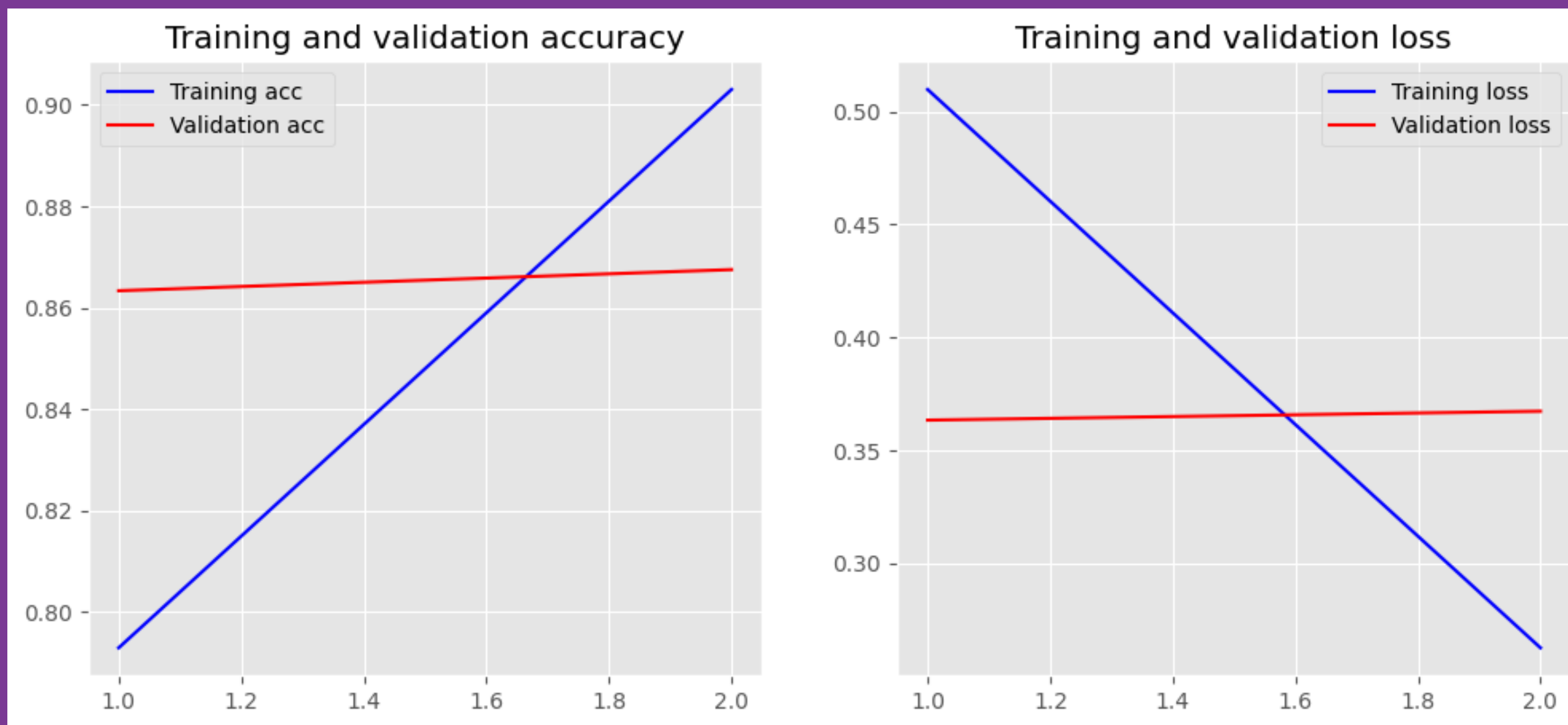
	precision	recall	f1-score	support
0	0.85	0.76	0.81	770
1	0.72	0.76	0.74	251
2	0.89	0.93	0.91	1379
accuracy			0.86	2400
macro avg	0.82	0.82	0.82	2400
weighted avg	0.86	0.86	0.86	2400

Nilai akurasi rata-rata
validitas silang:

86%

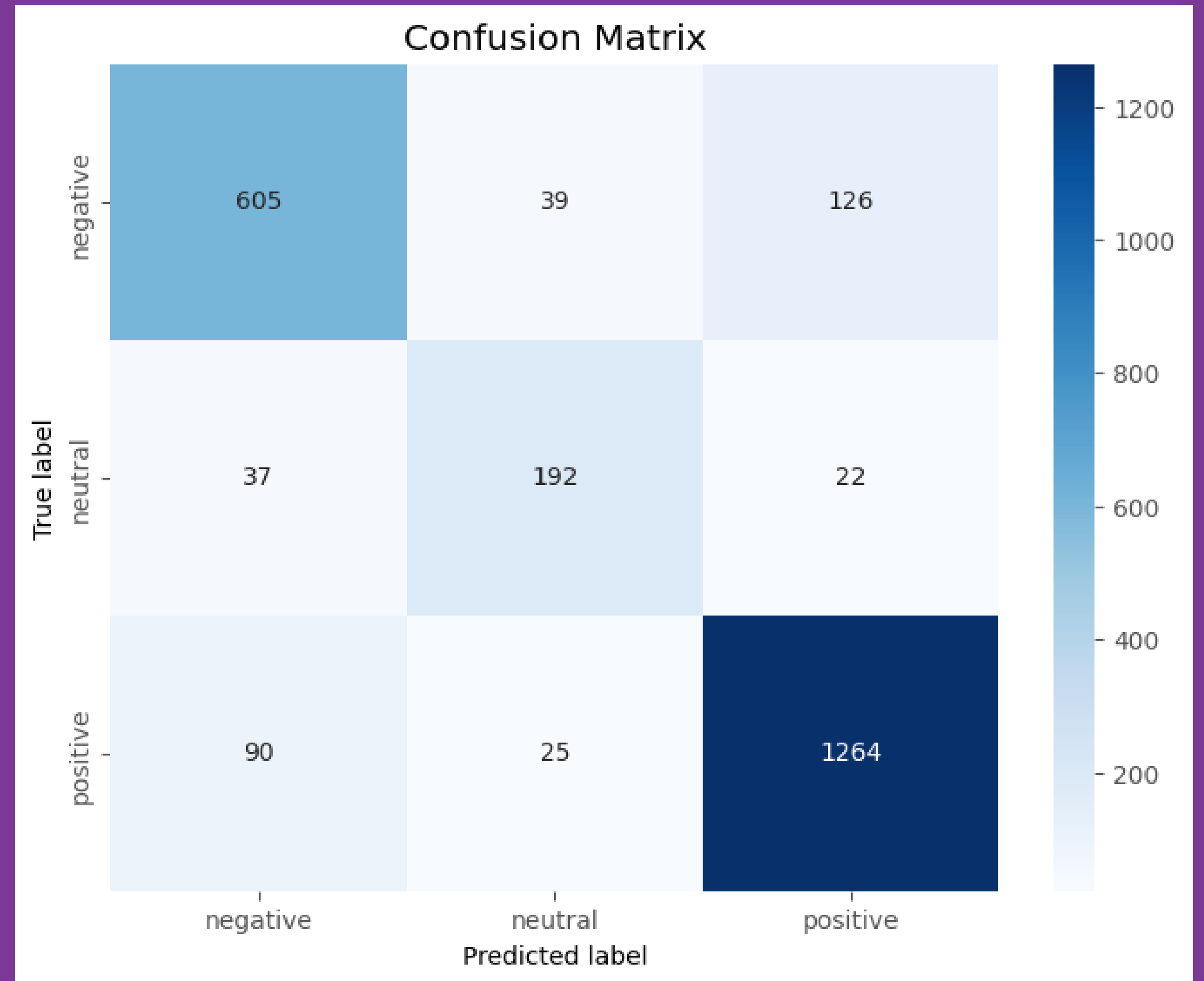


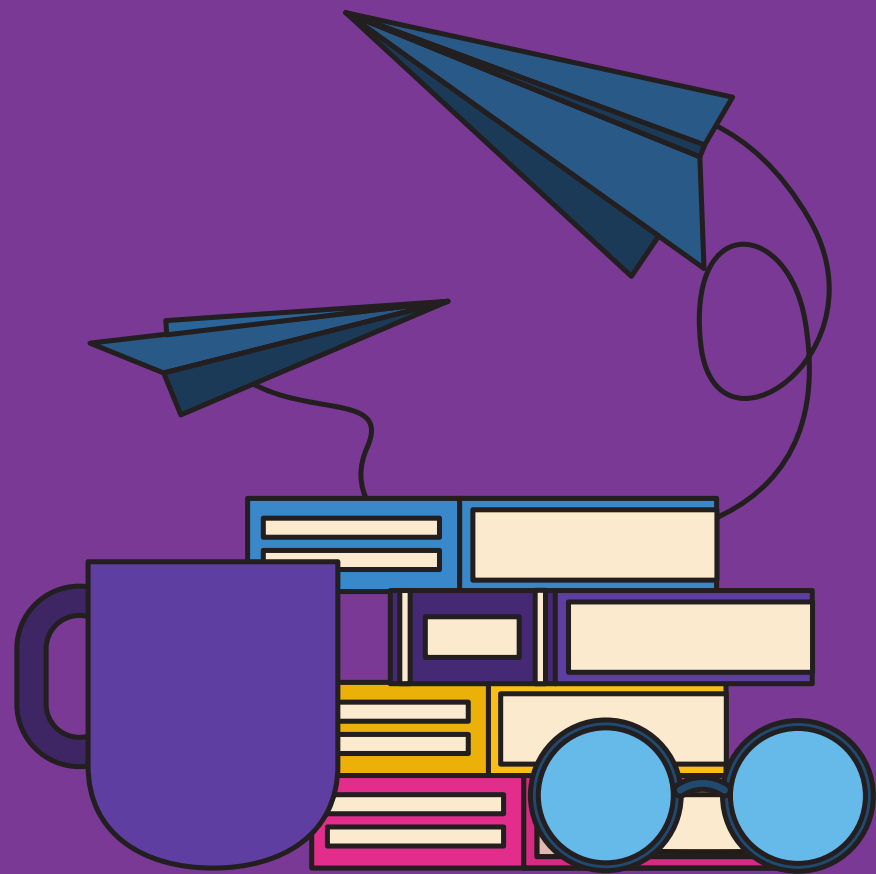
TRAINING MODEL LSTM (3)





CONFUSION MATRIX





KESIMPULAN & REKOMENDASI

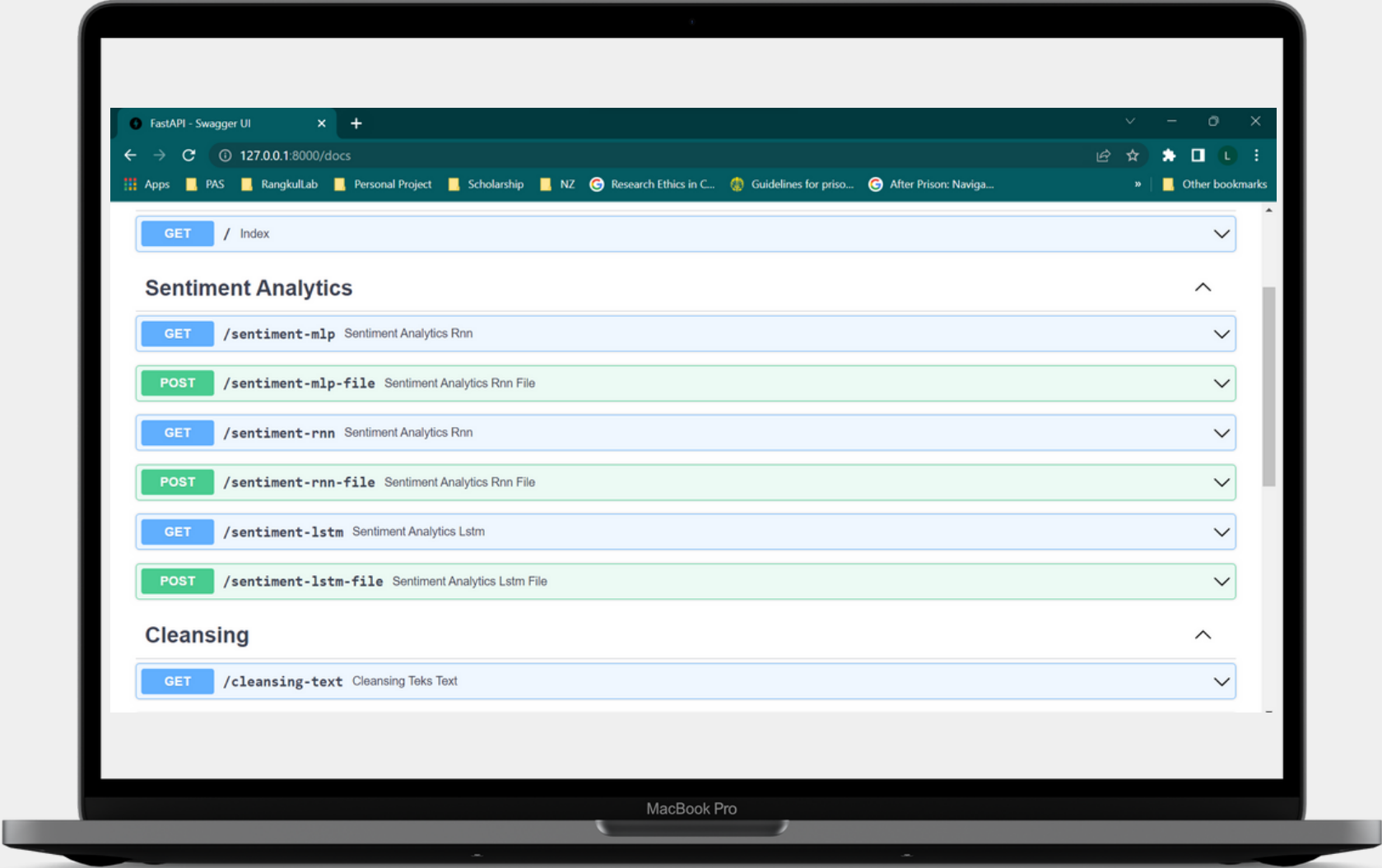
KESIMPULAN

Model LSTM memiliki tingkat akurasi yang sedikit lebih baik dibandingkan model RNN dalam menganalisis sentimen kalimat. Oleh karena itu, model LSTM lebih disarankan dalam penerapan analisis sentimen secara otomatis pada ekosistem Twitter.

REKOMENDASI

Pengembangan model deep learning dengan RNN dan LSTM dapat terus dilakukan dengan dukungan dataset training dan parameter yang lebih baik. Sehingga dapat menghasilkan model yang lebih akurat.

DEMO



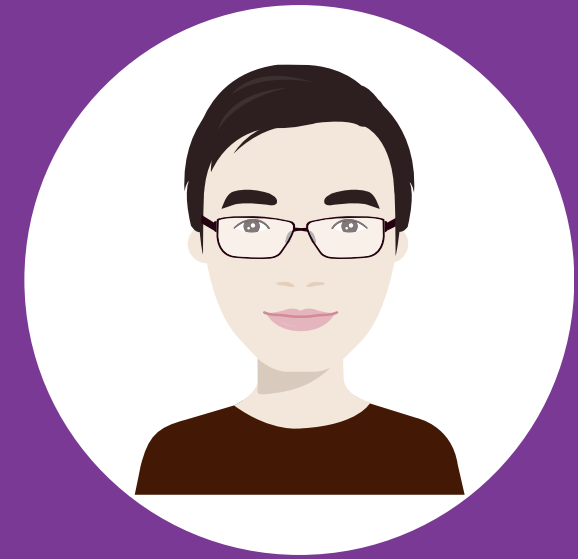
KELOMPOK 3



Yogi M.
Team Leader



Lilis Lisnawati
Team member



Fuguh Budi Utomo
Team member