

# MULTI-LABEL EMOTION CLASSIFICATION

on Text for the Shared Task of Emotion Recognition

INFO 557 — Fall 2025 · Shared Task Experiment · Codabench 11263

**Anggota Kelompok:** Herlina (25/571614/SPA/01154) · Findra Kartika Sari Dewi (25/571659/SPA/01172) · Marwan Ramdhany Edy (25/571637/SPA/01167) · Program Studi Doktor Ilmu Komputer, Fakultas MIPA, Universitas Gadjah Mada



UNIVERSITAS  
GADJAH MADA

LOCALLY ROOTED, GLOBALLY RESPECTED

## Shared Task Overview

ER

**Tugas:** mengklasifikasikan **7 emosi** (admiration, amusement, gratitude, love, pride, relief, remorse) yang dapat muncul **lebih dari satu** pada satu teks.

**Evaluasi:** Micro F1 Score pada data dev & test di platform Codabench.

**Fokus:** membandingkan berbagai arsitektur model NLP berbasis Transformer dan jaringan konvolusi.

## Task Definition & Examples

### Contoh:

**Input:** "Thanks for the reply! I appreciate your input. Please keep me in the loop, I'd love to be more active with this if possible."

**Output:** admiration, gratitude, love

### Label emosi:

- Admiration
- Amusement
- Gratitude
- Love
- Pride
- Relief
- Remorse

**Metrik:** Micro F1 Score.

## Experimental Setup

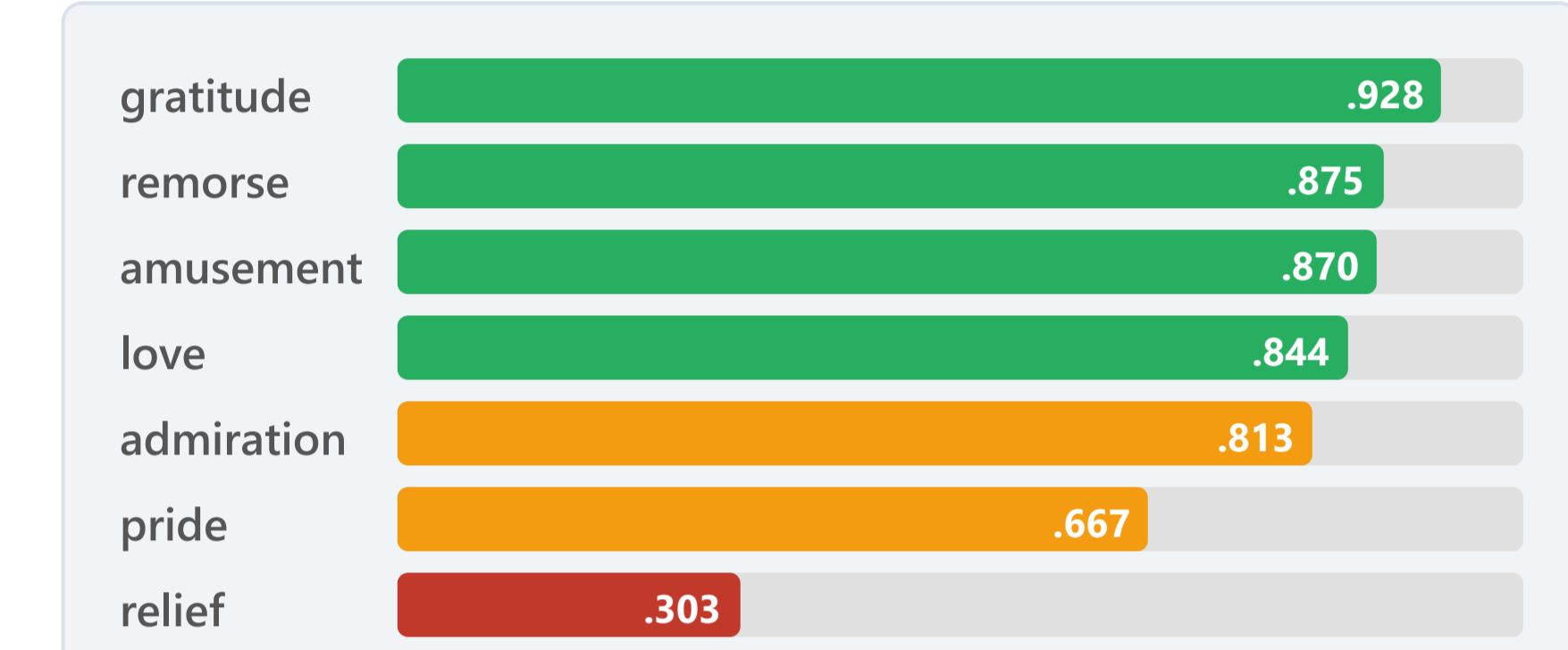
Dalam pengerjaan task ini dilakukan evaluasi berbagai arsitektur model NLP berbasis Transformer dan jaringan konvolusi:

- **DistilBERT:** Versi ringkas dari BERT, lebih cepat dan ringan.
- **DistilRoBERTa:** Versi ringkas dari RoBERTa, optimasi pre-training.
- **ALBERT:** Parameter BERT yang dikurangi (Lite BERT).
- **TextCNN:** baseline konvolusional klasik untuk klasifikasi teks.

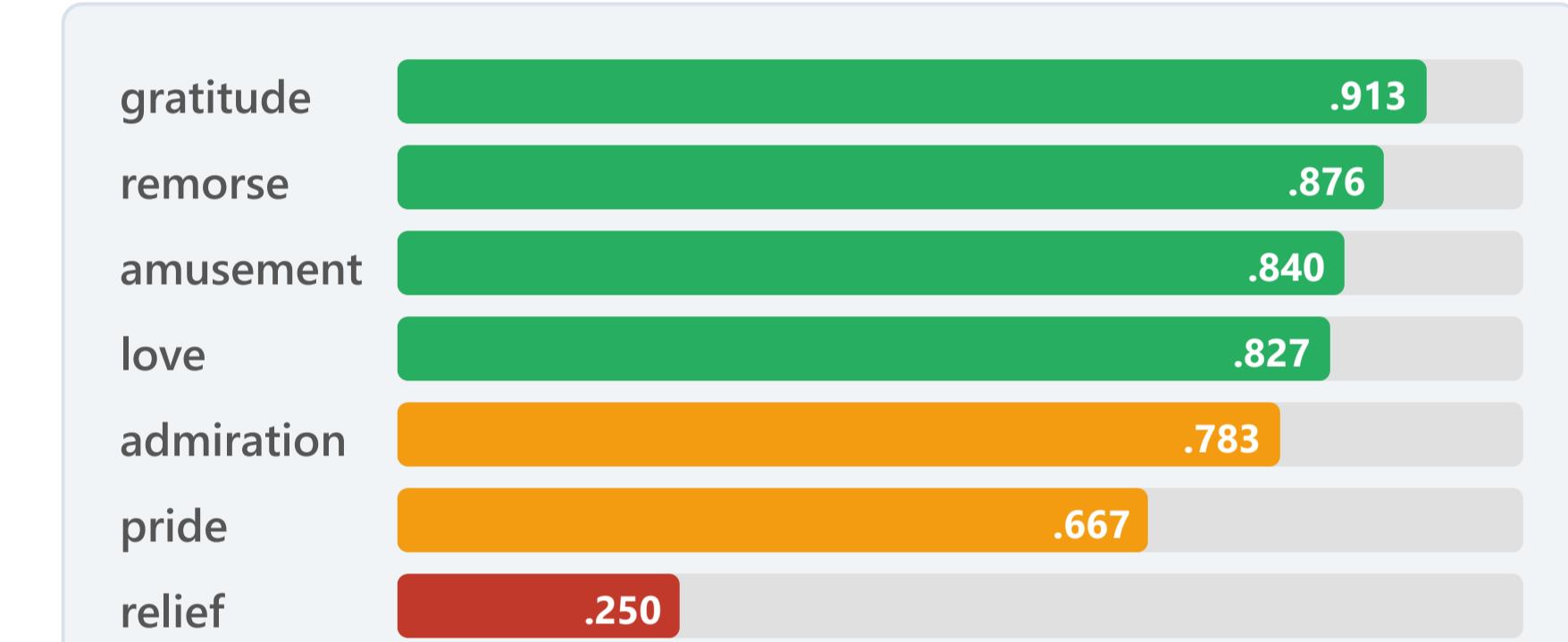
Pelatihan model dilakukan menggunakan data dari Shared Task of Emotion Recognition (INFO 557 Fall 2025) dengan Hyperparameter standar untuk klasifikasi multi-label.

## Top 2 Results

### 1. DistilBERT



### 2. DistilRoBERTa



## Error Analysis (DistilBERT)

Visualisasi performa per label (TP = True Positive, FP = False Positive, FN = False Negative).

### Admiration

**TP: 401**

FP: 97, FN: 87

### Gratitude

**TP: 323**

FP: 15, FN: 35

### Amusement

**TP: 275**

FP: 54, FN: 28

### Love

**TP: 231**

FP: 66, FN: 21

### Remorse

**TP: 63**

FP: 13, FN: 5

### Pride (Weak)

**TP: 8**

FP: 1, FN: 7

### Relief (Weak)

**TP: 5**

FP: 10, FN: 13

### Insight Utama:

- **Performa Stabil pada Label Dominan:** Mayoritas label (admiration, amusement, gratitude, love, remorse) menunjukkan kinerja yang konsisten dan kuat, didukung oleh ciri leksikal yang eksplisit pada data teks.
- **Kendala Ketidakseimbangan Data:** Dua label minoritas, *pride* dan *relief*, memiliki skor yang jauh lebih rendah ( $F1 < 0.35$ ). Hal ini mengindikasikan masalah ketidakseimbangan kelas (class imbalance) yang parah, di mana model gagal mempelajari pola karena kurangnya contoh positif.
- **Temuan Kesalahan Spesifik:**
  - *Admiration* cenderung memiliki tingkat *False Positive* yang tinggi (97), menunjukkan model terlalu agresif memprediksi label ini pada kalimat positif umum.
  - *Relief* memiliki rasio *False Negative* yang tinggi (13) dibanding *True Positive* (5), menandakan model sering gagal mendeteksi keberadaan emosi ini (missed detection).

## Leaderboard & Comparison

| # | Model / Participant  | Micro F1    | Adm         | Amuse       | Grat        | Love        | Pride       | Relief      | Remorse     |
|---|----------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1 | DistilBERT (Ours)    | <b>0.85</b> | <b>0.81</b> | <b>0.87</b> | <b>0.93</b> | <b>0.84</b> | <b>0.67</b> | <b>0.30</b> | <b>0.88</b> |
| 2 | grenganathan (Top 1) | 0.84        | 0.80        | 0.86        | 0.92        | 0.83        | 0.50        | 0.29        | 0.90        |
| 3 | DistilRoBERTa (Ours) | 0.83        | 0.78        | 0.83        | 0.91        | 0.82        | 0.61        | 0.15        | 0.84        |
| 4 | TextCNN (Ours)       | 0.74        | 0.60        | 0.80        | 0.93        | 0.81        | 0.00        | 0.00        | 0.83        |
| 5 | ALBERT (Ours)        | 0.69        | 0.60        | 0.70        | 0.84        | 0.66        | 0.125       | 0.00        | 0.65        |

**Analisis Kompetitif:** Model **DistilBERT (Ours)** berhasil mengungguli performa partisipan teratas (*grenganathan*) dengan selisih Micro F1 sebesar **+0.01**. Meskipun skor per label relatif seimbang, DistilBERT unggul signifikan dalam mendeteksi *Pride* (0.67 vs 0.50), menunjukkan kemampuan generalisasi yang lebih baik pada data minoritas.

## Conclusion & Future Work

- DistilBERT dan DistilRoBERTa menunjukkan performa superior dibandingkan model baseline konvolusional (TextCNN), membuktikan efektivitas *attention mechanism* untuk teks pendek.
- Performa tinggi pada label mayoritas kontras dengan hasil rendah pada label minoritas (Relief & Pride), menegaskan bahwa kuantitas data adalah hambatan utama.

### Future Works

- **Advanced Data Augmentation:** Menggunakan *Back-translation* atau generasi data sintetik berbasis LLM (seperti GPT) khusus untuk kelas *Pride* dan *Relief*.
- **Loss Function Optimization:** Menerapkan *Focal Loss* atau *Class-Weighted Cross Entropy* untuk memberikan penalti lebih besar pada kesalahan klasifikasi kelas minoritas.
- **Threshold Tuning:** Mengoptimalkan ambang batas probabilitas (threshold) yang berbeda untuk setiap label, alih-alih menggunakan threshold tunggal 0.5.