

Nama :Rahmi Agustin

Nim : 12145032

Kelas : RB

Tugas Terakhir NLP

## 1. Dataset

- **Ukuran Dataset:** Dataset yang lebih besar dapat membantu meningkatkan akurasi dan generalisasi, terutama jika model yang digunakan adalah kompleks seperti Transformer atau BERT. Dataset kecil sering kali tidak cukup untuk memanfaatkan kapasitas model ini.
- **Implikasi:**
  - Dataset besar diperlukan untuk model berbasis BERT atau Transformer agar mampu belajar fitur yang lebih mendalam.
  - Untuk model seperti LSTM atau GRU, dataset yang lebih kecil sering cukup jika preprocessing dan tokenisasi dilakukan dengan benar.
- **Rekomendasi:**
  - Jika dataset Anda kecil, pertimbangkan penggunaan **teknik augmentasi data** atau pre-trained embeddings (seperti GloVe atau FastText) untuk meningkatkan performa tanpa perlu dataset besar.

## 2. Waktu dan Sumber Daya Komputasi

- **Komputasi:**
  - **LSTM/GRU:** Memerlukan lebih sedikit komputasi dibanding Transformer atau BERT.
  - **Transformer/BERT:** Membutuhkan GPU/TPU untuk melatih dengan waktu yang masuk akal.
- **Kecepatan Pelatihan:**
  - Model LSTM lebih cepat untuk dataset kecil hingga sedang.
  - Model seperti BERT memerlukan waktu lebih lama, tetapi mampu menangkap hubungan kontekstual yang lebih baik dalam data teks.
- **Rekomendasi:**
  - Jika sumber daya terbatas, gunakan LSTM atau GRU dengan pre-trained embeddings.
  - Untuk mempercepat pelatihan, pastikan menggunakan **early stopping**, **batch size besar**, dan GPU.

## 3. Generalisasi

- **Kemampuan Generalisasi:**
  - **LSTM/GRU:** Baik untuk menangkap hubungan jangka pendek dan cukup untuk dataset dengan ukuran kecil hingga sedang.
  - **Transformer:** Baik untuk menangkap hubungan jangka panjang dan konteks, cocok untuk dataset besar.

- **BERT**: Performa generalisasi yang sangat baik karena pre-training pada data besar, tetapi membutuhkan fine-tuning yang hati-hati.
- **Hasil:**
  - Model berbasis BERT biasanya memberikan akurasi, recall, dan F1-score yang lebih tinggi pada dataset besar. Berikut merupakan hasil akurasi dari beberapa model yang menunjukkan model bert mendapatkan akurasi lebih baik dari model lainnya:

Model/Hasil	Precision	Recall	F1 Score	Akurasi
<b>GRU</b>	0.2279	0.2528	0.1100	0.2597
<b>Fasttext</b>	0.4474	0.4161	0.4167	0.4176
<b>Transformer</b>	0.4474	0.4161	0.4167	0.4176
<b>Bert</b>	0.5282	0.4606	0.4352	0.4648

Berdasarkan tabel hasil akurasi diperoleh bahwa model BERT memiliki performa terbaik dalam menangani data tersebut. Hal ini ditunjukkan oleh nilai Precision (0.5282), Recall (0.4606), dan F1 Score (0.4352) yang lebih tinggi dibandingkan model lainnya. Akurasi BERT (0.4648) juga menunjukkan keunggulan dibandingkan model GRU, FastText, dan Transformer. Nilai-nilai ini menunjukkan BERT lebih baik

- **Rekomendasi:**
  - Untuk generalisasi yang baik pada dataset kecil, gunakan pre-trained embeddings seperti GloVe atau FastText.
  - Untuk dataset besar, gunakan Transformer atau BERT dengan fine-tuning.

### Perbandingan Parameter Model

Model	Kelebihan	Kekurangan	Cocok untuk
<b>LSTM</b>	Baik untuk hubungan jangka pendek.	Sulit menangkap konteks global.	Dataset kecil hingga sedang.
<b>GRU</b>	Lebih cepat dibanding LSTM.	Performanya sedikit di bawah LSTM.	Dataset kecil hingga sedang.
<b>FastText</b>	Cepat, ringan, pre-trained tersedia.	Tidak menangkap konteks.	Aplikasi sederhana.
<b>Transformer</b>	Baik untuk hubungan jangka panjang.	Membutuhkan banyak sumber daya komputasi.	Dataset besar.
<b>BERT</b>	Generalisasi luar biasa dengan fine-tuning.	Pelatihan lambat, membutuhkan banyak data.	Dataset besar.