**Laporan Analisis Sentimen Menggunakan Model Recurrent Neural Network (RNN)**

**Iqbal Maulana**

**Teknik Informatika**

**A1 | 442023611094**

# Deskripsi Proyek

Proyek ini bertujuan untuk mencari hasil implementasi model Recurrent Neural Network (RNN) untuk analisis sentimen. Analisis sentimen adalah proses untuk menentukan polaritas emosional (positif, negatif, atau netral) dari suatu teks. Dalam konteks ini, model RNN dilatih untuk mengklasifikasikan ulasan teks sebagai sentimen positif atau negatif.

Tujuan utama dari proyek ini adalah untuk membangun dan melatih model RNN untuk klasifikasi sentimen, mengevaluasi kinerja model dalam memprediksi sentimen dari teks baru dan mendemonstrasikan proses *preprocessing* teks dan penggunaan model terlatih untuk inferensi.

# Dataset

Dataset yang digunakan dalam proyek ini adalah kumpulan ulasan sentimen yang terbagi menjadi dua kategori utama: positif dan negatif, dengan pelabelan 1 untuk positif, dan 0 untuk negatif.

# Pra-pemrosesan Data

Langkah-langkah pra-pemrosesan data untuk melatih model meliputi:

* Tokenisasi: Teks dipecah menjadi unit-unit yang lebih kecil (token)
* Pembuatan Urutan (Sequencing): Token-token diubah menjadi urutan numerik berdasarkan indeks dalam kamus kata (vocabulary)
* Padding: Urutan yang dihasilkan disamakan panjangnya dengan menambahkan nol (padding) agar sesuai dengan panjang input maksimum yang diharapkan oleh model (dalam kasus ini, maxlen=20)

**Arsitektur Model: Recurrent Neural Network (RNN)**

Model yang digunakan adalah model Recurrent Neural Network. Model RNN untuk analisis sentimen melibatkan:

* Embedding Layer: Mengubah input indeks kata menjadi representasi vektor padat.
* RNN Layer (misalnya, SimpleRNN, LSTM, GRU): Memproses urutan input dan menangkap dependensi temporal.
* Dense Layer(s): Lapisan fully connected untuk klasifikasi akhir.
* Output Layer: Lapisan dengan fungsi aktivasi sigmoid untuk klasifikasi biner (positif/negatif).

**Pelatihan Model**

Model dikompilasi dengan menentukan optimizer, fungsi *loss*, dan metrik evaluasi:

* Optimizer: Tidak disebutkan secara spesifik, tetapi *optimizer* seperti Adam atau RMSprop sering digunakan karena efektivitasnya dalam konvergensi model *deep learning*.
* Fungsi Loss: Untuk klasifikasi biner, *binary cross-entropy* adalah fungsi *loss* yang standar. Fungsi ini mengukur seberapa jauh prediksi model dari label sebenarnya.
* Metrik Evaluasi: Akurasi biasanya digunakan sebagai metrik utama untuk memantau kinerja model selama pelatihan dan evaluasi.

Proses pelatihan melibatkan *forward pass* (menghitung prediksi), *backward pass* (menghitung gradien *loss*), dan pembaruan bobot model menggunakan *optimizer* selama sejumlah *epoch* (iterasi penuh atas seluruh dataset pelatihan).

**Hasil dan Analisis**

Untuk mengevaluasi kemampuan model dalam mengklasifikasikan sentimen, dua contoh kalimat ulasan diuji:

Contoh 1: "The food was cheap and delicious and also the service"

1. Pra-pemrosesan Teks: Setelah pembersihan dan tokenisasi (dan mungkin penghapusan *stop words*), kalimat diubah menjadi "food cheap delicious also service". Ini menunjukkan bahwa model mampu mengekstrak kata-kata kunci yang membawa makna sentimen.
2. Urutan Input: Teks ini kemudian diubah menjadi urutan numerik dan di-*padded* hingga panjang 20.
3. Nilai Prediksi Mentah: Model menghasilkan nilai prediksi mentah sebesar 0.7031135559082031.
4. Prediksi Sentimen: Berdasarkan ambang batas 0.5, nilai 0.7031 ini diklasifikasikan sebagai Positif.
5. Analisis: Prediksi ini sangat konsisten dengan sentimen yang melekat pada kalimat tersebut. Frasa "cheap and delicious" dan "service" secara jelas mengindikasikan pengalaman yang positif. Model dengan tepat menangkap nuansa ini, menunjukkan keberhasilan dalam mengidentifikasi sentimen positif.

Contoh 2: "what a waste of money"

1. Pra-pemrosesan Teks: Teks ini kemungkinan diubah menjadi "waste money" atau sejenisnya, setelah penghapusan *stop words*.
2. Nilai Prediksi Mentah: Meskipun nilai numerik spesifik tidak ditampilkan dalam cuplikan output, prediksi untuk kalimat ini adalah Negatif.
3. Analisis: Frasa "waste of money" secara intrinsik adalah ekspresi ketidakpuasan atau sentimen negatif. Model berhasil mengidentifikasi hal ini, yang mengindikasikan kemampuannya untuk mengenali pola-pola bahasa yang terkait dengan sentimen negatif.

Kedua contoh kasus ini menunjukkan bahwa model yang telah dilatih memiliki pemahaman dasar yang kuat tentang bagaimana kata-kata dan frasa berkontribusi pada sentimen keseluruhan suatu ulasan. Tingkat kepercayaan (nilai prediksi mentah) yang dihasilkan juga memberikan indikasi kualitas prediksi.

# Evaluasi Model

Untuk meningkatkan kinerja model secara keseluruhan diperlukan datasaet yang lebih besar. Dengan melatih model pada dataset yang lebih besar dan lebih beragam akan meningkatkan kemampuan generalisasi model untuk menghadapi variasi bahasa dan konteks yang lebih luas.

# Kesimpulan

Model Recurrent Neural Network yang diimplementasikan dalam proyek ini telah berhasil menunjukkan kemampuannya dalam melakukan analisis sentimen biner pada ulasan teks. Dengan serangkaian langkah pra-pemrosesan yang efektif dan arsitektur RNN yang sesuai, model mampu mengidentifikasi sentimen positif dan negatif dengan akurasi yang menjanjikan pada contoh-contoh yang diberikan. Keberhasilan ini menyoroti potensi *deep learning* dalam tugas-tugas NLP yang kompleks.