DAFTAR ISTILAH

Accuracy	:	Persentase prediksi yang benar dari total jumlah data yang diuji.
Activation Function	:	Fungsi matematis dalam neuron jaringan saraf untuk menentukan output berdasarkan input, misalnya sigmoid atau ReLU.
Annotated Data	:	Data yang telah diberi label (misal *aware/not aware*) sebagai acuan untuk pelatihan atau evaluasi model.
Awareness	:	Tingkat kesadaran atau pemahaman individu terhadap suatu isu, dalam konteks ini adalah kesehatan mental.
Aware	:	Label untuk komentar yang menunjukkan kesadaran terhadap isu kesehatan mental.
Bag of Words	:	Representasi teks sebagai kumpulan kata tanpa memperhatikan urutan atau konteks kata.
Batch Size	:	Jumlah sampel data yang diproses sekaligus sebelum memperbarui parameter model.
Bidirectional LSTM	:	LSTM yang memproses data dalam dua arah—maju dan mundur—untuk menangkap konteks yang lebih luas.
Classification Report	:	Laporan hasil evaluasi model klasifikasi, berisi metrik seperti accuracy, precision, recall, dan F1-score.
Class Imbalance	:	Ketidakseimbangan jumlah data antara dua atau lebih kelas dalam dataset.
Comment Corpus	:	Sekumpulan komentar teks dari TikTok yang dianalisis dalam penelitian.
Confusion Matrix	:	Matriks evaluasi performa model klasifikasi yang menggambarkan perbandingan antara hasil prediksi dan kondisi aktual.
Corpus	:	Sekumpulan dokumen teks yang digunakan sebagai bahan analisis atau pelatihan model.
Cross-Validation	:	Teknik evaluasi model dengan membagi data ke dalam beberapa subset untuk menguji generalisasi model.
Dropout	:	Teknik regulasi dalam neural network untuk mencegah overfitting dengan menghilangkan sebagian neuron secara acak saat pelatihan.
Early Stopping	:	Mekanisme untuk menghentikan pelatihan model secara otomatis ketika nilai parameter tertentu tidak membaik lagi.
Embedding Layer	:	Lapisan dalam neural network untuk mengubah kata menjadi vektor berdimensi tetap.
Epoch	:	Satu iterasi penuh ketika seluruh dataset digunakan dalam pelatihan model.
Evaluation Metrics	:	Ukuran atau indikator performa model, seperti accuracy, recall, precision, dan F1-score.
FastText	:	Model *word embedding* yang mempertimbangkan sub-kata (subword) untuk menangani kata-kata tidak dikenal.

Feature Extraction	:	Proses mengekstraksi informasi penting dari data mentah untuk digunakan dalam pelatihan model.
F1-Score	:	Harmonic mean dari precision dan recall; menggambarkan keseimbangan antara keduanya.
FN *(False Negative)*	:	Prediksi model yang salah mengklasifikasikan data positif sebagai negatif.
FP *(False Positive)*	:	Prediksi model yang salah mengklasifikasikan data negatif sebagai positif.
Gensim	:	*Library Python untuk NLP, terutama untuk topic modelling seperti LDA.*
Hyperparameter	:	Parameter eksternal model yang ditentukan sebelum pelatihan, seperti jumlah neuron, learning rate, dll.
KDD *(Knowledge Discovery in Database)*	:	Proses eksplorasi dan ekstraksi pengetahuan dari data yang besar dan kompleks.
Labelling	:	Proses memberi anotasi atau label pada data, misalnya komentar TikTok diberi label aware atau not aware.
LDA *(Latent Dirichlet Allocation)*	:	Algoritma probabilistik untuk mengidentifikasi topik dalam kumpulan dokumen teks.
Learning Rate	:	Kecepatan pembaruan bobot dalam proses pelatihan model neural network.
Lexicon	:	Kumpulan kata atau frasa yang digunakan sebagai dasar pelabelan berdasarkan kamus kata.
Loss Function	:	Fungsi yang digunakan untuk mengukur kesalahan antara prediksi model dan label aktual.
LSTM *(Long Short-Term Memory)*	:	Jenis jaringan saraf berulang (RNN) yang mampu menyimpan informasi dalam jangka waktu panjang, cocok untuk data teks berurutan.
Not Aware	:	Label untuk komentar yang tidak menunjukkan kesadaran terhadap isu kesehatan mental.
Overfitting	:	Kondisi ketika model terlalu cocok dengan data pelatihan hingga gagal melakukan generalisasi terhadap data baru.
Preprocessing	:	Proses pembersihan dan transformasi data mentah sebelum dianalisis lebih lanjut.
Precision	:	Rasio antara jumlah prediksi positif yang benar dibandingkan dengan total prediksi positif.
PyLDAVis	:	Tool visualisasi untuk model LDA yang menampilkan *intertopic distance map*.
Recall	:	Rasio antara jumlah prediksi positif yang benar dibandingkan dengan total data positif aktual.
Regularization	:	Teknik untuk mencegah overfitting dengan menambahkan penalti dalam fungsi loss.
RNN *(Recurrent Neural Network)*	:	Jaringan saraf yang digunakan untuk memproses data sekuensial dengan mempertimbangkan urutan input.
Sigmoid	:	Fungsi aktivasi yang digunakan pada output layer untuk menghasilkan probabilitas antara 0 dan 1.

Stopwords	:	Kata-kata umum yang dihilangkan dalam pemrosesan teks karena dianggap tidak memiliki makna penting dalam konteks analisis (misal: "dan", "yang").
Text Normalization	:	Proses merapikan teks dari bentuk tidak baku menjadi bentuk standar, seperti menghapus simbol atau angka.
Tokenization	:	Proses memecah teks menjadi unit terkecil seperti kata atau token.
Topic Modelling	:	Teknik untuk mengidentifikasi struktur tematik tersembunyi dari kumpulan dokumen teks.
TP *(True Positive)*	:	Data positif yang berhasil diklasifikasikan dengan benar sebagai positif oleh model.
TN *(True Negative)*	:	Data negatif yang berhasil diklasifikasikan dengan benar sebagai negatif oleh model.
Tuning	:	Proses penyesuaian hyperparameter untuk mendapatkan performa model terbaik.
Val Accuracy	:	Akurasi yang diperoleh saat model diuji terhadap data validasi, digunakan untuk memantau overfitting.
Wordcloud	:	Visualisasi kata dalam bentuk awan di mana ukuran kata menunjukkan frekuensi kemunculannya.