

Dokumentasi Awal Penelitian

Date: 06/10/2025

1. Judul Penelitian

Peningkatan Model Analisis Sentimen Berbasis Aspek terhadap Tokoh Politik Indonesia Menggunakan Transformer dan Data Pemilu 2024.

2. Latar Belakang

Analisis sentimen berbasis aspek (Aspect-Based Sentiment Analysis / ABSA) merupakan pendekatan yang lebih mendalam dibanding analisis sentimen umum karena tidak hanya mengidentifikasi polaritas sentimen, tetapi juga menentukan aspek yang menjadi sasaran opini. Dalam konteks politik Indonesia, opini publik di media sosial menjadi sumber informasi penting yang dapat menggambarkan persepsi masyarakat terhadap tokoh politik. Penelitian sebelumnya menggunakan LSTM namun masih terbatas karena ukuran dataset kecil dan distribusi aspek tidak seimbang. Penelitian ini bertujuan untuk meningkatkan performa ABSA dengan memanfaatkan model Transformer seperti IndoBERT.

3. Tujuan Penelitian

- a. Mengembangkan model deteksi aspek berbasis Transformer untuk teks opini politik Indonesia.
- b. Meningkatkan akurasi klasifikasi sentimen pada tiap aspek dengan fine-tuning model bahasa pra-latih.
- c. Membandingkan performa model Transformer dengan baseline TF-IDF dan LSTM.
- d. Menganalisis pengaruh ketidakseimbangan data terhadap performa model ABSA.

4. Dataset Penelitian

Dataset berisi teks opini masyarakat terhadap tokoh politik Indonesia dari periode Pemilu 2014 dan sebagian diperbarui untuk Pemilu 2024. Dataset mencakup anotasi aspek dan sentimen untuk setiap teks.

5. Analisis Awal Dataset

Hasil eksplorasi awal menunjukkan dataset sangat tidak seimbang. Aspek integritas dan kapabilitas mendominasi data, sedangkan aspek empati dan kontinuitas sangat jarang muncul. Kondisi ini berpotensi menurunkan akurasi model pada aspek minor.

6. Permasalahan Penelitian

- a. Bagaimana meningkatkan performa deteksi aspek pada dataset yang imbalanced?

- b. Bagaimana model Transformer dapat meningkatkan akurasi klasifikasi sentimen per aspek?
- c. Bagaimana membangun pipeline ABSA yang efisien untuk teks opini politik Indonesia?

7. Rancangan Pipeline Penelitian

Pipeline penelitian terdiri dari tahapan berikut:

a. Tahap 1 – Data Preparation

- Gabungkan dataset aspek dan sentimen berdasarkan kolom content.
- Bersihkan teks dari URL, mention, tagar, dan karakter khusus.
- Normalisasi teks dan ubah ke huruf kecil.
- Bagi data menjadi train-validation-test (80:10:10).

b. Tahap 2 – Eksperimen Baseline

- Representasi teks menggunakan **TF-IDF**.
- Klasifikasi multi-label untuk aspek menggunakan **One-vs-Rest Logistic Regression**.
- Klasifikasi sentimen per aspek menggunakan model **Logistic Regression (binary)**.
- Evaluasi menggunakan *precision*, *recall*, dan *F1-score*.

c. Tahap 3 – Penerapan Transformer

- Gunakan model **IndoBERT-base-p1** sebagai model dasar.
- Fine-tune model untuk:
 - Multi-label classification (deteksi aspek)
 - Binary classification (sentimen per aspek)
- Gunakan *BCEWithLogitsLoss* dengan *pos_weight* untuk menangani ketidakseimbangan label.

d. Tahap 4 – Evaluasi dan Analisis

- Bandingkan performa baseline dengan model Transformer.
- Gunakan metrik:
 - **Micro-F1** dan **Macro-F1** untuk deteksi aspek.
 - **F1-score per aspek** untuk sentimen.
- Lakukan *error analysis* terhadap hasil prediksi yang salah.
- Analisis dampak *data imbalance* terhadap performa model.

e. Tahap 5 – Strategi Peningkatan

- Eksperimen *oversampling* dan *text augmentation* untuk minor class.
- Tuning threshold per label untuk optimalisasi F1.
- (Opsional) Multi-task learning: deteksi aspek dan sentimen dalam satu model.

8. Inisialisasi Project dan Kolaborasi

a. *Setup Repository GitHub*

Inisialisasi repositori GitHub dengan nama:

ABSA-Transformer-Pemilu2024

Tautan: <https://github.com/luthfilkhairi/ABSA-Transformer-Pemilu2024>

Langkah-langkah yang dilakukan:

1. Membuat repositori di GitHub dan menghubungkannya dengan project lokal pada VS Code.
2. Mengatasi konflik awal antar cabang lokal dan remote dengan git pull origin main --allow-unrelated-histories dan penyelesaian konflik pada README.md.
3. Melakukan push awal (initial commit) sehingga project tersinkron antara lokal dan remote repository.

b. *Struktur Folder Awal*

Struktur direktori awal project disusun untuk mendukung alur penelitian **Aspect-Based Sentiment Analysis (ABSA)** berbasis Transformer, sebagai berikut:

```
ABSA-Transformer-Pemilu2024/
|
└── data/                      # Dataset mentah dan hasil preprocessing
└── notebooks/                 # Eksperimen Jupyter Notebook
└── src/                        # Kode utama (preprocessing, modeling, evaluasi)
└── reports/                   # Dokumentasi dan hasil eksperimen
└── results/                    # Output model dan metrik evaluasi
└── README.md
└── requirements.txt
```

c. *Setup Kolaborasi Tim*

Tim penelitian terdiri dari:

1. Luthfil Khairi, S.Kom., M.Cs. (Ketua)
2. Ajeng Dwi Asti, M.Kom. (Rekan Peneliti)
3. Muhammad Zaki Al Hafiz (Asisten Riset)

Rencana kolaborasi di GitHub:

1. Penggunaan branch individu (dev-khairi, dev-ajeng, dev-zaki)
2. Branch utama main untuk hasil akhir yang sudah diverifikasi
3. Kolaborasi melalui fitur *Pull Request* untuk review perubahan

9. Rencana Jadwal Penelitian

Minggu	Periode	Kegiatan Utama	Target Capaian
--------	---------	----------------	----------------

1	16–22 Sept	Pembersihan dan eksplorasi dataset	Dataset siap digunakan
2	23–29 Sept	Pembuatan baseline TF-IDF + Logistic Regression	Baseline performa awal
3	30 Sept–6 Okt	Fine-tuning model Transformer (deteksi aspek)	Model aspek valid
4	7–13 Okt	Fine-tuning model Transformer (sentimen)	Model sentimen valid
5	14–20 Okt	Evaluasi, perbandingan model, dan visualisasi hasil	Nilai evaluasi & grafik performa
6	21–27 Okt	Error analysis & peningkatan model	Perbaikan akurasi minor aspect
7	28 Okt–3 Nov	Dokumentasi hasil & laporan kemajuan	Draft laporan selesai
8	4–14 Nov	Finalisasi laporan & presentasi	Siap untuk pelaporan 14 Nov

10. Luaran yang Diharapkan

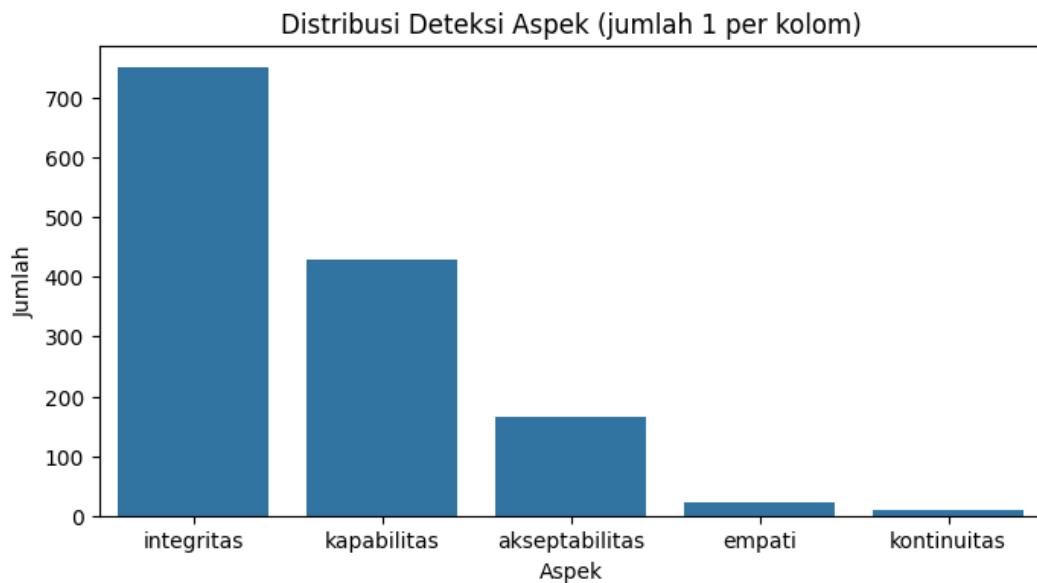
- a. Model ABSA berbasis Transformer dengan performa lebih baik.
- b. Analisis pengaruh ketidakseimbangan data.
- c. Dataset bersih dan terdokumentasi.
- d. Visualisasi hasil evaluasi dan error analysis.

11. Dokumentasi Tambahan

- a. **Tanggal Mulai:** 16 September 2025
- b. **Tanggal Pelaporan Tengah:** 14 November 2025
- c. **Progress Saat Ini:** Eksplorasi dataset, identifikasi imbalance, dan perancangan pipeline penelitian.
- d. **Persentase Capaian:** ±15%
- e. **Dokumen Pendukung:**
 - Screenshot grafik distribusi aspek
 - Plot distribusi sentimen per aspek
 - Notebook eksplorasi dataset (EDA)

Dokumen Pendukung

1. Screenshot grafik distribusi aspek



2. Plot distribusi sentimen per aspek

Distribusi Sentimen per Aspek (Angka):

a. Aspek: integritas

Integritas	
-1	594
1	408
0	343

b. Aspek: kapabilitas

Kapabilitas	
-1	917
1	233
0	195

c. Aspek: empati

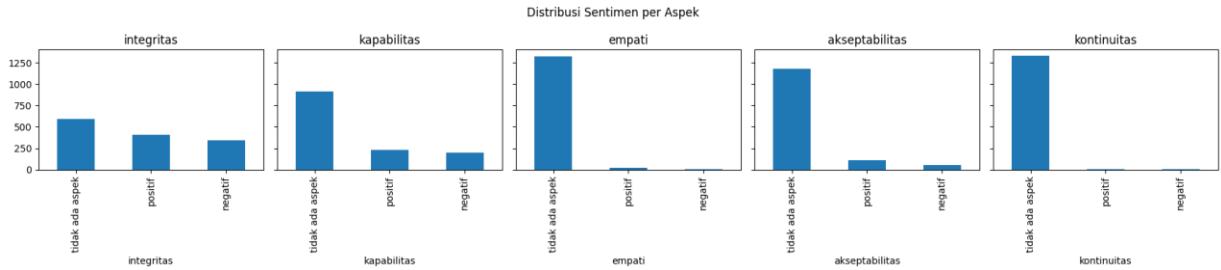
Kapabilitas	
-1	917
1	233
0	195

d. Aspek: akseptabilitas

Akseptabilitas	
-1	1180
1	110
0	55

e. Aspek: kontinuitas

Kontinuitas	
-1	1336
1	6
0	3



3. Notebook eksplorasi dataset (EDA)

Dataset Aspek: (1345, 6)

Dataset Sentimen: (1345, 6)

Apakah semua baris content sama dan urut? **False**

Distribusi jumlah aspek terdeteksi:

- a. integritas 750
- b. kapabilitas 428
- c. akseptabilitas 165
- d. empati 23
- e. kontinuitas 9

Kesimpulan Sementara:

- a. Jumlah total data: 1345
- b. Rata-rata aspek per teks: 1.0223048327137547
- c. Aspek paling sering muncul: integritas
- d. Aspek paling jarang muncul: kontinuitas