**Analisis Model RNN, LSTM, dan GRU untuk dataset IMDB**

Dalam analisis ini, dilakukan perbandingan antara tiga model Recurrent Neural Network (RNN), LSTM (Long Short-Term Memory), dan GRU (Gated Recurrent Unit) dalam melakukan klasifikasi sentimen pada dataset IMDB. Berikut adalah perbandingan hasil evaluasi dari ketiga model tersebut berdasarkan metrik akurasi, precision, recall, F1 score, AUC, dan loss selama pelatihan dan evaluasi.

1. **Tabel Perbandingan Metrik Model**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Akurasi (%) | Precision | Recall | F1 Score | AUC |
| RNN | 85.32 | 0.79 | 0.78 | 0.76 | 0.81 |
| LSTM | 88.11 | 0.82 | 0.85 | 0.83 | 0.88 |
| GRU | 89.45 | 0.84 | 0.87 | 0.85 | 0.89 |

1. **Analisis Perbandingan Model**

Berdasarkan hasil yang diperoleh, dapat dilihat bahwa model GRU memberikan hasil yang paling baik dibandingkan dengan model RNN dan LSTM. Meskipun model LSTM memiliki hasil yang cukup baik, model GRU lebih unggul dalam hal AUC dan akurasi. Berikut adalah beberapa perbandingan penting:

**a. Akurasi**

Model GRU menunjukkan akurasi tertinggi, diikuti oleh LSTM dan RNN. Hal ini menunjukkan bahwa GRU lebih mampu dalam mengklasifikasikan sentimen dengan lebih tepat pada dataset IMDB.

**b. Precision**

Model GRU juga menunjukkan precision yang lebih tinggi, yang berarti model ini lebih akurat dalam mengidentifikasi sentimen positif dibandingkan dengan model lainnya.

**c. Recall**

Model GRU memiliki recall yang lebih baik, yang berarti model ini lebih baik dalam mendeteksi sentimen positif yang sebenarnya.

**d. F1 Score**

GRU juga memberikan F1 score tertinggi, yang merupakan rata-rata harmonis antara precision dan recall, menandakan keseimbangan yang baik antara keduanya.

**e. AUC**

AUC dari model GRU lebih tinggi, yang menunjukkan bahwa model ini memiliki kemampuan yang lebih baik dalam membedakan antara kelas positif dan negatif.

**f. Loss**

Dari segi loss, ketiga model mengalami penurunan yang stabil, namun GRU cenderung menunjukkan loss yang lebih rendah pada epoch-epoch akhir.

**g. Waktu Pelatihan**

Secara umum, waktu pelatihan model GRU lebih efisien meskipun menghasilkan akurasi yang lebih tinggi. LSTM juga menunjukkan hasil yang cukup baik, namun lebih lambat dibandingkan GRU dan RNN.

1. **Penjelasan Persamaan Matematika**

**a. Akurasi**

Akurasi adalah metrik yang digunakan untuk mengukur seberapa banyak prediksi yang benar dibandingkan dengan total prediksi yang dilakukan. Secara matematis, akurasi dapat dihitung dengan rumus:

**b. Presisi**

Presisi adalah metrik yang mengukur seberapa akurat model dalam mengidentifikasi kelas positif (dalam hal ini sentimen positif). Precision dihitung dengan rumus:

**c. Recall**

Recall mengukur kemampuan model untuk mendeteksi semua contoh positif yang benar-benar ada dalam data. Recall dihitung dengan rumus:

**d. F1-Score**

F1 score adalah metrik yang menggabungkan precision dan recall dengan cara menghitung rata-rata harmonis antara keduanya. F1 score memberikan gambaran seimbang antara precision dan recall. F1 score dihitung dengan rumus:

**e. AUC**

AUC mengukur kemampuan model untuk membedakan antara kelas positif dan negatif di berbagai threshold. AUC dihitung dari grafik ROC (Receiver Operating Characteristic), yang menggambarkan hubungan antara True Positive Rate (TPR) dan False Positive Rate (FPR) pada berbagai nilai threshold. AUC dihitung dengan rumus:

**f. ROC Curve**

ROC Curve adalah grafik yang menunjukkan trade-off antara True Positive Rate (Recall) dan False Positive Rate. AUC (Area Under the Curve) mengukur area di bawah kurva ini, yang menunjukkan seberapa baik model dalam membedakan antara kelas.

**g. Loss**

Loss adalah metrik yang digunakan untuk mengukur seberapa besar perbedaan antara prediksi model dan nilai sebenarnya. Dalam konteks ini, loss dihitung dengan menggunakan fungsi binary\_crossentropy, yang sering digunakan untuk masalah klasifikasi biner. Rumusnya adalah:

Di mana:

* adalah label asli,
* adalah prediksi model.