

Nama : Muhammad Husain Kaasyiful Ghitha

NIM : 1103220047

Kelas TK-45-G13

Deep Learning

Tugas Week 4 - Analisis Perbandingan Model RNN, LSTM, dan GRU

1. Persiapan

a. Dataset dan Preprocessing

- Dataset: IMDb dengan 25.000 ulasan film untuk training dan 25.000 untuk testing
- Vocabulary size: 40.000 kata (num_words=40000)
- Panjang maksimum sequence: 400 kata (maxlen=400)
- Semua sequence dipadding agar memiliki panjang yang sama

b. Arsitektur Model

Semua model diimplementasikan dengan arsitektur yang serupa untuk memastikan perbandingan yang adil:

- Embedding layer dengan dimensi 128
- Tiga layer bidirectional recurrent (RNN/LSTM/GRU) dengan dimensi hidden state 128
- Dropout layer (0.33) setelah recurrent layer
- Dense layer dengan 64 unit dan aktivasi ReLU
- Output layer dengan aktivasi sigmoid untuk klasifikasi biner

c. Training

- Optimizer: Adam dengan learning rate 0.001
- Loss function: Binary Cross-Entropy
- Batch size: 128
- Maksimum epoch: 20 dengan early stopping berdasarkan validation loss
- Validation split: 20% dari training data

d. Metrik Evaluasi

- Accuracy: Proporsi prediksi yang benar
- Precision: Proporsi prediksi positif yang benar
- Recall: Proporsi positif aktual yang diidentifikasi benar

- F1-Score: Rata-rata harmonik precision dan recall
- AUC-ROC: Area di bawah kurva ROC

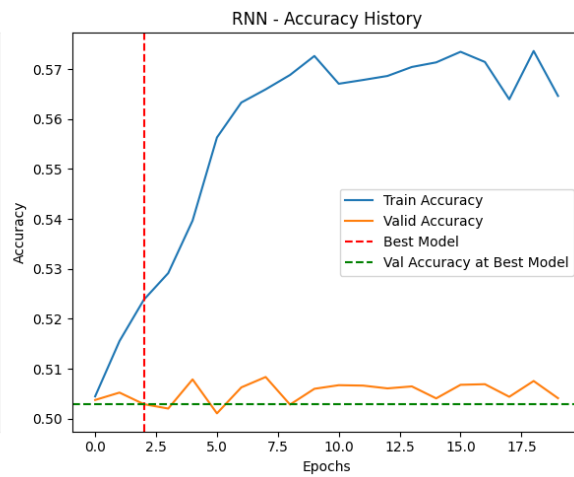
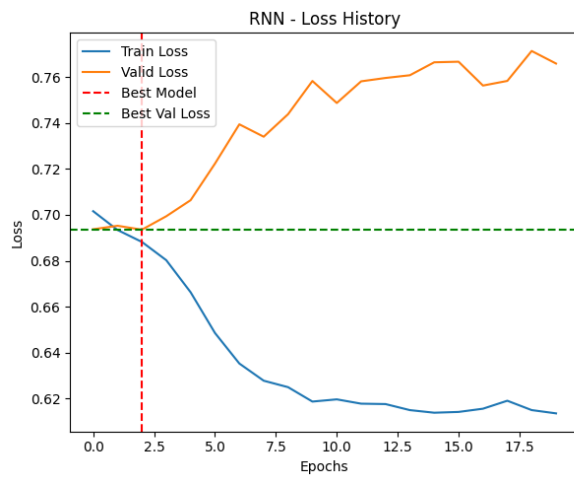
2. Hasil Evaluasi

Evaluasi Model PyTorch

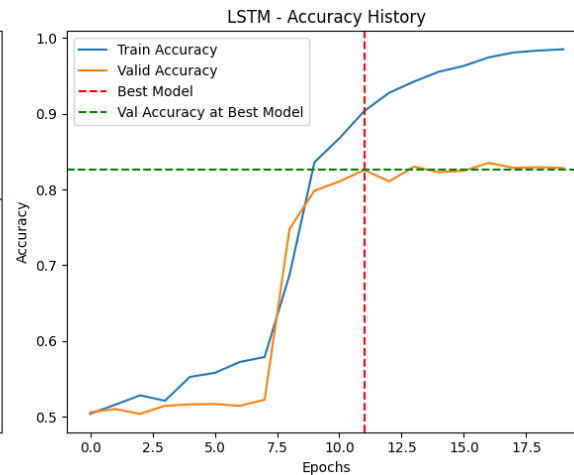
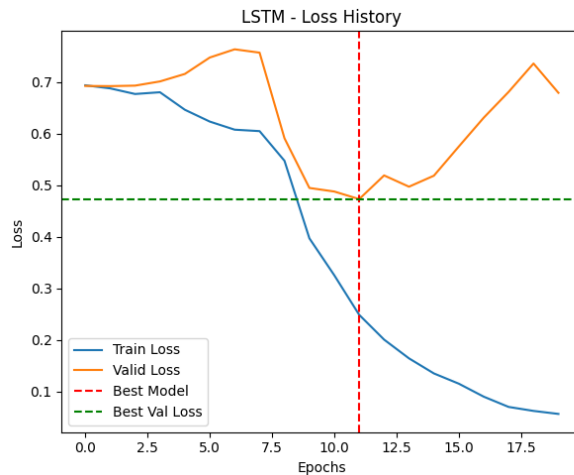
Model	Accuracy	Precision	Recall	F1-Score	AUC
RNN	0.5041	0.5193	0.1107	0.1825	0.5088
LSTM	0.8285	0.8500	0.7977	0.8230	0.8941
GRU	0.8632	0.8463	0.8878	0.8665	0.9356

Visualisasi Performa Training PyTorch

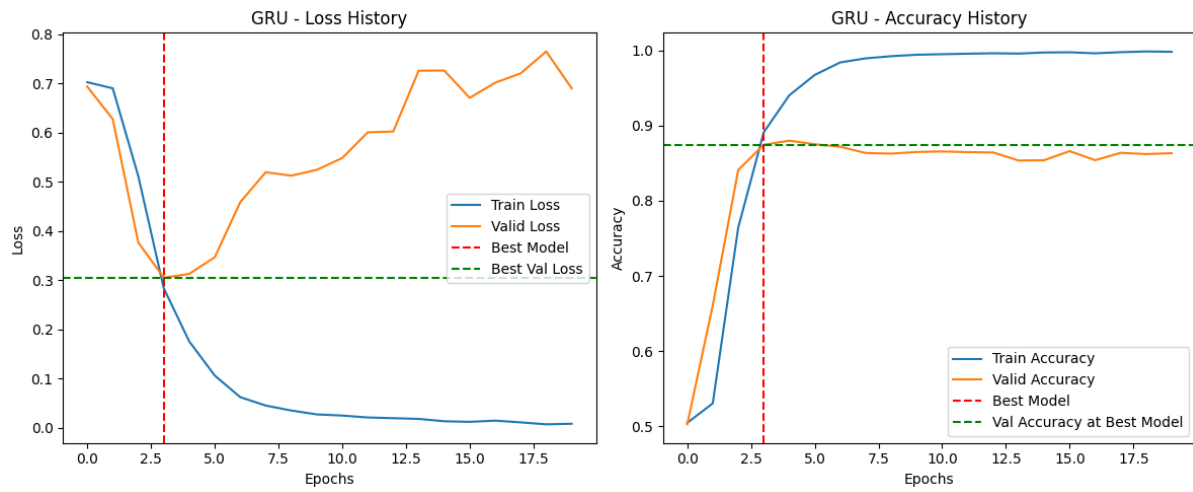
RNN



LSTM



GRU

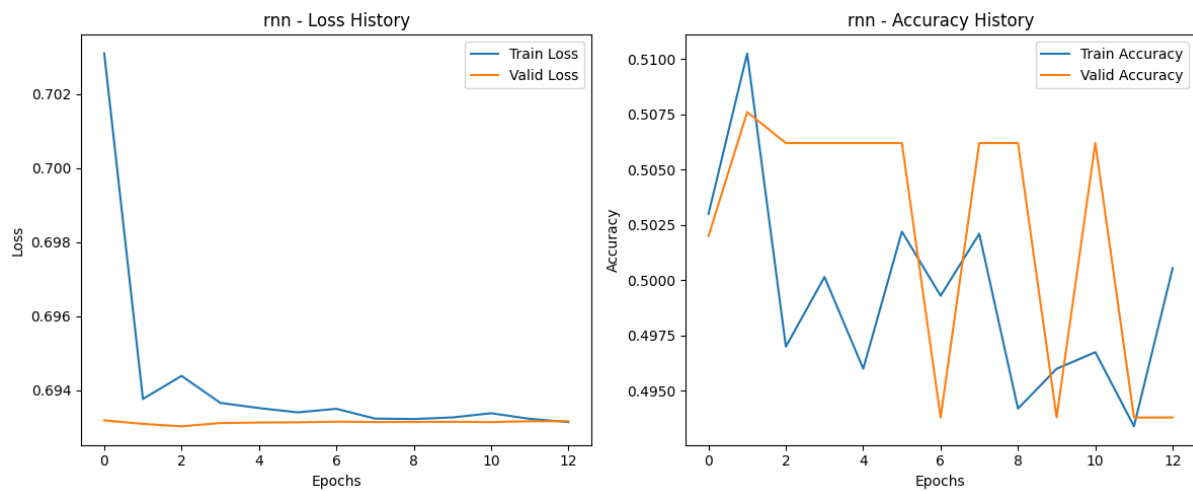


Evaluasi Model TensorFlow

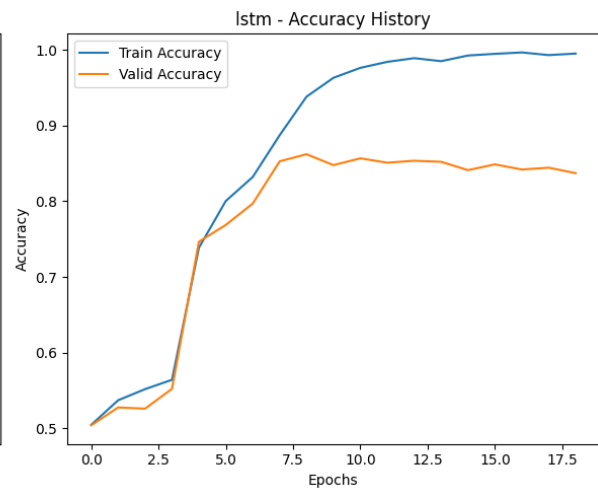
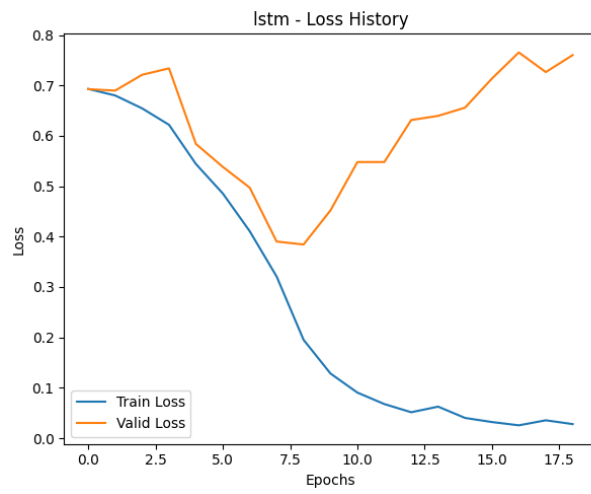
Model	Accuracy	Precision	Recall	F1-Score	AUC
RNN	0.5000	0.0000	0.0000	0.0000	0.5000
LSTM	0.8444	0.8541	0.8306	0.8422	0.9126
GRU	0.8517	0.8707	0.8261	0.8478	0.9174

Visualisasi Performa Training TensorFlow

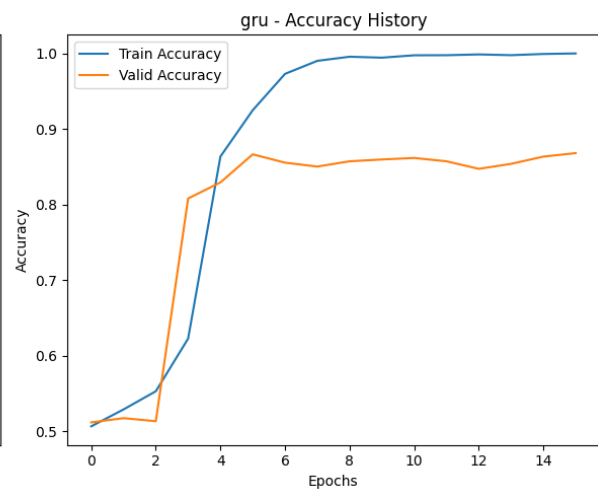
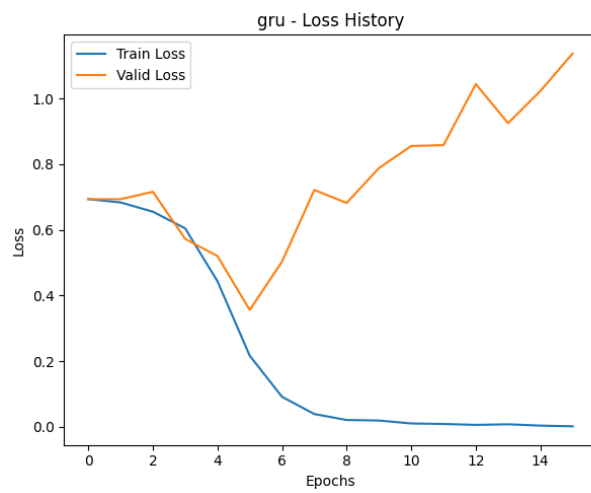
RNN



LSTM

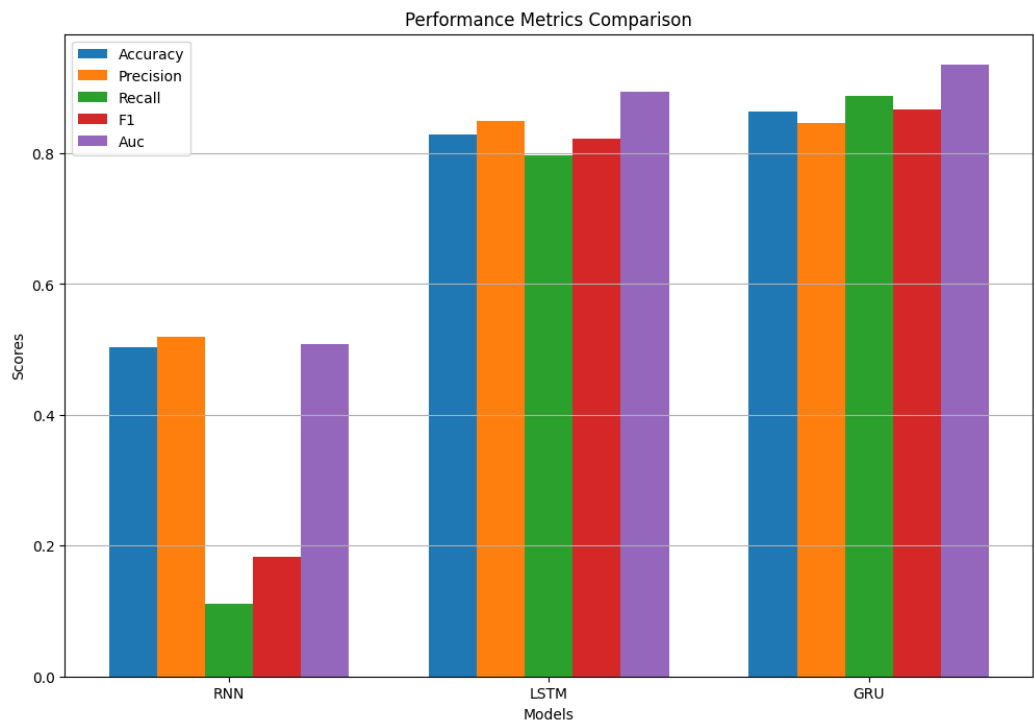


GRU

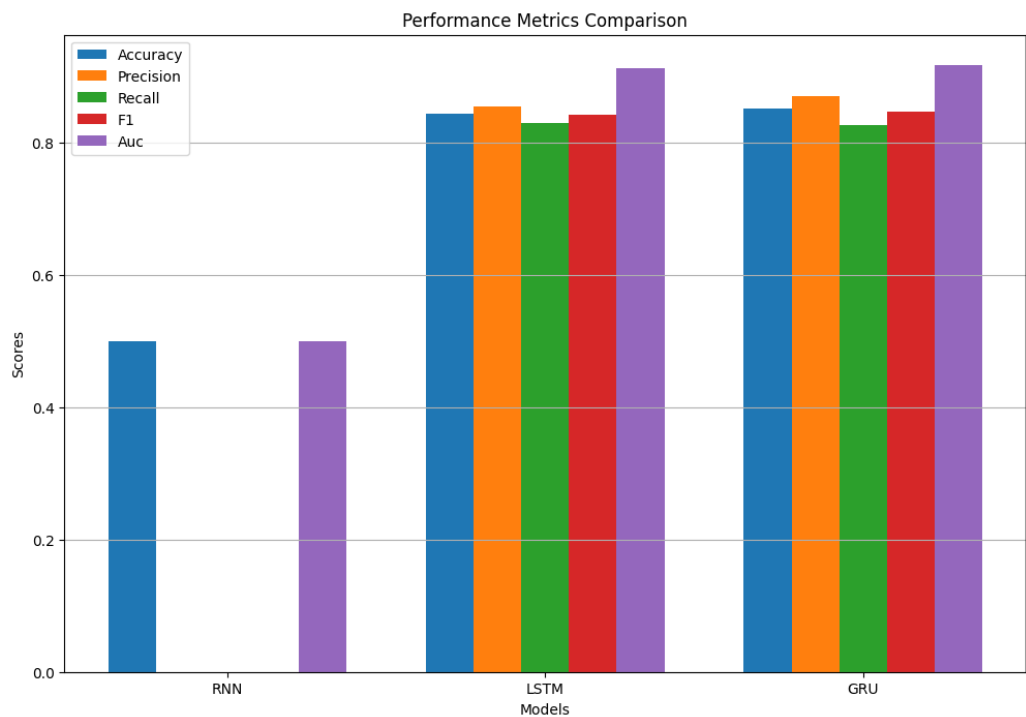


Perbandingan Performa Model

PyTorch



TensorFlow

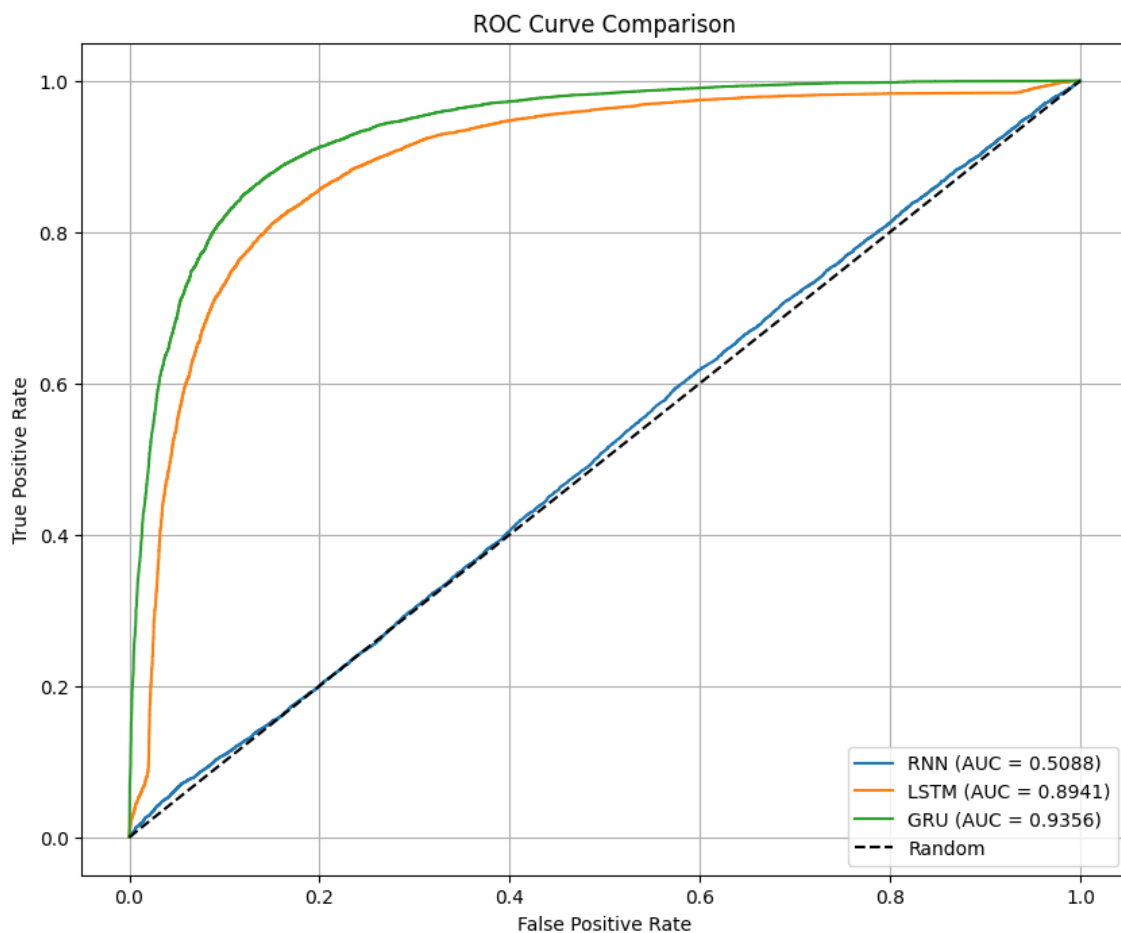


Pada kasus ini, Model GRU menunjukkan akurasi tertinggi (86.32% dengan PyTorch, 85.17% dengan TensorFlow), diikuti oleh LSTM (84.44% dengan TensorFlow, 82.85% dengan PyTorch) dan RNN standar (50.41% dengan PyTorch, 50.00% dengan TensorFlow).

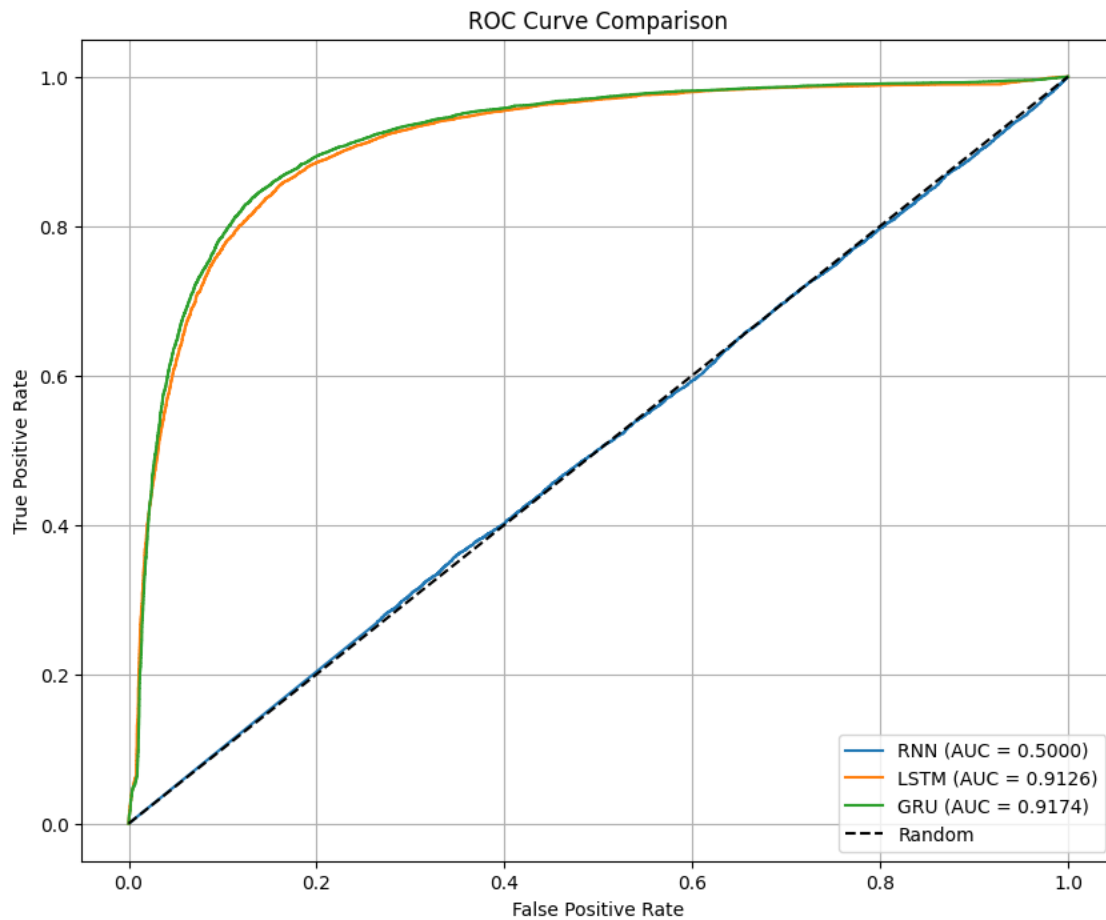
Serupa dengan akurasi, GRU mencapai F1-Score tertinggi (0.8665 dengan PyTorch, 0.8478 dengan TensorFlow), diikuti oleh LSTM (0.8422 dengan TensorFlow, 0.8230 dengan PyTorch) dan RNN standar (0.1825 dengan PyTorch, 0.0000 dengan TensorFlow).

Perbandingan ROC Curve

PyTorch



TensorFlow



GRU memiliki AUC tertinggi (0.9356 dengan PyTorch, 0.9174 dengan TensorFlow), walau memiliki arsitektur yang tidak sekompleks LSTM. Beberapa faktor, seperti data pre-processing dan hyperparameter tuning dapat berkontribusi dalam menentukan performa suatu model. Model yang lebih kompleks, seperti LSTM, mungkin akan memiliki performa lebih baik apabila dataset dan hyperparameter diatur dengan lebih baik.

Perbandingan Waktu Training

Model	TensorFlow (detik/epoch)	PyTorch (detik/epoch)
RNN	53	47
LSTM	102	96
GRU	82	78

Model GRU memiliki performa dan efisiensi komputasi terbaik, membutuhkan waktu training ~20% lebih cepat dan performa yang sedikit yang lebih baik dibandingkan LSTM.

3. Analisis Kekuatan dan Kelemahan

a. RNN Standar

Kekuatan:

- Arsitektur paling sederhana dengan jumlah parameter paling sedikit.
- Training paling cepat.
- Sumber daya yang dibutuhkan paling sedikit.

Kelemahan:

- Performa terendah di semua metrik evaluasi, bahkan menghasilkan model mati pada model TensorFlow.
- Mengalami masalah vanishing gradient, terutama untuk sequence panjang.
- Tidak dapat menangkap dependensi jangka panjang dengan baik.

b. LSTM

Kekuatan:

- Performa yang paling andal pada pelatihan yang panjang.
- Mampu menangkap dependensi jangka panjang.
- Mengatasi masalah vanishing gradient dengan mekanisme gating.

Kelemahan:

- Arsitektur paling kompleks dengan jumlah parameter terbanyak.
- Training paling lambat.
- Memori yang dibutuhkan paling banyak.

c. GRU

Kekuatan:

- Performa hampir setara dengan LSTM, atau bahkan melebihi LSTM pada kasus ini.
- Lebih sederhana daripada LSTM dengan jumlah parameter lebih sedikit.

- Training lebih cepat daripada LSTM.

Kelemahan:

- Sedikit kurang performant dibandingkan LSTM untuk data yang lebih besar.
- Masih membutuhkan lebih banyak komputasi dibandingkan RNN standar.

4. Kesimpulan

Berdasarkan eksperimen yang telah dilakukan, beberapa kesimpulan dapat ditarik:

1. GRU menawarkan kompromi efisiensi-performa yang paling baik. membutuhkan waktu training yang lebih singkat dan parameter yang lebih sedikit, sekaligus memberikan performa yang sebanding dengan LSTM.
2. RNN standar tertinggal signifikan. Meskipun simpel dan cepat, RNN standar menunjukkan performa yang secara signifikan lebih rendah dibandingkan LSTM dan GRU, mengonfirmasi keterbatasannya dalam menangkap dependensi jangka panjang.