

RISET INFORMATIKA

PENGEMBANGAN MODEL NLP UNTUK ANALISIS SENTIMEN BERBASIS EMOSI PADA TEKS: STUDI KASUS BAHASA INDONESIA

Rangga Widiasmara

NPM : 21081010085

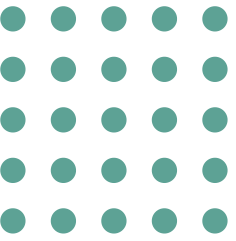
PROGRAM STUDI INFORMATIKA

Latar Belakang

Analisis sentimen tidak hanya mencakup klasifikasi polaritas (positif, negatif, netral), tetapi juga emosi mendalam seperti marah, senang, sedih, dll. Dalam Natural Language Processing (NLP), analisis emosi dapat digunakan di berbagai bidang seperti:

- Layanan Pelanggan: Untuk menganalisis kepuasan pelanggan.
- Kesehatan Mental: Untuk mendeteksi potensi masalah emosional dari teks pasien.
- Sosial Media: Untuk memahami dinamika emosi pengguna secara real-time.

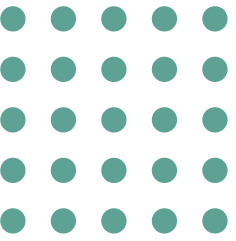
Research Gap



01. Pendekatan NLP tradisional masih kurang optimal untuk menangkap konteks dalam bahasa non-Inggris seperti Bahasa Indonesia.
02. Kurangnya kombinasi metode dalam menangkap pola semantik dan temporal dari teks untuk analisis emosi.
03. Kebanyakan penelitian berhenti pada tahap deteksi emosi tanpa menjelaskan bagaimana hasil analisis emosi dapat diintegrasikan secara langsung ke dalam proses pengambilan keputusan.



Rumusan Masalah



01. Apa metode terbaik untuk menangkap konteks semantik dan temporal dalam bahasa alami menggunakan dataset Bahasa Indonesia?
02. Apa metrik evaluasi terbaik untuk membandingkan model berbasis deep learning dengan pendekatan tradisional dalam analisis emosi pada teks?
03. Apa tantangan utama dalam implementasi analisis emosi otomatis untuk mendukung pengambilan keputusan di bidang tertentu



Mind Mapping

01.

Analisis Sentimen Berbasis Emosi:

- Referensi: Penelitian yang memanfaatkan LSTM, CNN, dan transformer untuk emosi.
- Gap: Fokus lebih pada bahasa Inggris; kurang pada Bahasa Indonesia.

02.

Teknologi NLP:

- Embedding semantik: Word2Vec, GloVe, FastText.
- Model: LSTM, CNN, Transformer (BERT, RoBERTa).

03.

Dataset:

- Dataset emosi (baik publik maupun buatan sendiri).
- Tantangan: Ukuran dataset yang kecil untuk bahasa lokal.

Metode yang Diusulkan



Pra-Pemrosesan Teks

- Tokenisasi, stemming, dan stopword removal.
- Representasi teks menggunakan word2vec atau FastText.



Arsitektur Model NLP

- Pendekatan Tradisional: SVM atau Naïve Bayes untuk baseline.
- Pendekatan Lanjutan:
 - CNN: Untuk ekstraksi fitur spasial teks.
 - LSTM: Untuk menangkap pola temporal.



Data

- Dataset teks emosi dalam Bahasa Indonesia.
- Penggunaan teknik augmentasi data untuk meningkatkan kualitas dataset kecil.



Metrik Pengujian

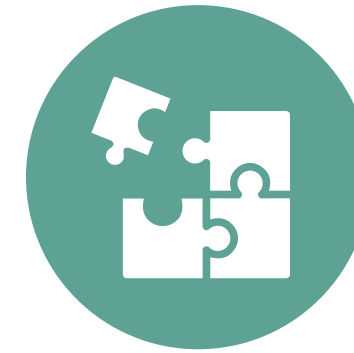
- Accuracy: Evaluasi keseluruhan performa model.
- Precision, Recall, F1-Score: Untuk menilai performa pada masing-masing kategori emosi.

Progres Riset



Pengumpulan Dataset

- Dataset yang terkumpul sekitar 3000 data dengan 5 emosi
- Sedang melakukan Augmentasi data agar dataset bervariasi



Implementasi Model

- Masih Beres eksperimen dengan arsitektur LSTM jika dipadukan dengan CNN



Terima Kasih