

Nama : Hamdan Syaifuddin Zuhri
NIM : 1103220220
Kelas : TK-45-G13 – Deep Learning

Analisis PyTorch Model RNN, LSTM, dan GRU Pada Dataset IMDb

Berdasarkan hasil evaluasi model RNN pada dataset IMDb, kita bisa melihat bahwa model berhasil mengklasifikasikan data dengan baik. Mari kita breakdown lebih lanjut metrik evaluasi dan memberikan analisis dari hasil yang didapatkan.

1. Akurasi (Accuracy)

Akurasi: 0.8616

Artinya, model berhasil mengklasifikasikan sekitar 86.16% data dengan benar. Angka ini cukup baik untuk tugas klasifikasi biner, mengingat kita hanya menggunakan model RNN tanpa optimasi yang lebih lanjut. Akurasi yang tinggi menunjukkan bahwa model memiliki kemampuan yang baik untuk membedakan antara kelas 0 dan kelas 1.

2. Presisi (Precision)

Presisi untuk kelas 1: 0.8533

Presisi mengukur proporsi prediksi positif yang benar di antara seluruh prediksi positif yang dibuat oleh model. Dalam hal ini, model memiliki presisi sekitar 85.33% dalam memprediksi kelas 1 (positif). Artinya, ketika model memprediksi kelas 1, 85.33% dari prediksi tersebut adalah benar.

3. Recall

Recall untuk kelas 1: 0.8732

Recall mengukur seberapa baik model dalam menemukan semua contoh positif yang sebenarnya. Dalam hal ini, model berhasil mengidentifikasi 87.32% dari seluruh contoh kelas 1 yang sebenarnya. Dengan kata lain, model tidak banyak melewatkan data kelas 1, namun masih ada sedikit data yang terlewat (False Negatives).

4. F1-Score

F1-Score untuk kelas 1: 0.8632

F1-Score adalah rata-rata harmonis antara presisi dan recall. Nilai 0.8632 menunjukkan bahwa model mencapai keseimbangan yang baik antara presisi dan recall. F1-score yang tinggi ini mencerminkan bahwa model RNN tidak hanya mengidentifikasi sebagian besar contoh positif (recall tinggi), tetapi juga mengurangi kesalahan prediksi positif (presisi tinggi).

5. AUC (Area Under the Curve)

AUC: 0.8616

AUC mengukur seberapa baik model dalam membedakan antara kelas positif dan negatif. Nilai AUC 0.8616 menunjukkan bahwa model cukup baik dalam memisahkan kedua kelas tersebut. AUC yang mendekati 1 menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam melakukan klasifikasi.

6. Macro Average dan Weighted Average

Macro Average: Rata-rata dari presisi, recall, dan f1-score untuk setiap kelas, tanpa mempertimbangkan ketidakseimbangan kelas. Dalam hal ini, nilai-nilai macro average menunjukkan bahwa model RNN memiliki performa yang konsisten antara kelas 0 dan kelas 1, dengan sedikit perbedaan antara kedua kelas tersebut.

Weighted Average: Ini adalah rata-rata tertimbang yang mempertimbangkan jumlah contoh di setiap kelas. Dengan jumlah support yang sama untuk kedua kelas (12,500 untuk masing-masing kelas), nilai weighted average ini sama dengan macro average, yang lebih menekankan bahwa kedua kelas diperlakukan dengan seimbang.

Kesimpulan

Secara keseluruhan, model RNN ini menunjukkan performa yang sangat baik pada tugas klasifikasi biner dengan dataset IMDB. Akurasi 86.16% menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar. Presisi, recall, dan F1-score yang tinggi menunjukkan bahwa model ini tidak hanya mampu mengenali banyak contoh kelas positif (kelas 1), tetapi juga meminimalkan kesalahan dalam prediksi.

Namun, meskipun model ini memiliki performa yang baik, masih ada ruang untuk perbaikan. Misalnya, menggunakan model LSTM atau GRU yang lebih kompleks dapat meningkatkan hasil lebih lanjut, terutama dalam menangani ketergantungan jangka panjang dalam data sekuensial seperti teks.