

KLASIFIKASI KOMENTAR CYBERBULLYING DENGAN IBM GRANITE

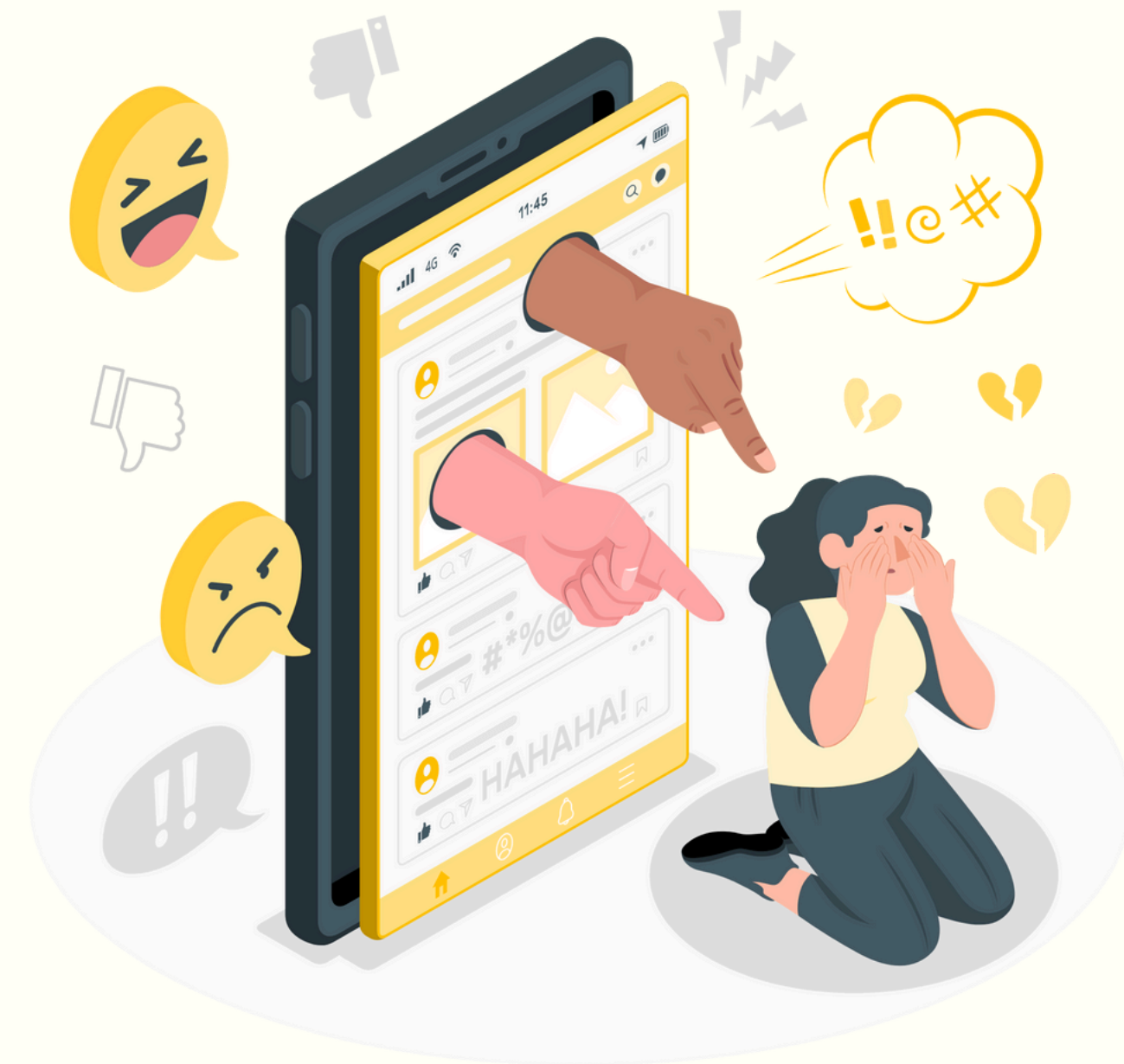
capstone project

Presented By Me

DATASET

cyberbullying-indonesia- dataset

Dataset ini merupakan kumpulan komentar dalam bahasa Indonesia yang dikumpulkan untuk keperluan analisis dan deteksi cyberbullying. Tujuan utama dari dataset ini adalah untuk menyediakan data autentik yang dapat digunakan dalam proyek machine learning, khususnya dalam klasifikasi dan pemrosesan bahasa alami (NLP).



PROJECT OVERVIEW

Latar Belakang

Cyberbullying di media sosial semakin sering terjadi, membutuhkan solusi otomatis untuk mendeteksi dan merangkum komentar yang berpotensi cyberbullying.

Tujuan Project

Mengklasifikasikan komentar media sosial sebagai "Bullying" atau "Bukan Bullying" dan merangkum komentar tersebut menggunakan model LLM (IBM Granite).

Pendekatan

IBM Granite digunakan untuk memahami dan memproses teks agar bisa secara akurat mengklasifikasi apakah komentar termasuk bullying atau bukan, serta merangkum poin utama komentar tersebut.

ANALYSIS PROCESS

1. Persiapan dan Konfigurasi

Lingkungan kerja disiapkan dengan menginstal library yang diperlukan seperti langchain_community, replicate, dan pandas. Model IBM Granite dikonfigurasi menggunakan API Key dari Replicate untuk mendukung tugas klasifikasi dan peringkasan.

2. Pengolahan dan Pembersihan Data

Dataset komentar dibaca dari file CSV, kemudian diproses untuk mengatasi encoding error. Langkah ini memastikan data bersih dan siap dianalisis oleh model IBM Granite.

3. Penerapan Model

Model dijalankan pada lima komentar teratas untuk menghasilkan ringkasan kalimat pendek dan klasifikasi "Bullying" atau "Bukan Bullying". Hasil tersebut disusun dalam bentuk DataFrame pandas agar mudah dianalisis dan divisualisasikan.

4. Evaluasi dan Pengujian Model

Performa model dievaluasi dengan membandingkan hasil klasifikasi terhadap label asli menggunakan akurasi, confusion matrix, dan classification report. Pengujian lanjutan dilakukan pada komentar baru untuk menunjukkan kemampuan model dalam skenario nyata.

INSIGHT DATA



Jumlah Data

Dataset terdiri dari 650 komentar, yang siap digunakan untuk pelatihan dan evaluasi model klasifikasi.





Kategori Data

Jumlah komentar "Bullying" dan "Non-bullying" seimbang — masing-masing 325 komentar, sehingga cocok untuk supervised learning.



Visualisasi Data

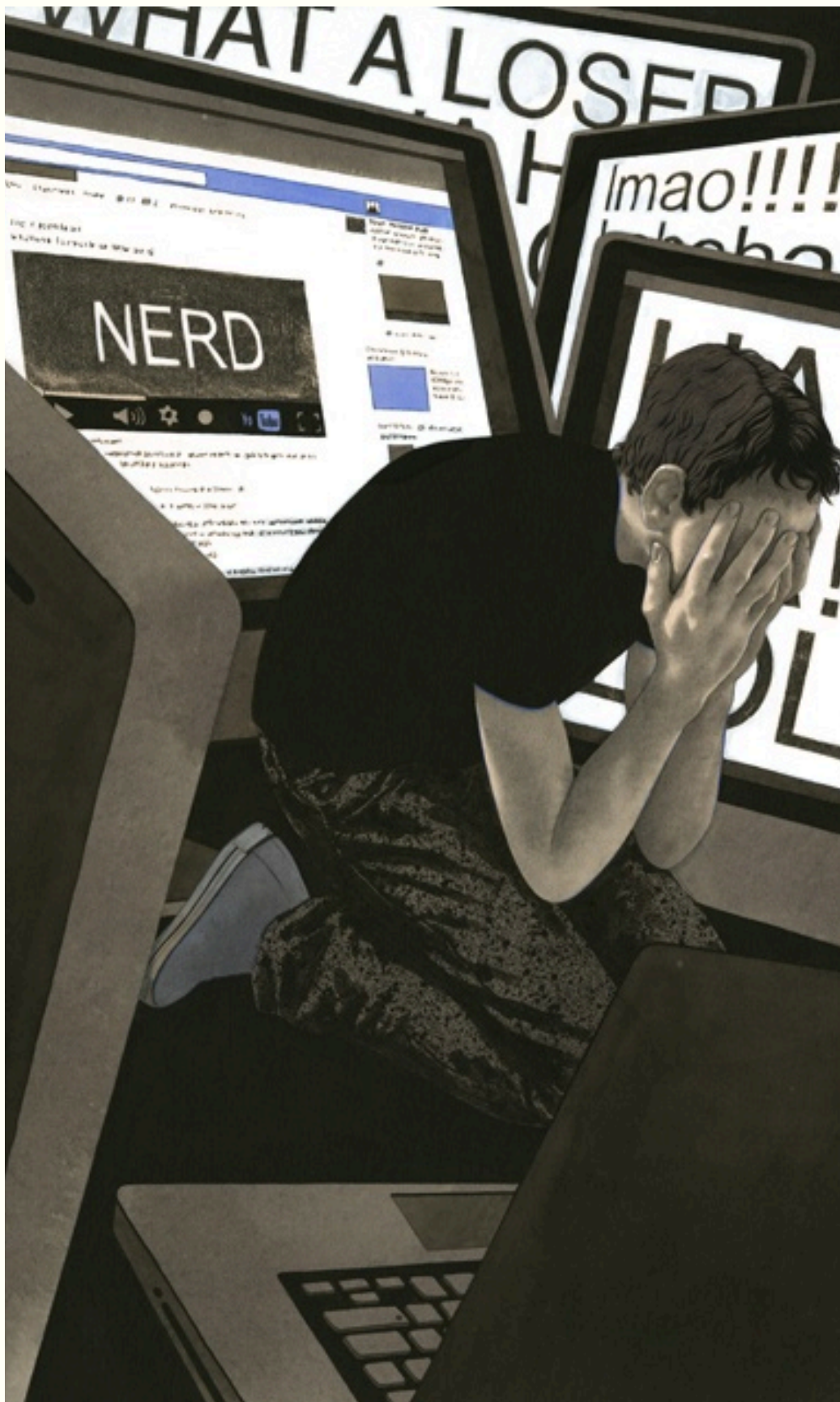
 Grafik distribusi kategori: Menunjukkan keseimbangan jumlah komentar.

 Grafik rata-rata panjang komentar: Menggambarkan perbedaan panjang teks antar kategori secara visual.

CONCLUSION

Proyek ini berhasil mendemonstrasikan penggunaan model IBM Granite untuk merangkum dan mengklasifikasikan komentar media sosial terkait cyberbullying. Meskipun hasil awal menunjukkan kemampuan model dalam mengidentifikasi komentar "Bullying", evaluasi terbatas pada lima komentar menunjukkan masih adanya kesalahan klasifikasi. Dataset yang digunakan memiliki distribusi kategori yang seimbang dan perbedaan panjang komentar antar kategori, yang dapat dimanfaatkan lebih lanjut dalam pengembangan model klasifikasi.





RECOMMENDATION

Untuk meningkatkan akurasi dan efektivitas sistem, disarankan melakukan evaluasi penuh pada seluruh dataset, menganalisis kesalahan klasifikasi, dan mengeksplorasi teknik prompt engineering atau fine-tuning. Selain itu, membandingkan performa IBM Granite dengan model lain serta mempertimbangkan pendekatan klasifikasi tradisional berbasis fitur teks dapat membantu membangun sistem deteksi cyberbullying yang lebih handal dan efisien, terutama untuk penggunaan dalam skenario dunia nyata.

AI SUPPORT

Klasifikasi & Peringkasan Otomatis

AI secara otomatis mengklasifikasikan komentar dan merangkumnya menjadi kalimat singkat, menggantikan proses manual yang lambat.

Efisiensi & Skalabilitas

Otomatisasi memungkinkan analisis data dalam jumlah besar secara cepat, cocok untuk skala media sosial.

Alat Bantu Analisis

AI mendukung proses analisis, namun tetap membutuhkan interpretasi manusia untuk hasil yang akurat dan kontekstual.

THANK YOU

Thanks For Watching

Presented By Me