UAS NLP

MEMBANDINGKAN PERFORMA RNN DAN LSTM UNTUK MEMPREDIKSI SENTIMEN DARI REVIEW MOVIE PADA DATASET IMDB

Ahmad Fauzan

NIM : 11220940000031

MATEMATIKA

Fadhilah Awal Rizqiansyah

NIM : 11220940000064

MATEMATIKA

Tujuan

01. Memprediksi sentimen ulasan film (positif atau negatif).

02. Membandingkan akurasi dan performa model RNN dan LSTM.

03. Menganalisis efektivitas pre-processing dan tokenisasi dalam NLP.

Table of Content

EDA

EDA

IMDb

Dataset

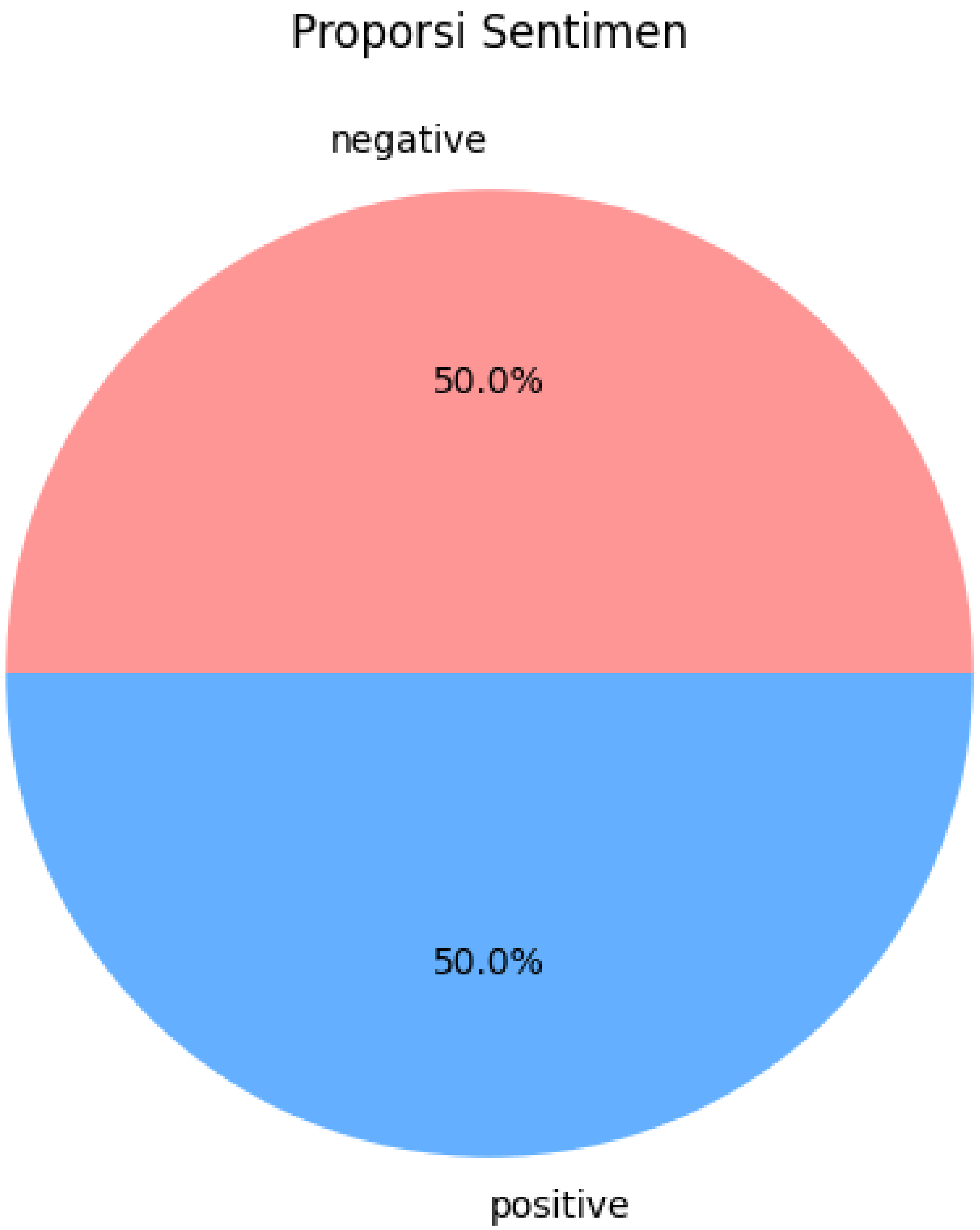
01. 50000 baris data dengan 2 column, review dan sentiment

02. Tidak ada Missing Value

03. Ada 418 duplikat data

EDA

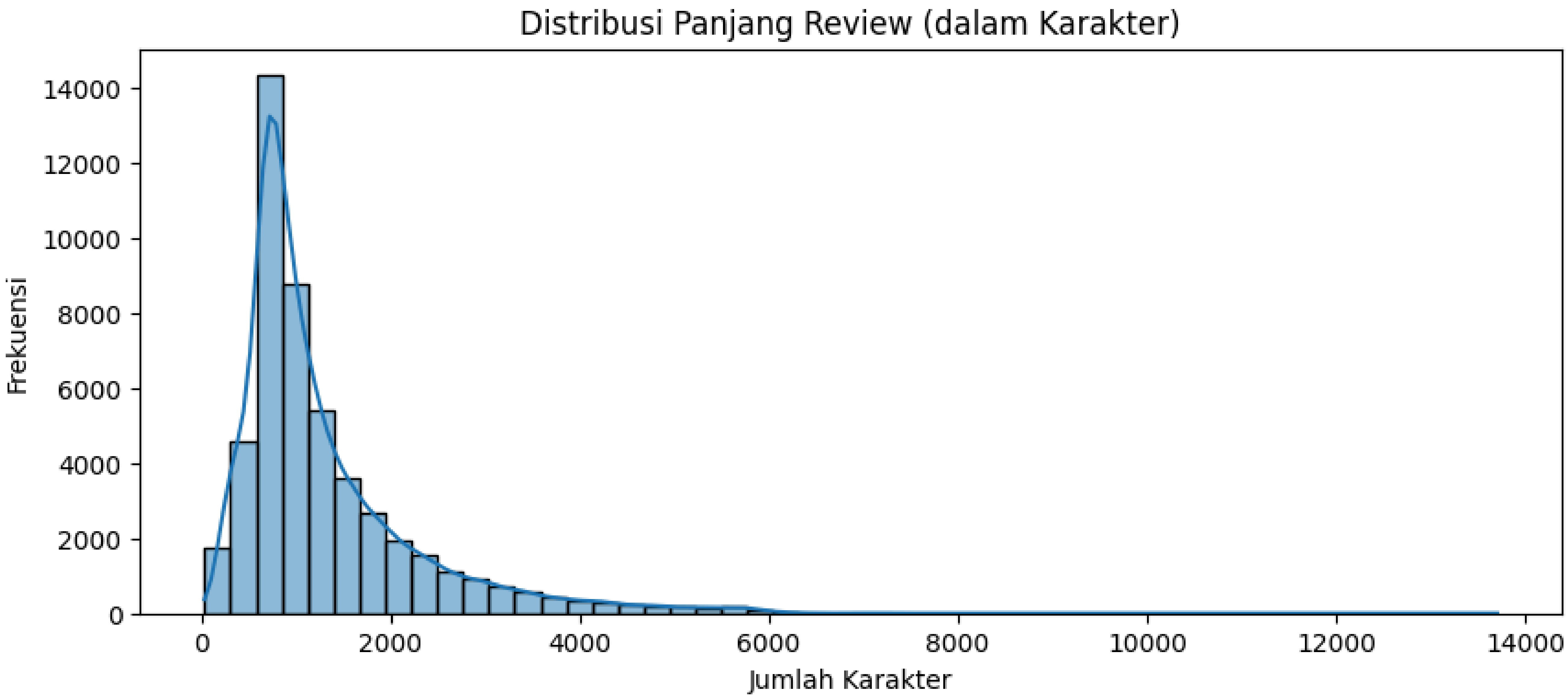
Proporsi Sentimen



Proporsi Sentimen

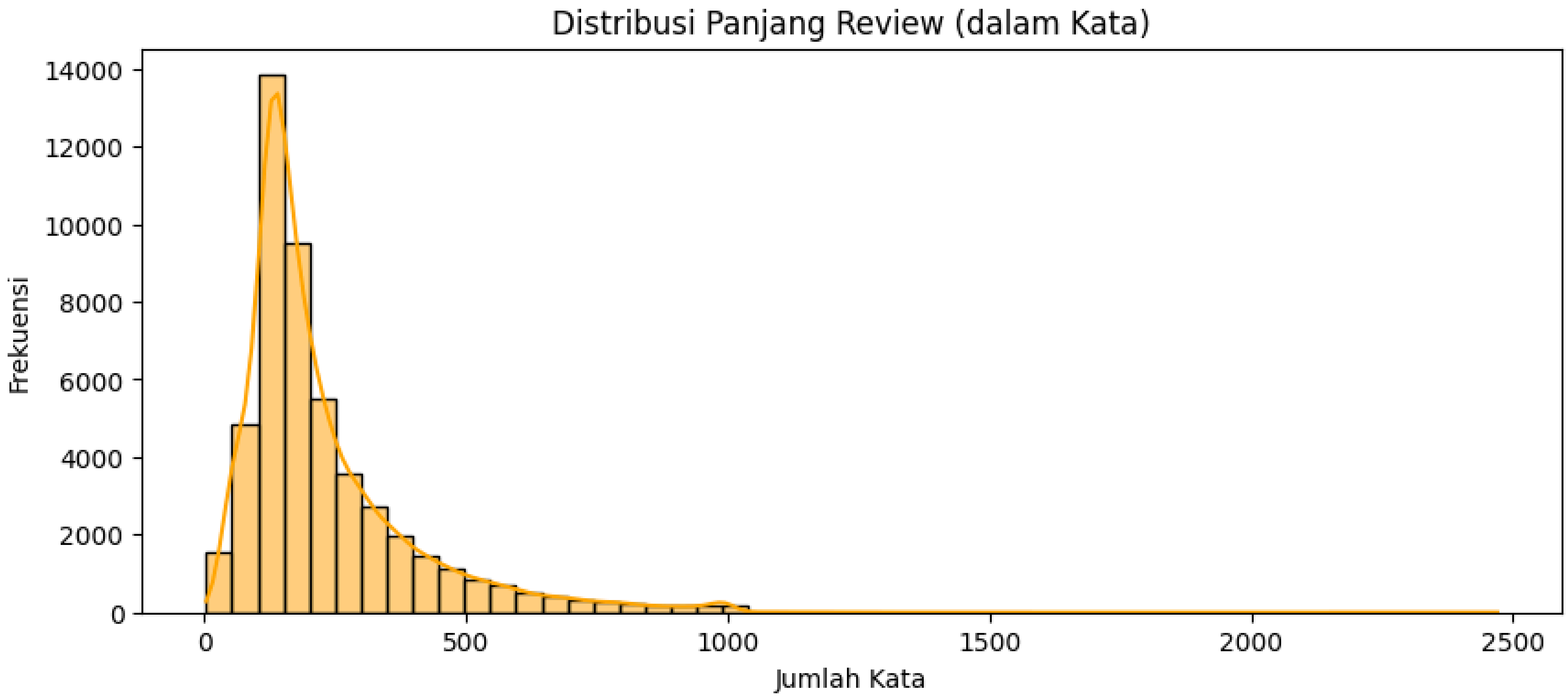
EDA

Distribusi Panjang Review (dalam Karakter)



EDA

Distribusi Panjang Review (dalam Kata)



Preprocess Data

Preprocess Data

01. Menghapus Duplikat Data

02. Mengubah Teks menjadi Huruf Kecil

03. Menghapus URL

04 Menghapus Teks HTML

05 Menghapus Space Berlebih

Split Data & Tokenisasi

Split Data & Tokenisasi

01. Data dibagi menjadi tiga yaitu : 80% Training, 10% Test, dan 10% Validation

02. Tokenizer: Top 20.000 kata
OOV token: <OOV>

03. Padding panjang: 260 kata

04 Sequence: X_train_pad, X_val_pad, X_test_pad

05 Seed ditetapkan untuk memastikan hasil konsisten setiap eksperimen.

Arsitektur Model

Arsitektur Model RNN dan LSTM

Arsitektur Model

- 📖 **Embedding Layer:** Menerjemahkan setiap kata menjadi representasi vektor numerik.
- 🧠**RNN/LSTM Layer:** Menganalisis urutan vektor untuk memahami konteks kalimat dan dependensi jangka panjang. Diproses dari dua arah untuk pemahaman yang lebih kaya.
- 👨‍⚖️ **Fully Connected Layer:** Menerima ringkasan konteks dari lapisan sebelumnya dan membuat keputusan akhir untuk klasifikasi.
- 👍/👎**Output:** Menghasilkan probabilitas sentimen (0-1) menggunakan fungsi aktivasi Sigmoid, yang kemudian diterjemahkan menjadi prediksi **Positif** atau **Negatif**.

Parameter

VOCAB_SIZE = len(tokenizer.word_index) + 1
EMBEDDING_DIM = 128
HIDDEN_DIM = 64
OUTPUT_DIM = 1
N_LAYERS = 2
BIDIRECTIONAL = True
DROPOUT = 0.5
LEARNING_RATE = 0.001
N_EPOCHS = 50
CLIP = 1.0
PATIENCE = 10

Workflow RNN

Input & Embedding

Proses: Teks ulasan mentah dibersihkan, diubah menjadi sekuens angka, lalu dipetakan menjadi urutan vektor makna oleh Embedding Layer.
Bentuk Input/Output: [Batch, Time] → [Batch, Time, Embedding Dim]

Pemrosesessan RNN

Proses: Memproses urutan kata menggunakan satu aliran memori saja, yaitu **hidden state**. Informasi dari awal sekuens rentan "hilang" atau "terlupakan" (vanishing gradient).
Bentuk: [B, T, E] → [B, H*2]

Penggabungan & Klasifikasi

Proses: Vektor ringkasan akhir dari kedua arah (bidirectional) digabungkan. Hasilnya kemudian dimasukkan ke Fully Connected (FC) Layer untuk menghasilkan prediksi akhir.
Bentuk Input/Output: [Batch, Hidden Dim * 2] → [Batch]

Workflow LSTM

Input & Embedding

Pemrosesan LSTM

Penggabungan & Klasifikasi

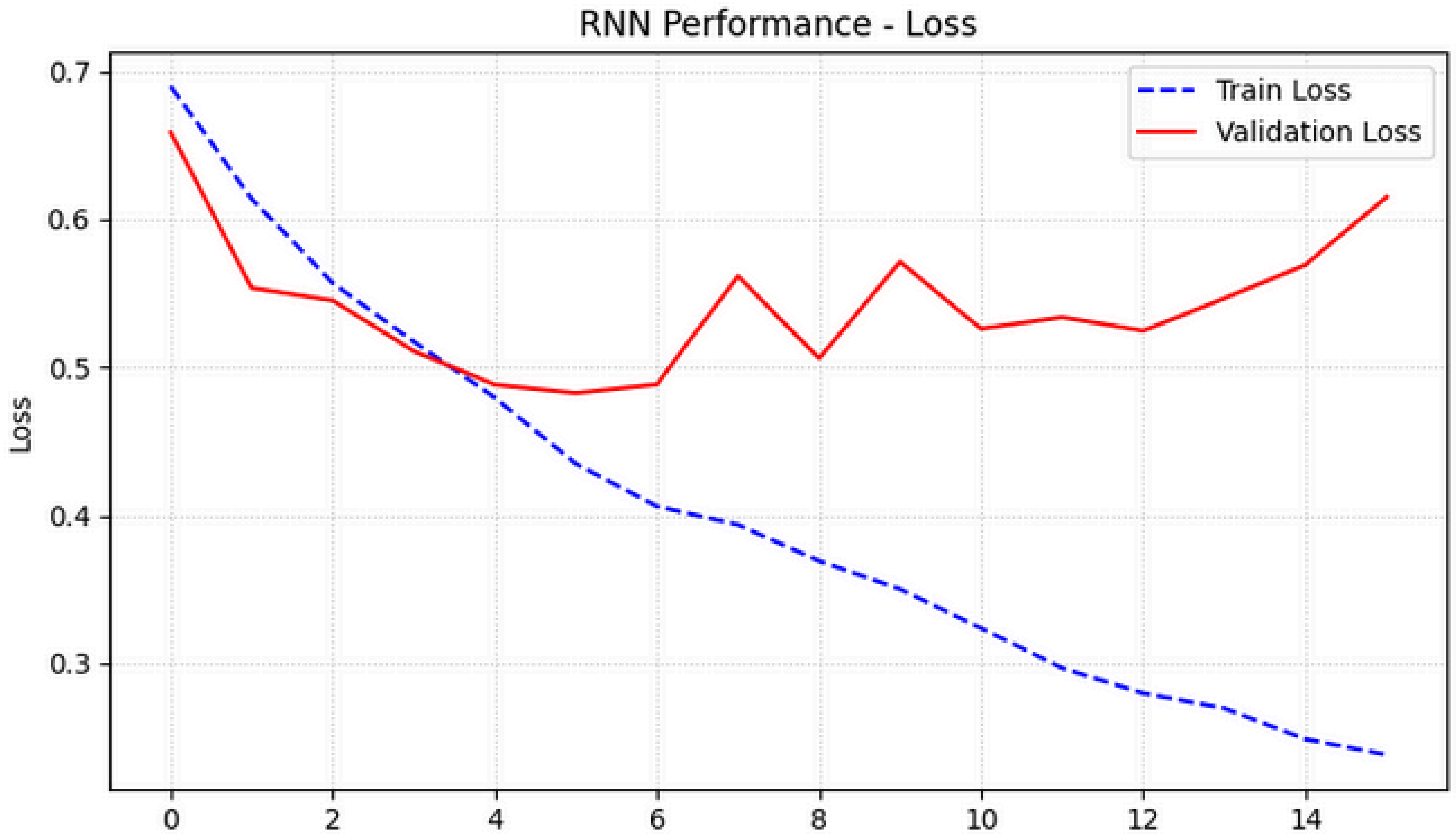
Proses: Teks ulasan mentah dibersihkan, diubah menjadi sekuens angka, lalu dipetakan menjadi urutan vektor makna oleh *Embedding Layer*.
Bentuk Input/Output: [Batch, Time] → [Batch, Time, Embedding Dim]

Proses: Memanfaatkan dua aliran memori: **hidden state** dan **cell state**. **Gerbang (gates)** secara aktif mengatur informasi apa yang harus diingat dan dilupakan, sehingga mampu menangani dependensi jangka panjang.
Bentuk: [B, T, E] → [B, H*2]

Proses: Vektor ringkasan akhir dari kedua arah (*bidirectional*) digabungkan. Hasilnya kemudian dimasukkan ke *Fully Connected (FC) Layer* untuk menghasilkan prediksi akhir.
Bentuk Input/Output: [Batch, Hidden Dim * 2] → [Batch]

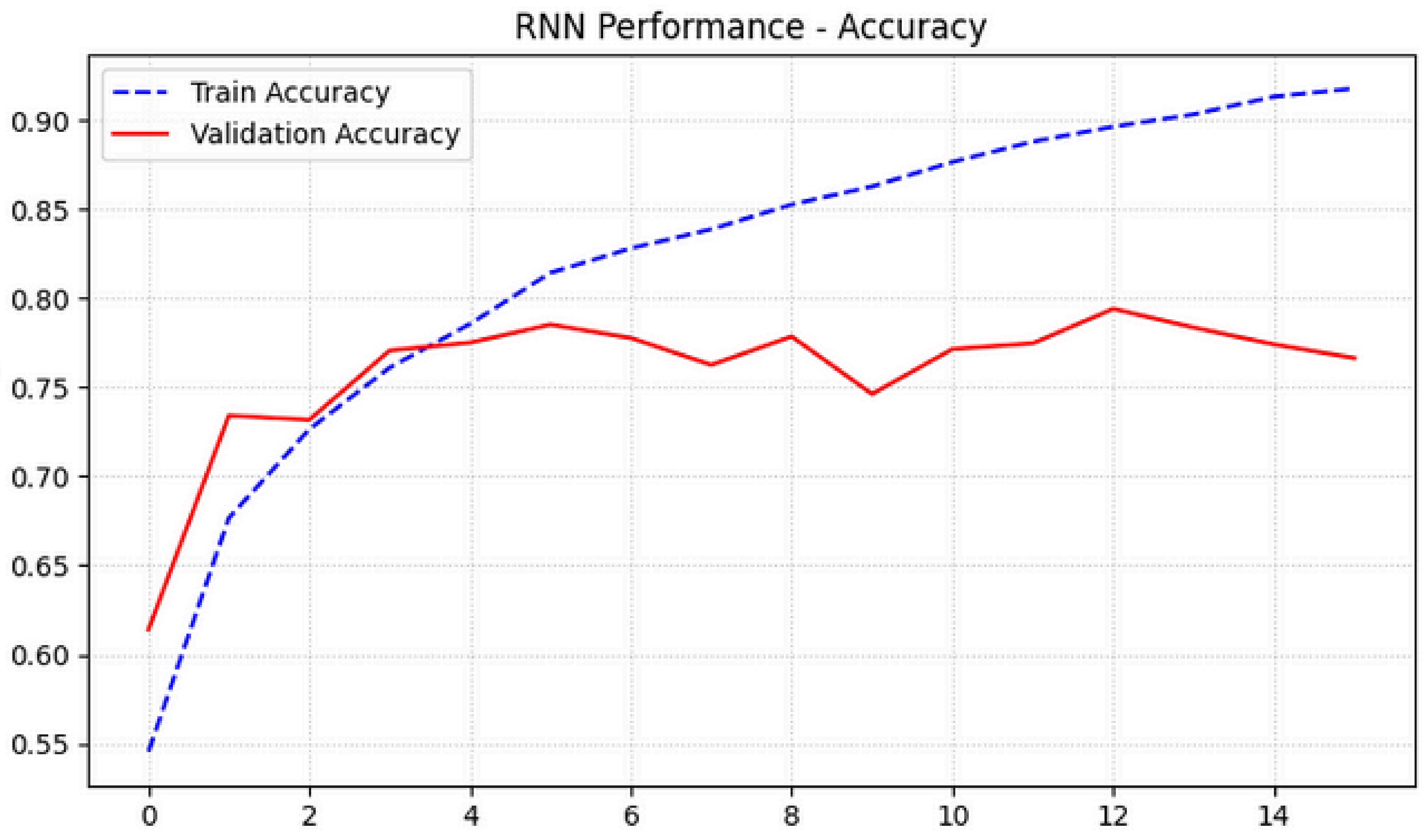
Evaluasi Model

RNN - Loss



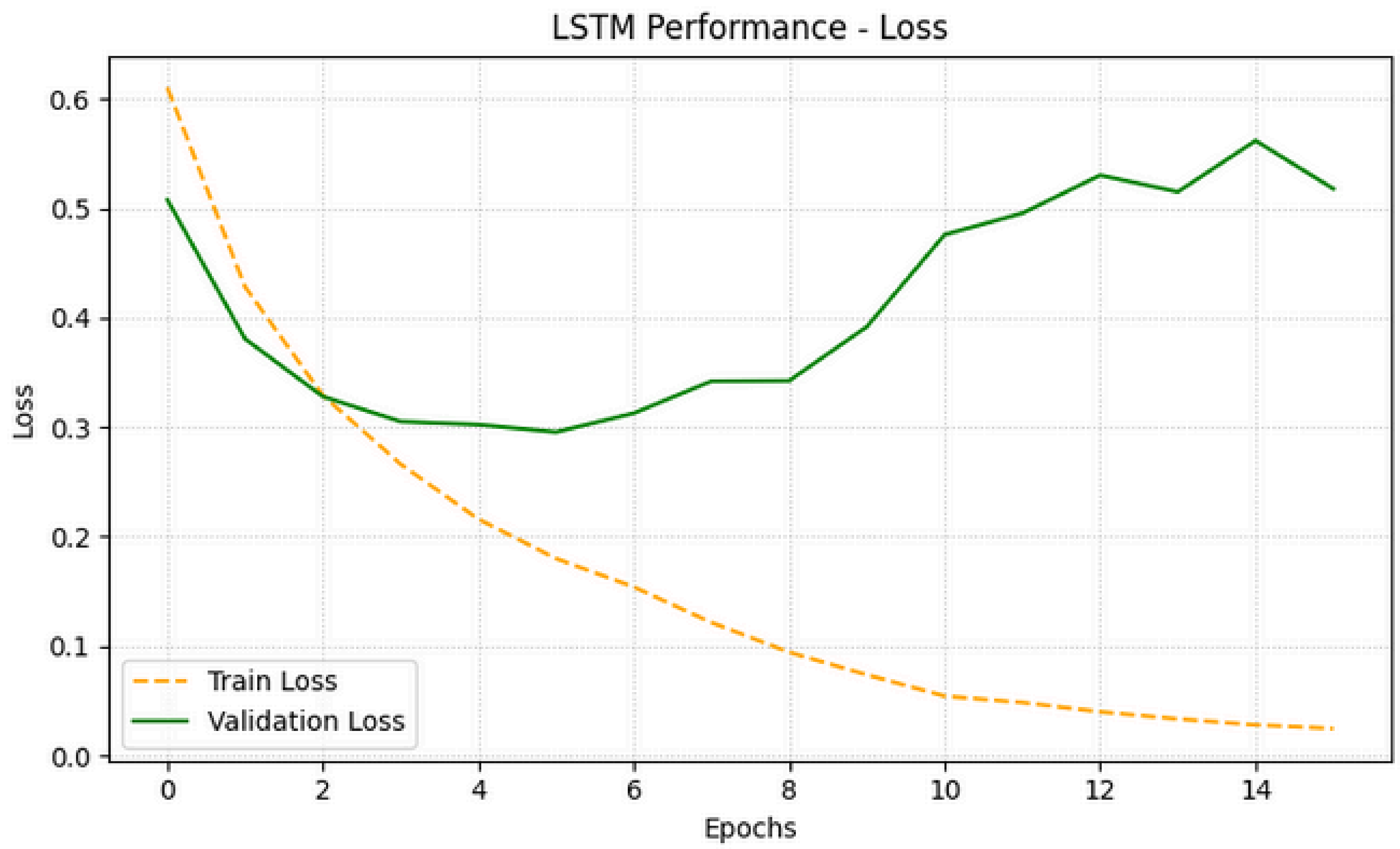
- **Train Loss:** Terus menurun selama 15 epoch, menunjukkan model belajar dari data train.
- **Validation Loss:** Cenderung stagnan dan bahkan naik setelah epoch ke-6, mengindikasikan potensi overfitting.

RNN - Accuracy



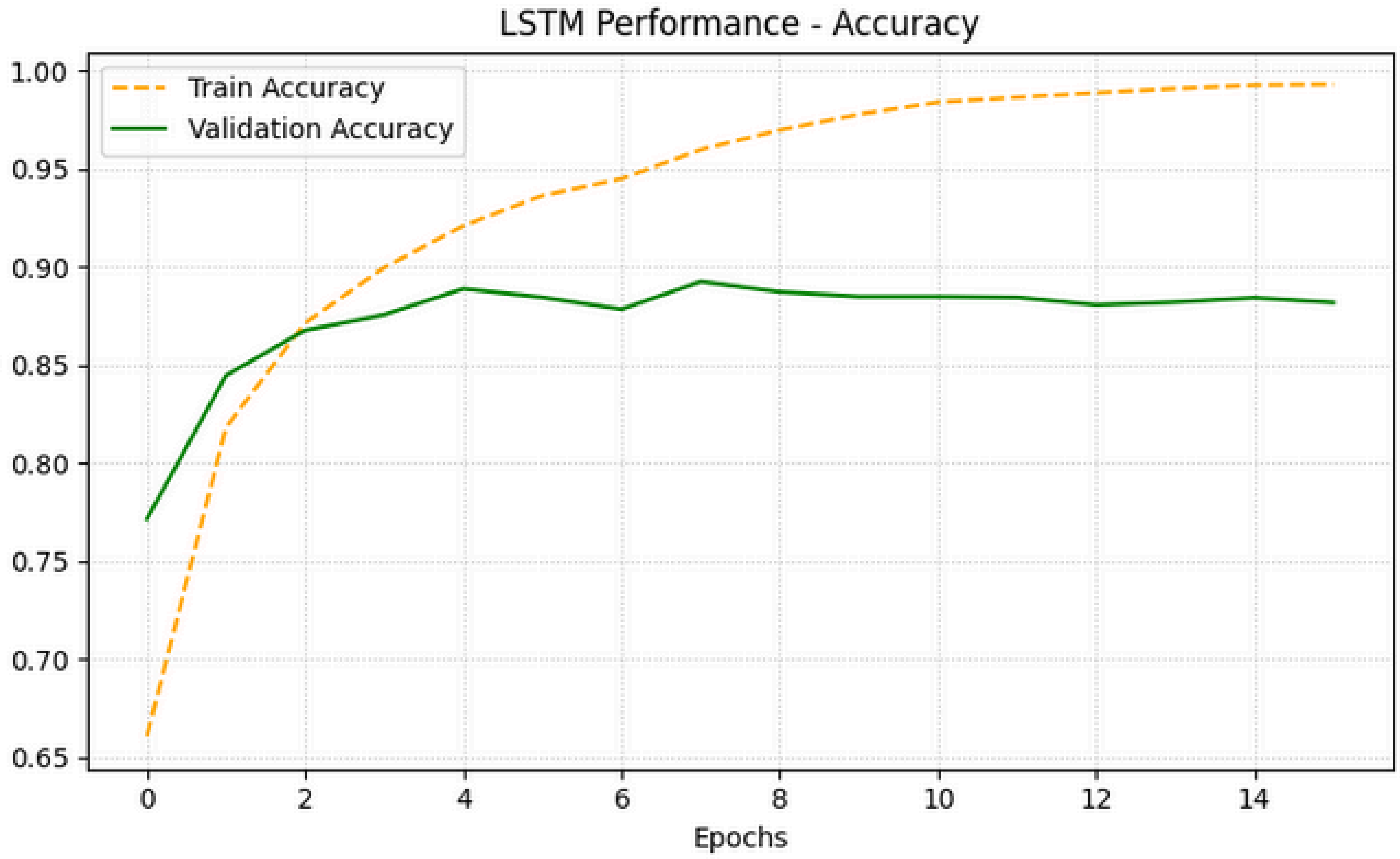
- **Train Accuracy:** Terus meningkat secara signifikan.
- **Validation Accuracy:** Tidak meningkat konsisten, berhenti sekitar 78–79%.

LSTM - Loss



- **Train Loss:** Turun tajam dan stabil hingga sangat rendah.
- **Validation Loss:** Turun sampai epoch ke-5, lalu sedikit naik tapi tetap lebih stabil dari RNN.

LSTM - Accuracy

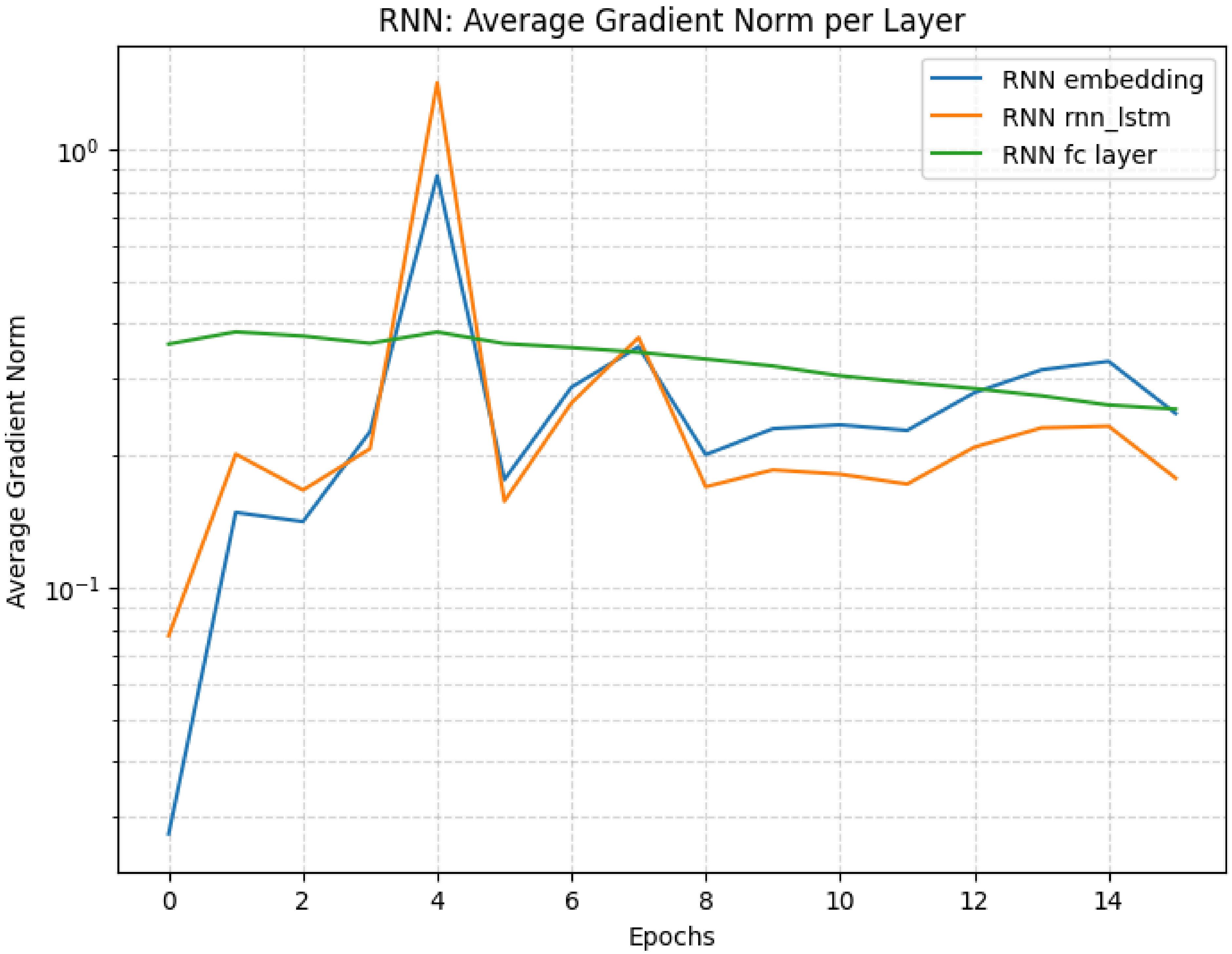


- **Train Accuracy:** Meningkat hingga hampir 99%.
- **Validation Accuracy:** Stabil hampir menyentuh 90%, menunjukkan generalisasi lebih baik daripada RNN.

Perbandingan

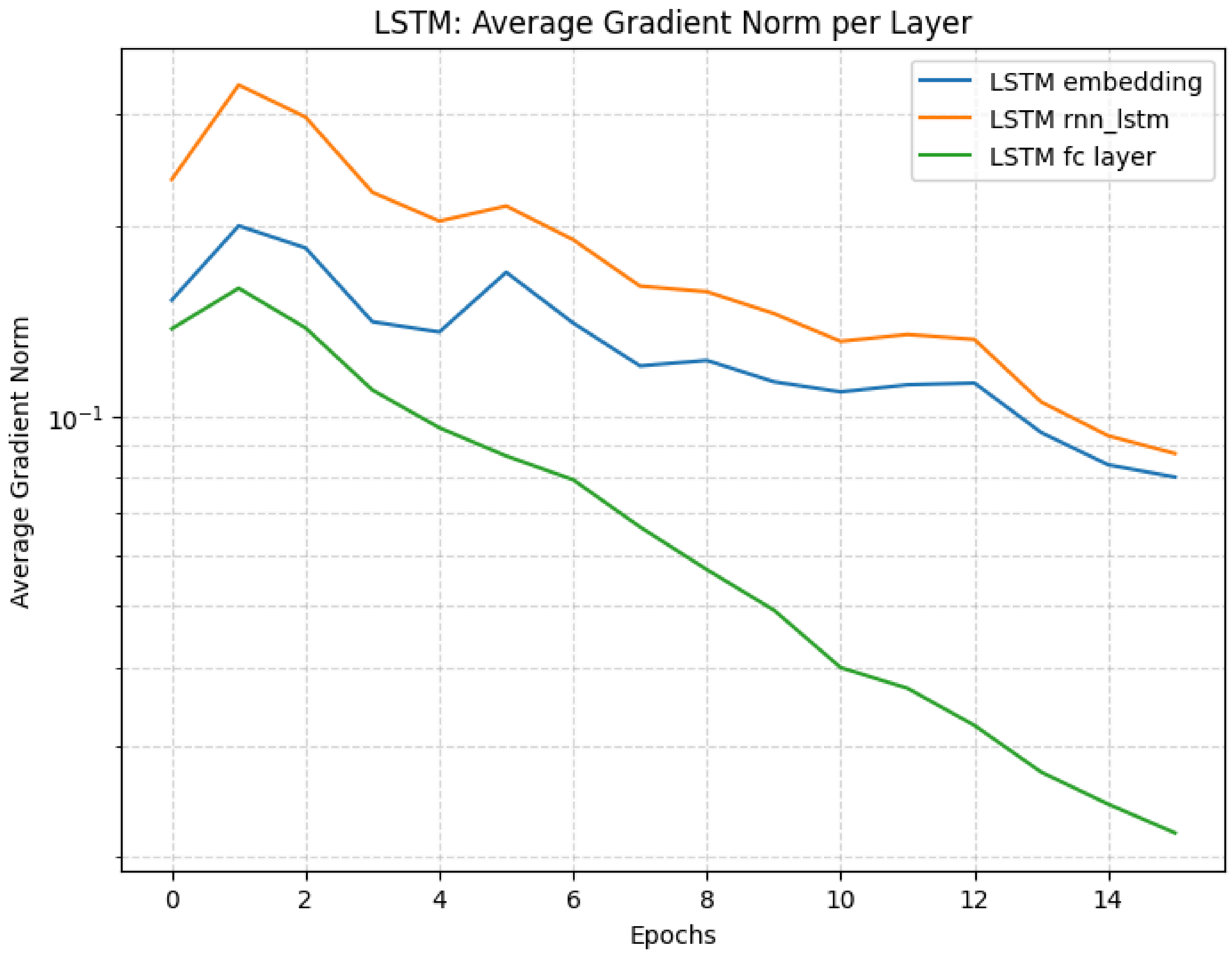
Training log	RNN	LSTM
Train Loss	Terus menurun secara stabil.	Menurun tajam dan sangat konsisten.
Val Loss	Overfitting sangat jelas setelah epoch ke-6.	Mulai overfitting setelah epoch ke-9
Train Accuracy	Terus meningkat hingga >90%.	Meningkat tajam hingga mendekati 99%.
Val Accuracy	Stagnan di sekitar 78–79%	Stabil hampir mendekati 90%.

RNN - Evaluasi Gradien Rata-rata per Layer



- **Layer Embedding & RNN:** Fluktuatif, terutama spike besar di epoch ke-4 → kemungkinan terjadi exploding gradient.
- **Layer FC (Linear):** Stabil tapi cenderung stagnan.

LSTM - Evaluasi Gradien Rata-rata per Layer



Semua Layer: Menunjukkan penurunan konsisten pada norm gradien.

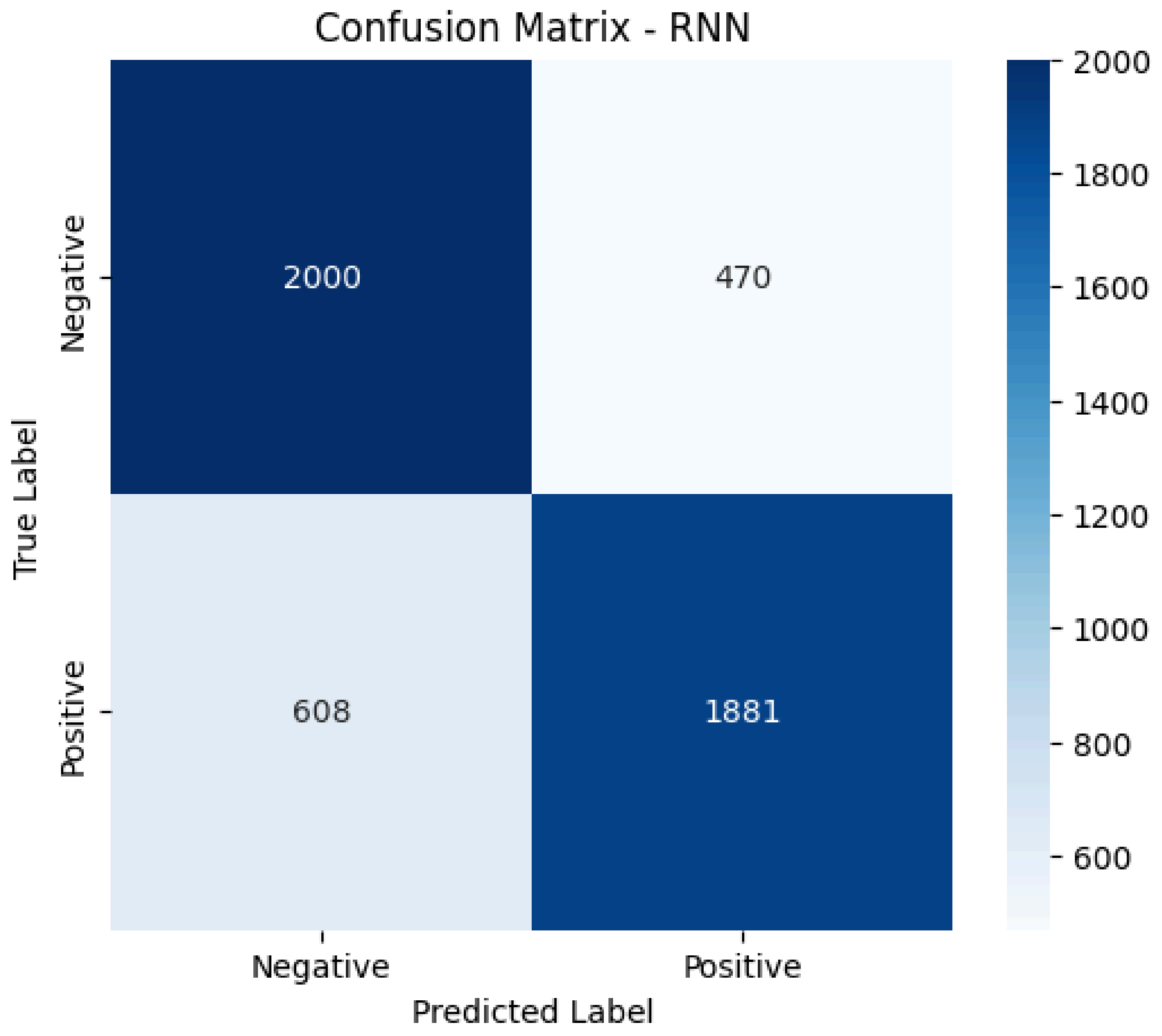
Lebih stabil dibandingkan RNN, tidak ada lonjakan tajam, menandakan LSTM lebih stabil dalam proses belajar.

Perbandingan

Layer	RNN	LSTM
Embedding Layer	Fluktuatif, cenderung tidak stabil, terlihat adanya spike besar.	Menurun secara konsisten dan stabil sepanjang epoch.
RNN/LSTM Layer	Sangat fluktuatif, ada lonjakan tajam di epoch ke-4 (exploding gradient).	Stabil, tidak ada lonjakan drastis proses belajar lebih terkontrol.
FC (Fully Connected)	Stabil namun stagnan, tidak banyak perubahan.	Turun secara perlahan, menunjukkan proses belajar tetap terjadi.

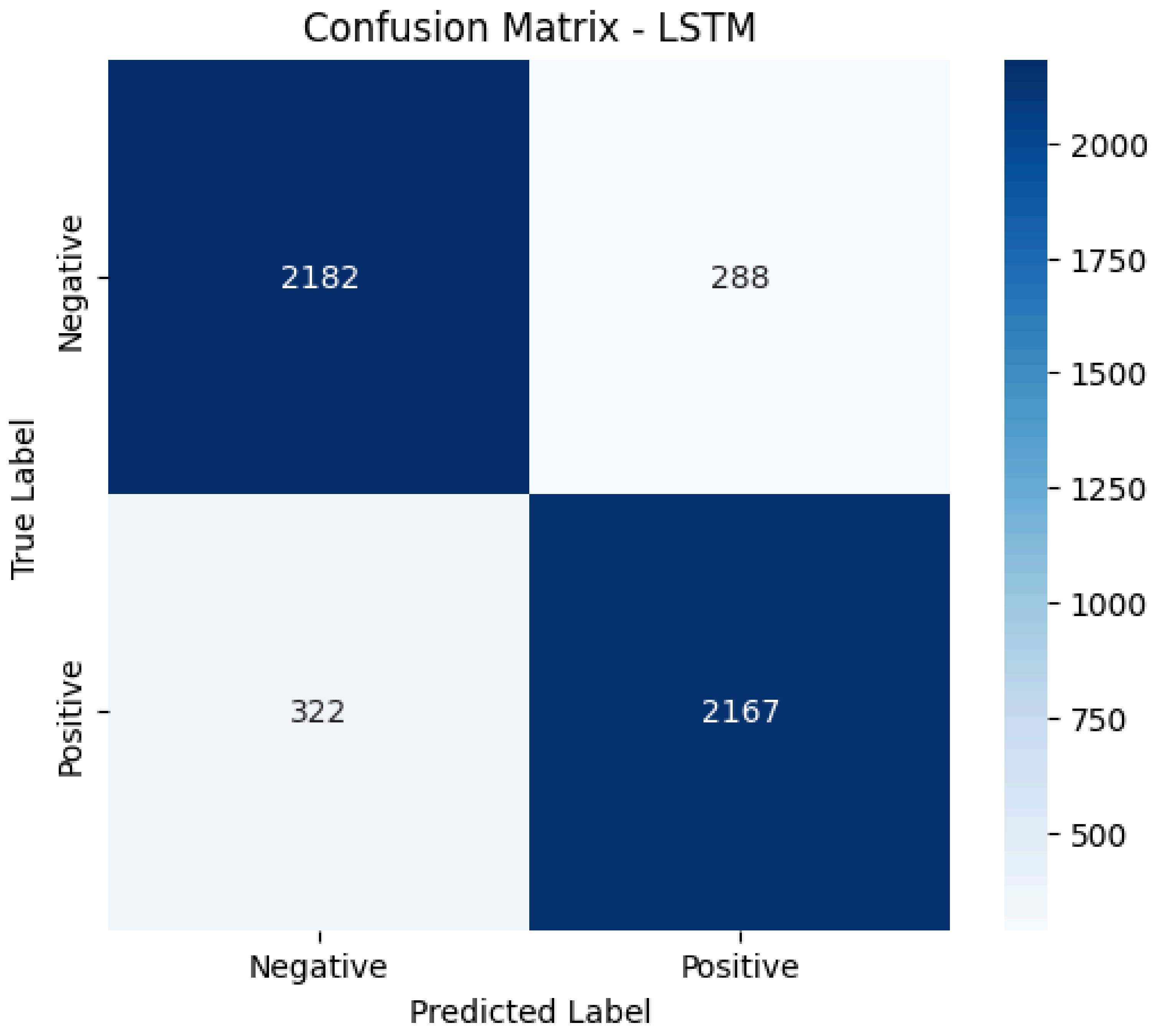
RNN

	Precision	Recall	F1-Score
Negative	0.77	0.81	0.79
Positive	0.80	0.76	0.78
Accuracy			0.78
Macro avg	0.78	0.78	0.78
Weighted avg	0.78	0.78	0.78



LSTM

	Precision	Recall	F1-Score
Negative	0.87	0.88	0.88
Positive	0.88	0.87	0.88
Accuracy			0.88
Macro avg	0.88	0.88	0.88
Weighted avg	0.88	0.88	0.88



Tabel Perbandingan

	RNN	LSTM
Accuracy	0.78	0.88
Precision (macro avg)	0.78	0.88
Recall (macro avg)	0.78	0.88
F1-Score (macro avg)	0.78	0.88

Ulasan Panjang Prediksi RNN dan LSTM Berbeda

index	clean_review	label_sebenarnya	prediksi_rnn	prediksi_lstm	jumlah_kata
6	when you put this dvd into your player and hit "play," you will experience a brief moment of silence and see a black screen as the laser is guided to the correct starting point in the center of the disc. cherish this moment. make sure you have some tylenol or something (preferably pm's so you can fall asleep), because you're going to have a massive headache once this movie starts.starring a bunch of big-breasted girls and with an opening that actually made me chuckle a bit, i thought i'd be in for a good time. sure, the opening sequence was a wee bit awkward and most of the jokes fell flat and it seemed like this was going to be a scream ripoff (by the way, my sole chuckle was from julie strain's final comment in this scene). but then i knew there was trouble... the opening sequence had a terrible rock song. during this terrible rock song, i looked over the dvd chapter titles and saw things that said "topless in the backyard!" and "better than sex!". i knew what the selling point of this movie was going to be.and that's the sad truth: the only good thing about this movie is the attractive cast. other than that, it's a sadly routine slasher film that throws in an "innovative" concept about murder clubs, which ends up being fake anyway. so, the whole movie then points in another direction to try to be confusing and this huge mystery, but it all just adds up to not being interesting at all and leaves you feeling like you don't care for any of the characters. i mean, when the main character of the movie is revealed to have murdered an innocent woman, can you really feel any sympathy towards her when she's in fear for her life? the scream influence is prevalent throughout, with a ghost face killer and some really terrible jokes. we're also treated to scenes of the main character talking to her mom and dad (lloyd kaufman! the only other cool part of the movie!) about an abortion or something. uh. yeah... this isn't a "so-bad-it's-good" movie, it's just bad.someone compared this to a troma film, but... you know, most any film that comes from full moon (or its offshoot, as this film proves) is horrible. not horrible in a troma sense -- i've seen many troma films, and i can honestly say they all offer something, anything that you can walk away with and tell your friends about later. however, this film has pretty much nothing at all enjoyable about it. beware.	Negative	Positive	Negative	434
126	...as valuable as king tut's tomb! (ok, maybe not that valuable, but worth hunting down if you can). i notice no one has commented on this movie for some years, and i hope a fresh post will spark some new comments. this is a film that i remembered only snippets of from childhood, and only saw recently when i tired of waiting for fox to honour its own past, and hunted down the korean dvd (in english, but with unremovable korean subtitles). i won't go through another long plot description - suffice to say that seeing it for the first time in its proper widescreen format left me agape at the vistas and the scope of the film. the matte paintings still hold up, and the palace sets are truly breathtaking. but it is the smaller scale details that lend this film its depth and richness, offering a glimpse into the lifestyles of egypt's poor as well as its elite. the bazaars, hovels, docks, embalming houses, and taverns are as fascinating as pharaoh's throne room. while errors abound on the large scale (most notably the dynastic succession), the details are more meticulously researched than the vast majority of hollywood's films. visually, it's not without its flaws - the interiors are often too overly lit and colourful to blend seamlessly with the exteriors. nevertheless, this is a movie that should be credited for being as audacious in the small as it is in the large. tedious? in parts, absolutely. overacted? underacted? yes, both - though 'understated' might be a more apt description. too long? absolutely not. i wished they had spent more time with sinuhe's experiences in the house of death, and among the hittites, and less with his 'romance' with nefer, though. historically inaccurate? yes, that too, but so was shakespeare. nobody chastises him for it. i appreciate historical accuracy as much as the next guy, but ultimately it has to be remembered that cinema is theater, not a history lesson.	Positive	Negative	Positive	331

Perbandingan Frekuensi Panjang Ulasan yang Salah Diprediksi

RNN Lebih Rentan Gagal di Ulasan Panjang:

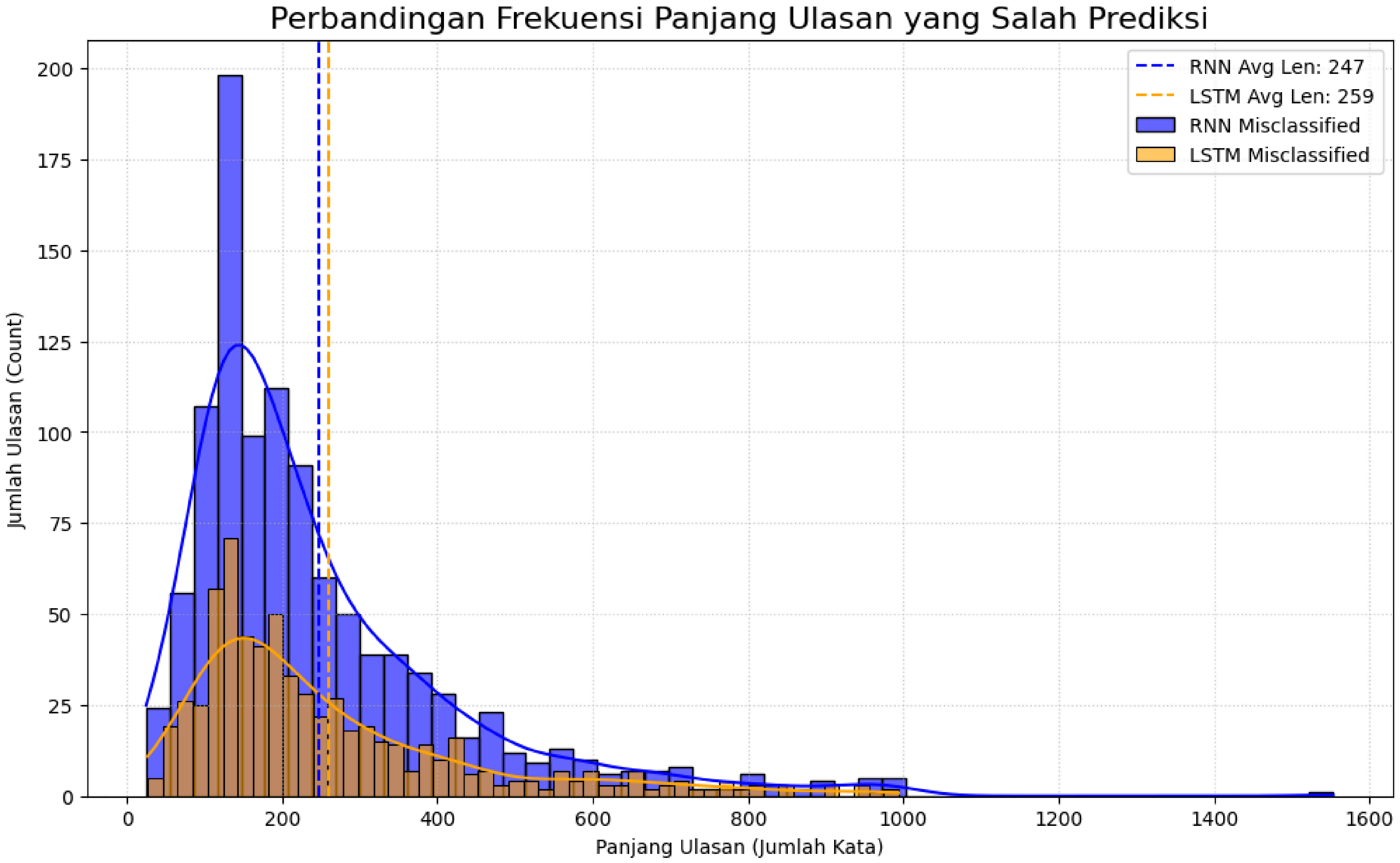
- Ulasan panjang (<400 kata) cenderung lebih sering salah diklasifikasi oleh RNN dibandingkan LSTM.
- Hal ini bisa dikaitkan dengan vanishing gradient dan ketidakmampuan RNN menangkap konteks panjang.

LSTM Lebih Stabil di Rentang Panjang:

- LSTM menunjukkan distribusi kesalahan yang lebih stabil dan tidak terlalu ekstrem di ulasan panjang.
- Ini menunjukkan kemampuannya mengingat konteks jangka panjang.

Rata-Rata Panjang Teks Salah Klasifikasi:

- 📉 RNN Avg Len: 247 kata
- 📈 LSTM Avg Len: 259 kata



Menunjukkan LSTM tetap mampu menangani ulasan sedikit lebih panjang sebelum melakukan kesalahan.

Distribusi Panjang Ulasan (LSTM Benar, RNN Salah)

LSTM Lebih Unggul di Rentang Panjang Sedang:

- Puncak distribusi berada di kisaran 150-300 kata.
- Artinya, LSTM dapat menangkap konteks pada panjang teks yang cukup kompleks, sedangkan RNN mulai kehilangan akurasi.

LSTM Lebih Konsisten pada Ulasan Panjang:

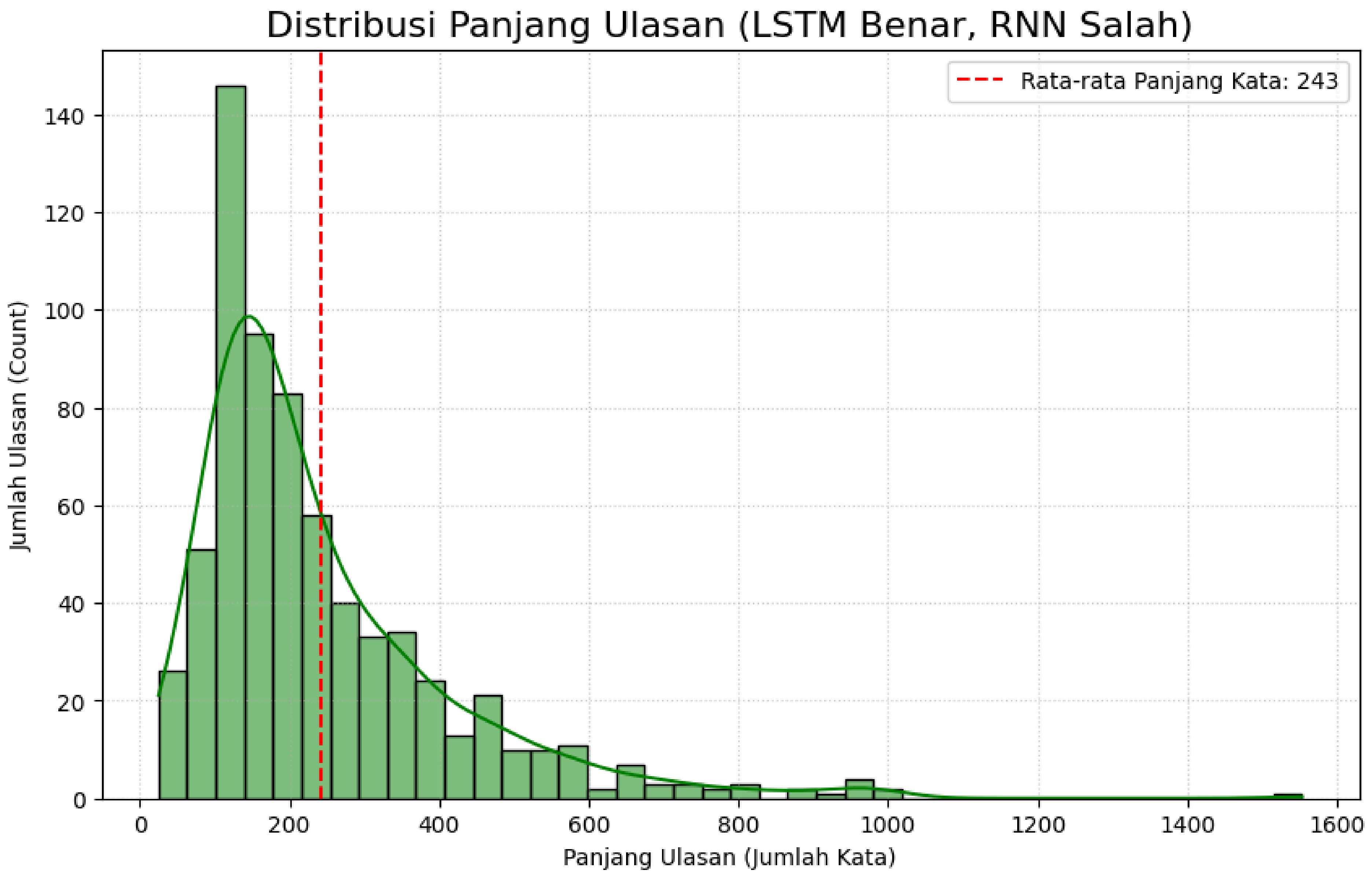
- LSTM tetap bisa menghasilkan prediksi benar hingga panjang >1000 kata, walau jumlah kasusnya sedikit.
- Hal ini menunjukkan keunggulan LSTM dalam menangani dependensi jangka panjang.

Bukti Keterbatasan RNN:

- RNN gagal pada sejumlah besar ulasan yang tidak terlalu pendek, menandakan kelemahan dalam mempertahankan informasi di sekuens panjang (vanishing gradient).

Distribusi Menyebar:

- Penyebaran panjang ulasan cukup bervariasi, menunjukkan bahwa LSTM mampu menangani berbagai panjang teks yang sulit ditangani oleh RNN.



Kesimpulan

Arsitektur LSTM Lebih Ulasan dari RNN:

- LSTM memiliki mekanisme gate yang menjaga informasi jangka panjang, sementara RNN lebih terbatas karena hanya menggunakan hidden state.
- LSTM lebih tahan terhadap vanishing gradient.

Distribusi Gradien Lebih Stabil pada LSTM:

- Gradien LSTM lebih terkontrol tanpa lonjakan ekstrem.
- RNN menunjukkan potensi exploding gradient di epoch awal.

Analisis Panjang Teks:

- Sebagian besar kesalahan terjadi pada ulasan pendek (<200 kata).
- LSTM juga terbukti lebih stabil pada ulasan yang panjang, menangani konteks lebih kompleks dengan baik

Terima Kasih