



MATH - FST - UIN JAKARTA

kelompok 4

Dosen Pengampu: Muhaza Liebelinto, M.Si.

UAS NLP 2025

PERBANDINGAN RNN DAN LSTM PADA KLASIFIKASI SENTIMEN IMDB

11220940000020

M. Abdul Ghofur

11220940000028

Alif Alamsyah

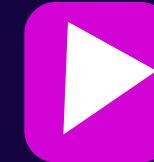




PROJECT BACKGROUND



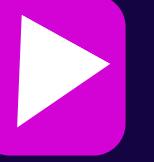
- Sentimen dalam review film penting bagi pengguna & industri.



- Review seringkali panjang → memerlukan model yang bisa memahami konteks jangka panjang.



- RNN kesulitan dengan long-term dependencies (vanishing gradient).



- LSTM dirancang untuk mengatasi masalah tersebut.





PROJECT OBJECTIVES



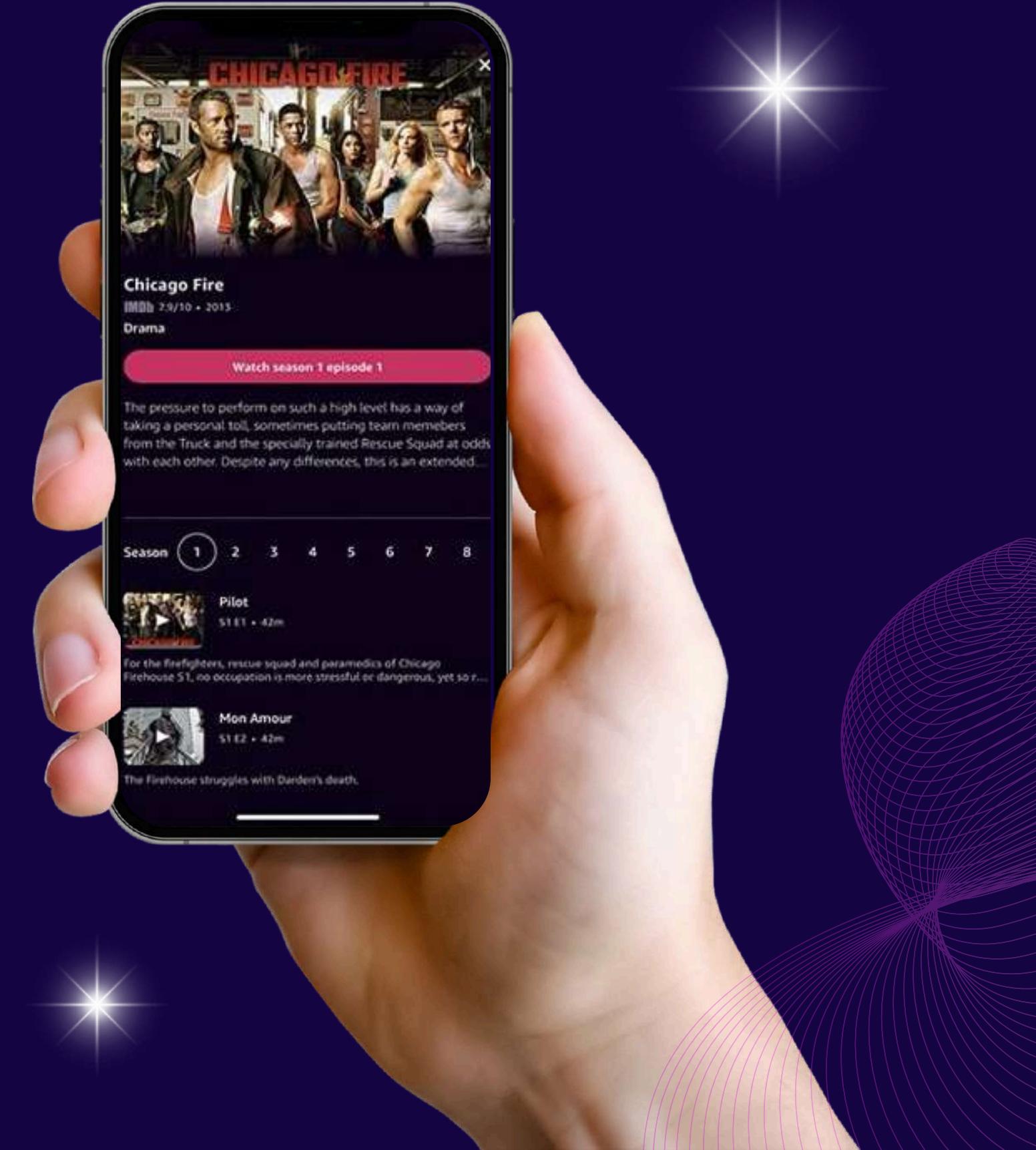
- Membandingkan performa RNN vs LSTM dalam klasifikasi sentimen IMDB.



- Mengevaluasi kemampuan model dalam menangani review panjang.



- Menganalisis stabilitas gradient selama training.





IMDB MOVIE REVIEWS

50.000 review, balance positif-negatif

IMDb

Negative (0)
50%

Positive (1)
50%

DATASET IMDB

Dataset ini terdapat review yang panjang dan kompleks (terdapat kata yang tidak baku dan simbol)

Review

Sentiment

This is a good film. This is very funny. Yet after this film there were no good Ernest films!

positive

Hated it with all my being. Worst movie ever. Mentally- scarred. Help me. It was that bad.TRUST ME!!!

negative



PROJECT FLOW

- ✓ **EDA & Preprocessing**
- ✓ **Data Split -> Tokenisasi -> Padding**
- ✓ **Model Building & Evaluation**
- ✓ **Analisis**
- ✓ **Kesimpulan**



MATH - FST - UIN JAKARTA

EXPLORATORY DATA ANALYSIS

POSITIVE

Word Cloud (Top 20) - Sentimen Positif



Word Cloud (Top 20) - Sentimen Negatif



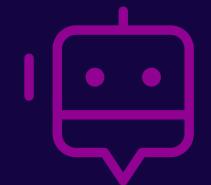
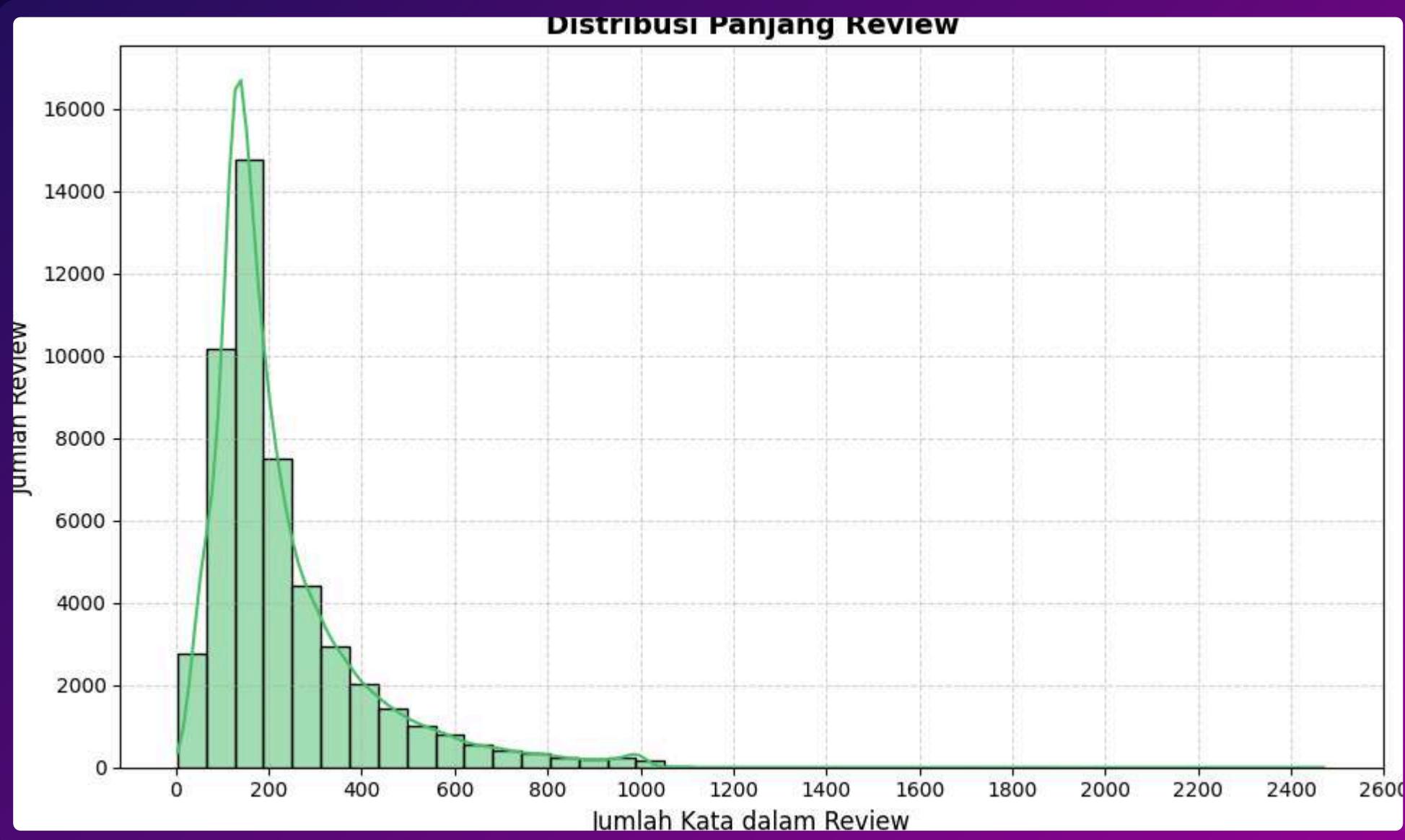
WordCloud

perbedaan kosakata antar sentimen



EXPLORATORY DATA ANALYSIS

Tentang Data



Distribusi Panjang Review berdasarkan kata



DATA PREPROCESSING



Hapus Duplikasi Data

Mencegah Data Menghafal

50000 Data

49582 Data

Cleaning Text using taudataNlpTm



Remove URL



Remove symbols



Unescape HTML



Filter karakter



Lower Case



Stopword removal



Replace newline



Lemmatisasi (Spacy)



DATA PREPROCESSING



Encoding Sentiment

Mengubah sentimen label menjadi biner

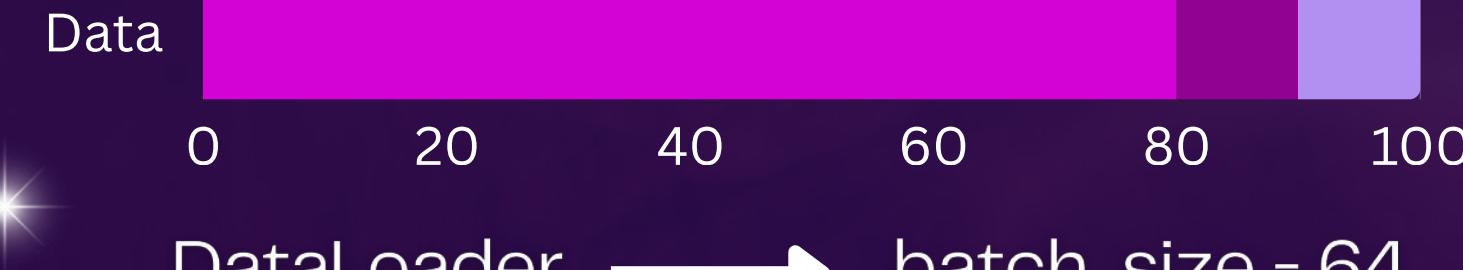


Tokenize Text

- Max vocab = 10000
- padding = 200
- Data teks diubah menjadi tensor numerik



Data Split





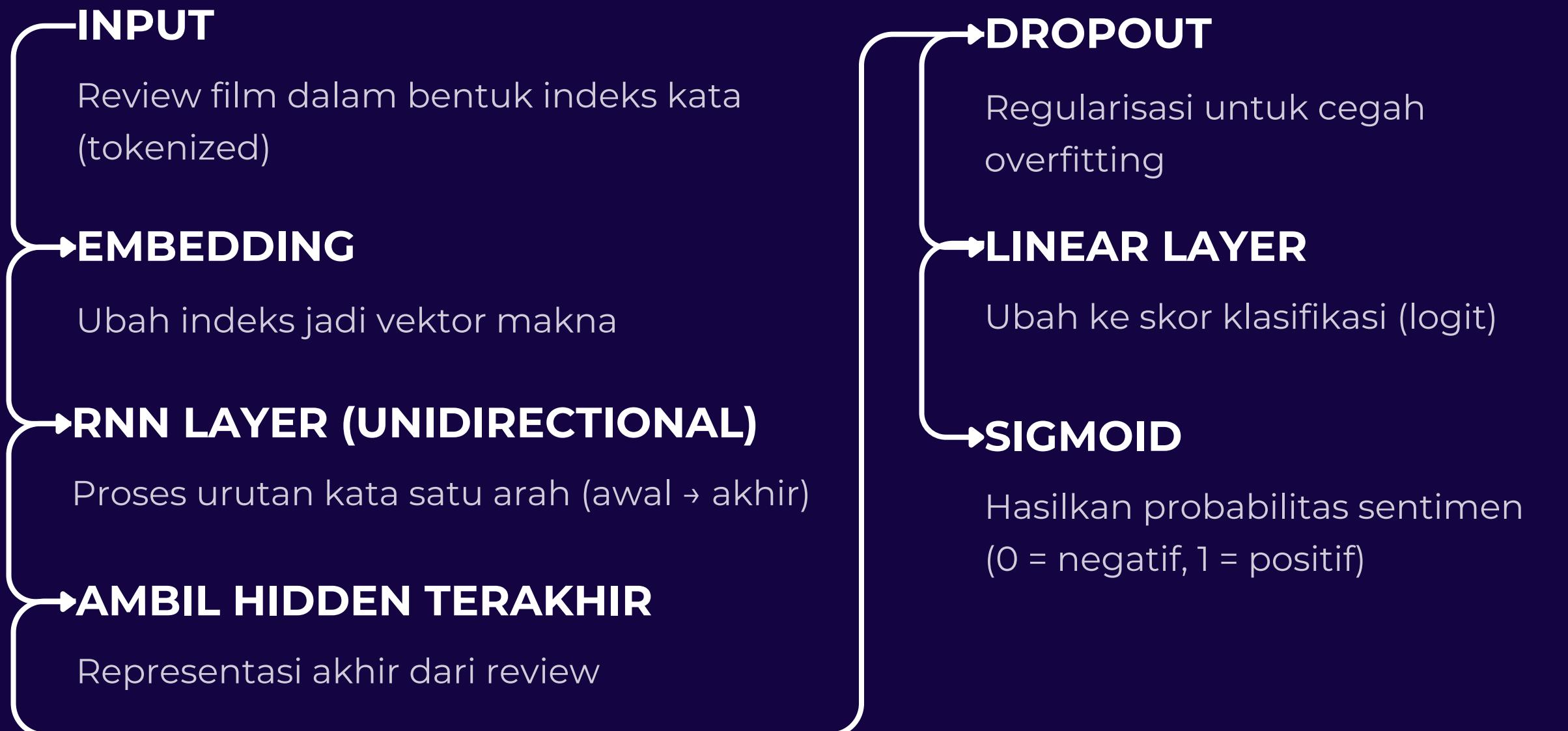
MATH - FST - UIN JAKARTA

kelompok 4

MODEL BUILDING



WORKFLOW RNN



**VOCAB_SIZE = 10000,
EMBEDDING_DIM = 100,
HIDDEN_DIM = 256,
OUTPUT_DIM = 1,
N_LAYERS = 2,
DROPOUT = 0.5**



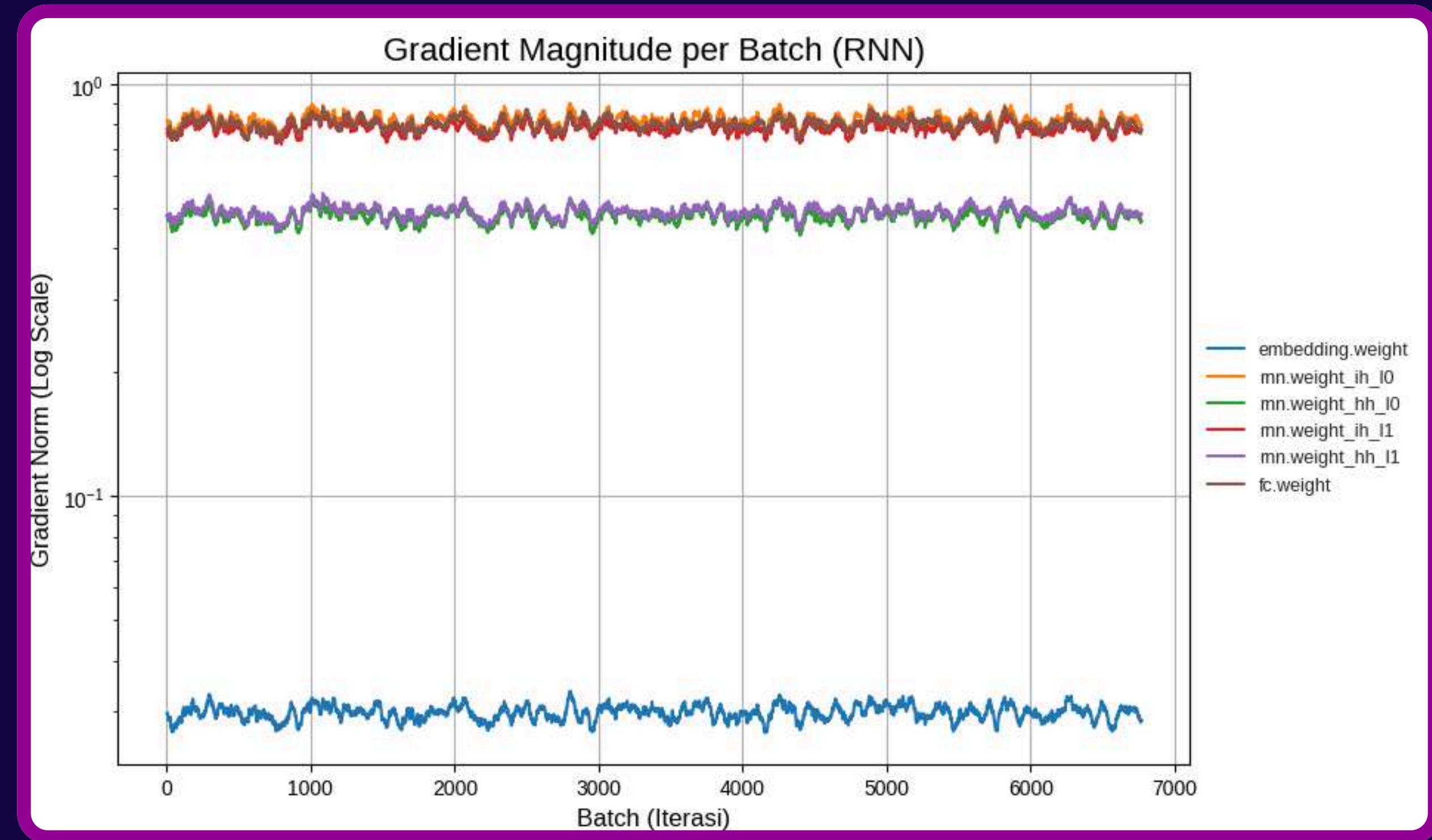
WORKFLOW LSTM



**VOCAB_SIZE = 10000,
EMBEDDING_DIM = 100,
HIDDEN_DIM = 256,
OUTPUT_DIM = 1,
N_LAYERS = 2,
DROPOUT = 0.5**

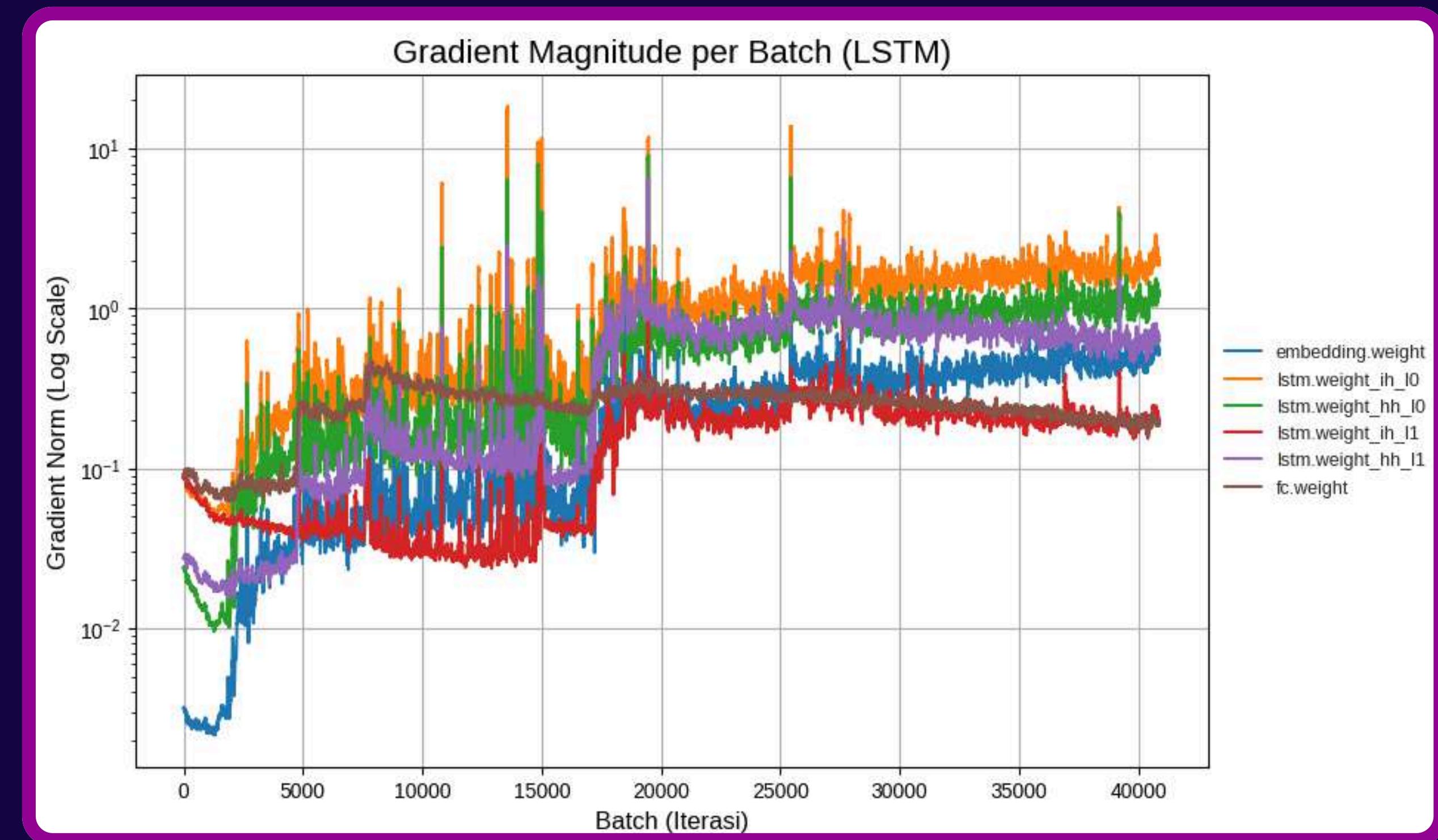


RNN'S GRADIENT





LSTM'S GRADIENT





GRADIENT ANALYSIS

RNN

Model ini mengalami vanishing gradient yang parah, terutama pada lapisan awal seperti embedding weight, di mana magnitudo gradien sangat rendah (kurang dari 10^{-1}), sementara hanya lapisan output (fc.weight) yang mendapat sinyal pembelajaran yang cukup kuat dengan magnitudo lebih tinggi dan stabil. Selain itu jelas bahwa tiap iterasi yang dilakukan bobot nya cenderung stagnan/tidak diperbarui.

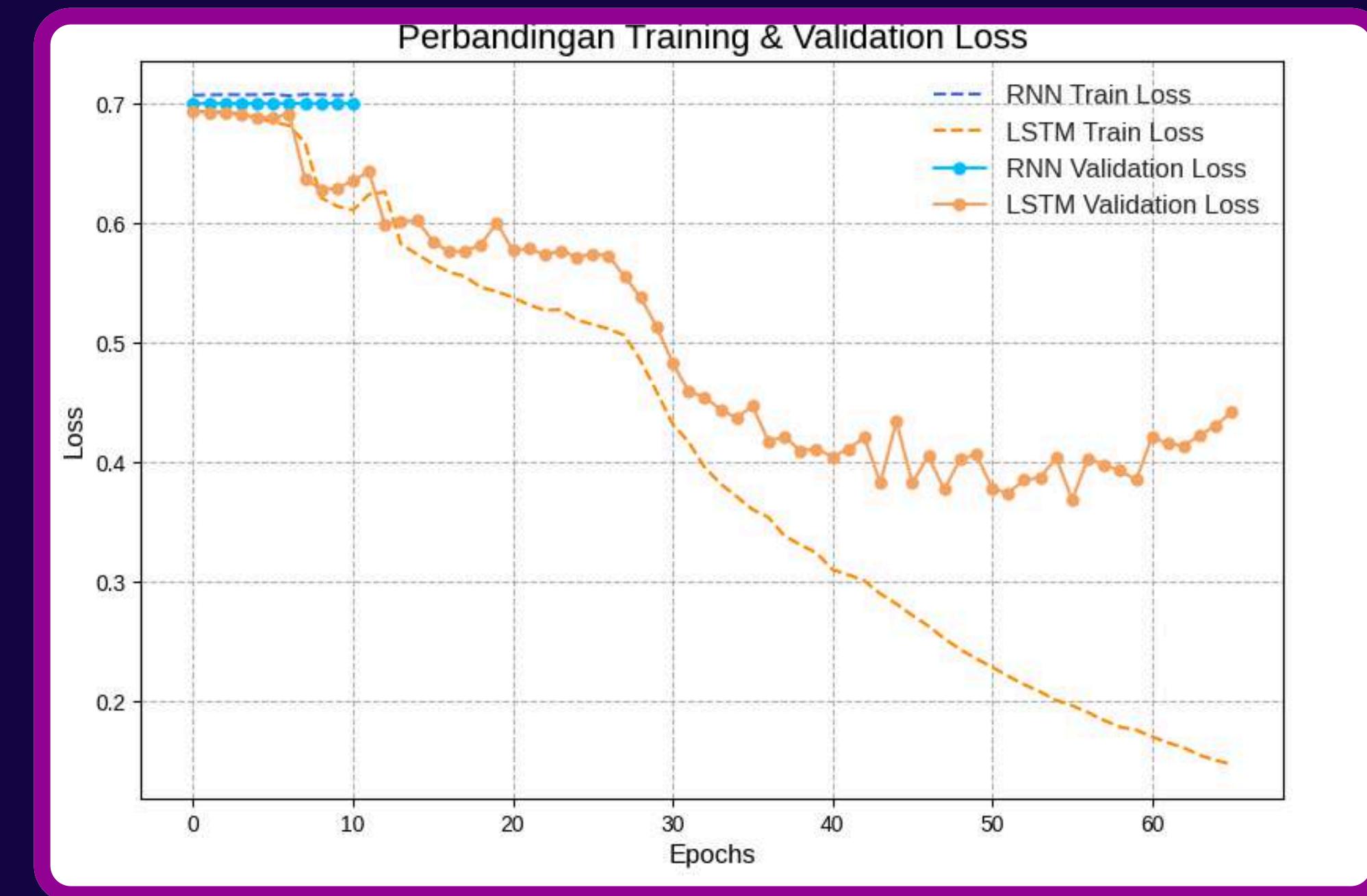
LSTM

Model ini menunjukkan kecenderungan stabil dalam penyaluran gradien seiring waktu. Meskipun terdapat indikasi vanishing gradient ringan di awal pelatihan, mekanisme gating LSTM membantu menjaga aliran gradien tetap efektif, khususnya pada tahap-tahap selanjutnya. Hal ini membuat model cocok untuk menangani long-term dependencies.



RNN VS LSTM PERFORMANCE

LOSS





LOSS

RNN

- Stabil di kisaran 0.706, tidak membaik signifikan selama training.

TRAIN

LSTM

- Menurun konsisten dari 0.693 ke 0.1468 hingga early stopping di epoch ke-66.

TEST

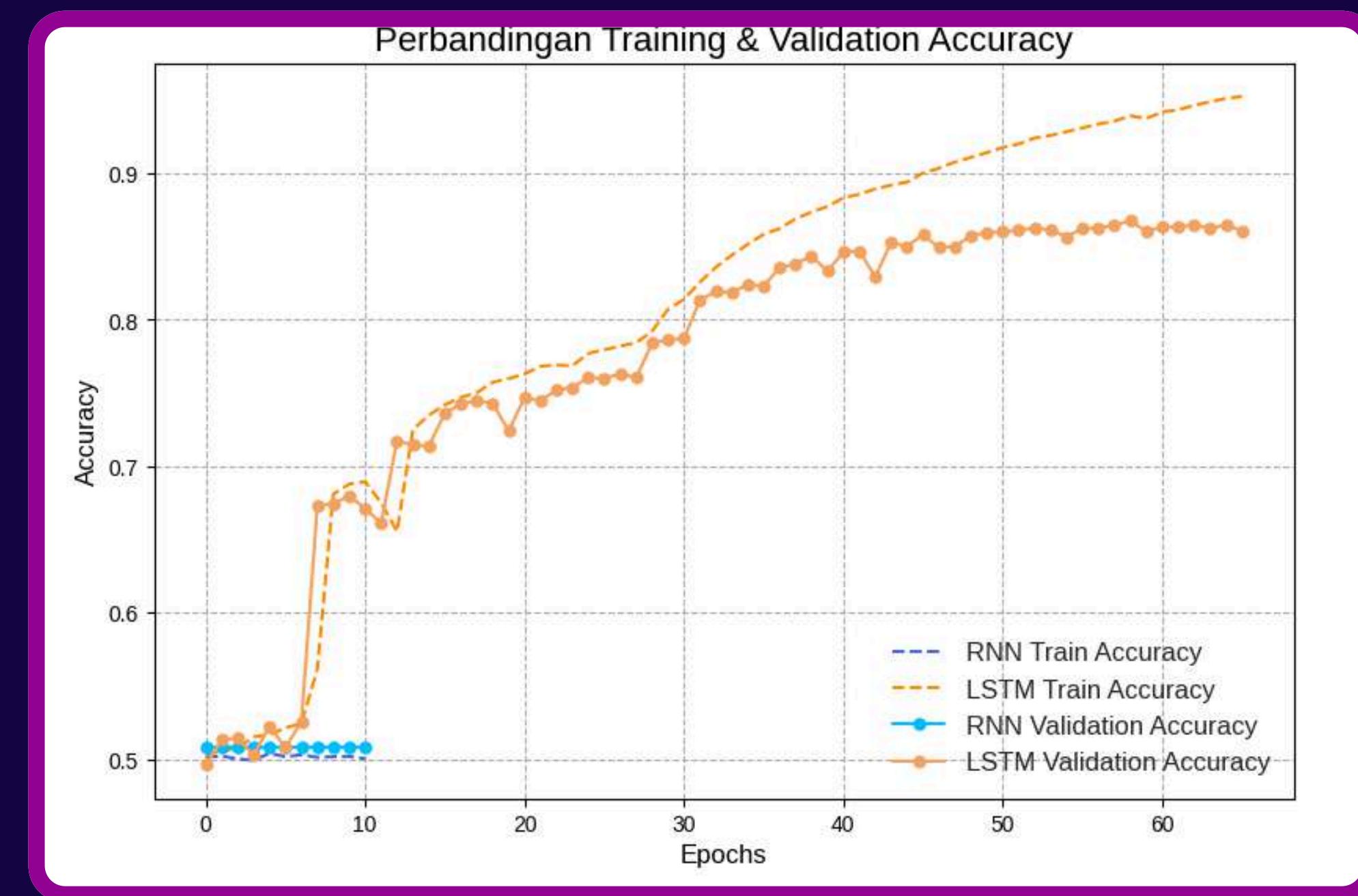
- Tetap stagnan di 0.7 tidak mengalami penurunan.

- Menurun signifikan hingga 0.3683 (epoch 56), kemudian mulai naik kembali (overfitting ringan).



RNN VS LSTM PERFORMANCE

ACCURACY





ACCURACY

RNN

- Meningkat lambat dan stagnan di kisaran 50%, tidak menunjukkan perkembangan berarti.

TRAIN

- Tetap stagnan di 50.83% selama 11 epoch.

TEST

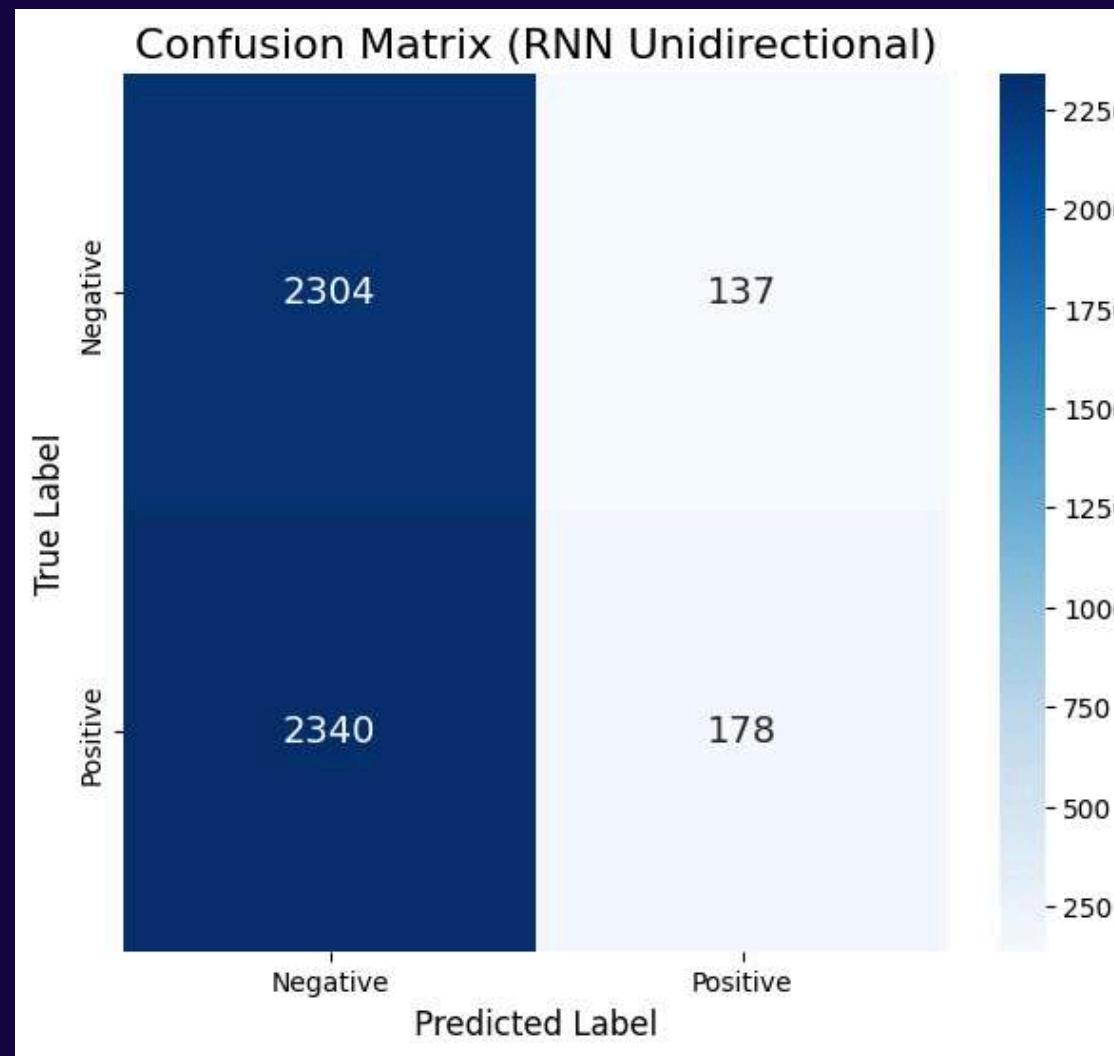
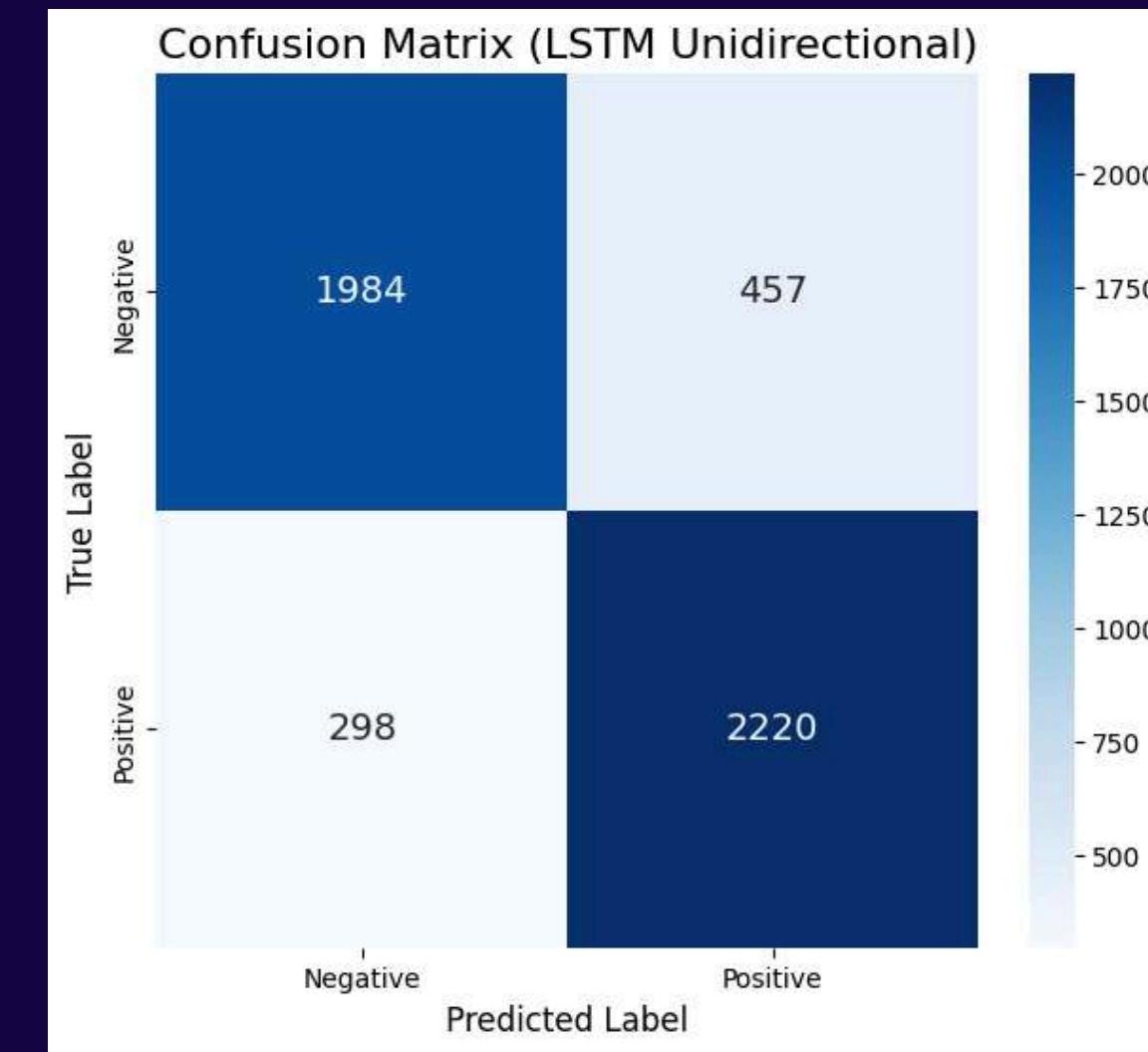
LSTM

- Meningkat pesat dari awal hingga mencapai 95.24% pada epoch ke-65 sebelum early stopping.

- Meningkat stabil dari 49.62% hingga 86.47%, mulai terlihat overfitting setelah epoch ke-52.



TEST DATA EVALUATION CONFUSION MATRIX

RNN**LSTM**



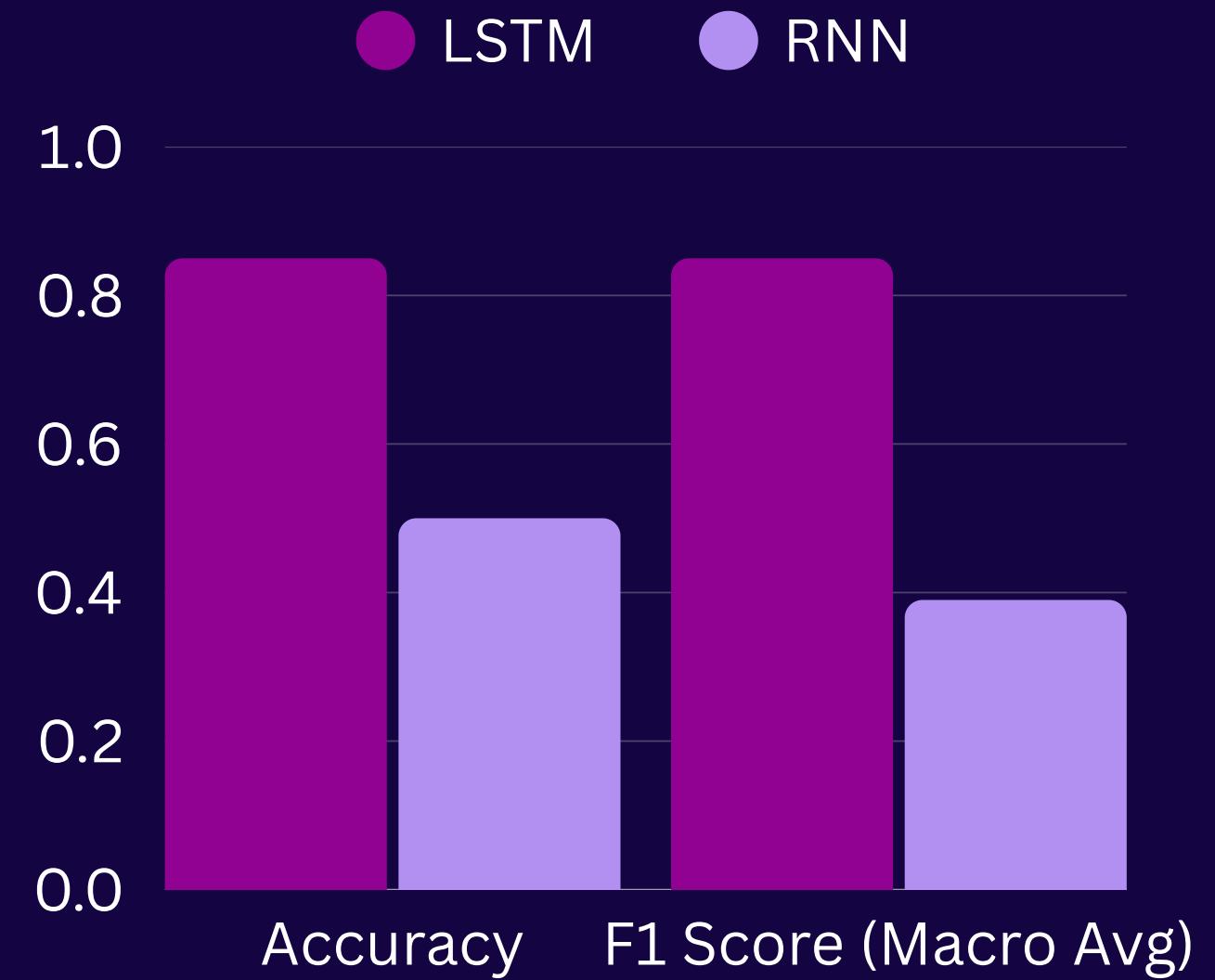
TEST DATA EVALUATION

RNN

Model sangat lemah terhadapa data text panjang.

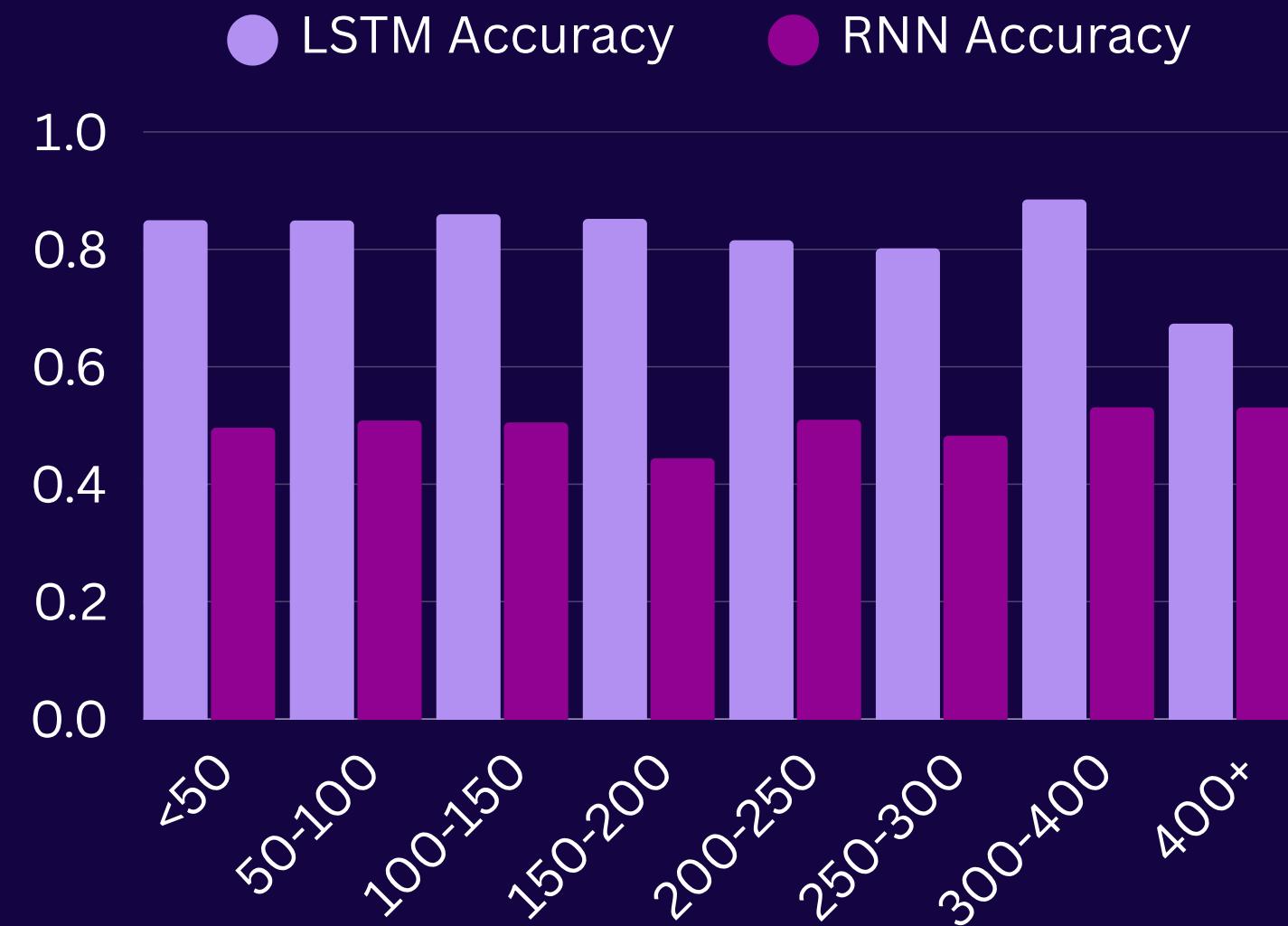
LSTM

Model jauh lebih kuat dibanding RNN.





ANALYSIS MISSCLASIFIED ACCURACY BY REVIEW LENGTH GROUP



- Akurasi RNN rendah di semua kelompok jumlah kata.
- Akurasi LSTM cukup stabil di >80% kecuali pada kata berjumlah 400 lebih.



KEY FINDINGS

- LSTM unidirectional mengungguli RNN unidirectional dalam klasifikasi sentimen pada dataset IMDB.
- RNN menunjukkan gejala vanishing gradient, terlihat dari nilai gradien yang kecil dan datar sepanjang training.
- LSTM memiliki gradien yang lebih dinamis dan stabil, membantu proses pembelajaran lebih efektif.
- Pada kelompok review panjang, LSTM mempertahankan akurasi jauh lebih baik dibanding RNN.
- Metrik evaluasi (akurasi, F1-score, loss) menunjukkan konsistensi performa LSTM di atas RNN.
- Meskipun LSTM unggul, review yang sangat panjang (>400 kata) tetap menjadi tantangan klasifikasi bagi model.



SUMMARY

- Terbukti LSTM adalah model yang lebih tepat digunakan untuk klasifikasi sentimen teks panjang seperti review IMDB.
- RNN cenderung gagal mempertahankan informasi dalam sequence panjang, karena keterbatasan arsitektur dan masalah vanishing gradient.



MATH - FST - UIN JAKARTA

kelompok 4



**TERIMA KASIH ATAS
PERHATIANNYA**

SAMPAI JUMPA LAGI