Dokumentasi Awal Penelitian

***Date: 06/10/2025***

# Judul Penelitian

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

# 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.

# Tujuan Penelitian

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

# 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.

# 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.

# Permasalahan Penelitian

1. Bagaimana meningkatkan performa deteksi aspek pada dataset yang imbalanced?
2. Bagaimana model Transformer dapat meningkatkan akurasi klasifikasi sentimen per aspek?
3. Bagaimana membangun pipeline ABSA yang efisien untuk teks opini politik Indonesia?

# Rancangan Pipeline Penelitian

Pipeline penelitian terdiri dari tahapan berikut:

1. **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).

#### **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*.

#### **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.

#### **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.

#### **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.

# 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 |

# Luaran yang Diharapkan

1. Model ABSA berbasis Transformer dengan performa lebih baik.
2. Analisis pengaruh ketidakseimbangan data.
3. Dataset bersih dan terdokumentasi.
4. Visualisasi hasil evaluasi dan error analysis.

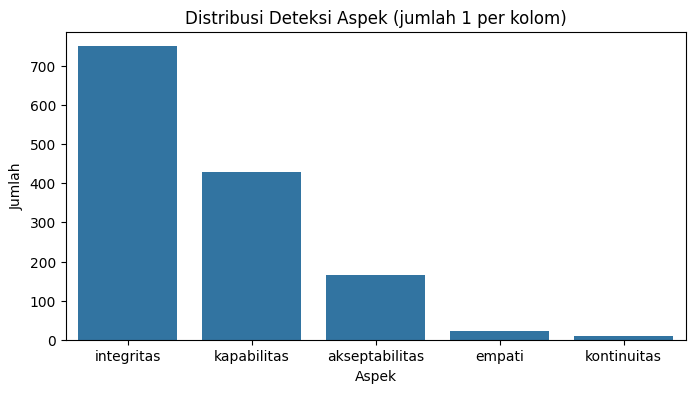
# Dokumentasi Tambahan

1. **Tanggal Mulai:** 16 September 2025
2. **Tanggal Pelaporan Tengah:** 14 November 2025
3. **Progress Saat Ini:** Eksplorasi dataset, identifikasi imbalance, dan perancangan pipeline penelitian.
4. **Persentase Capaian:** ±15%
5. **Dokumen Pendukung:**

* Screenshot grafik distribusi aspek
* Plot distribusi sentimen per aspek
* Notebook eksplorasi dataset (EDA)

# Dokumen Pendukung

1. Screenshot grafik distribusi aspek



1. Plot distribusi sentimen per aspek

Distribusi Sentimen per Aspek (Angka):

1. Aspek: integritas

|  |  |
| --- | --- |
| **Integritas** |  |
| -1 | 594 |
| 1 | 408 |
| 0 | 343 |

1. Aspek: kapabilitas

|  |  |
| --- | --- |
| **Kapabilitas** |  |
| -1 | 917 |
| 1 | 233 |
| 0 | 195 |

1. Aspek: empati

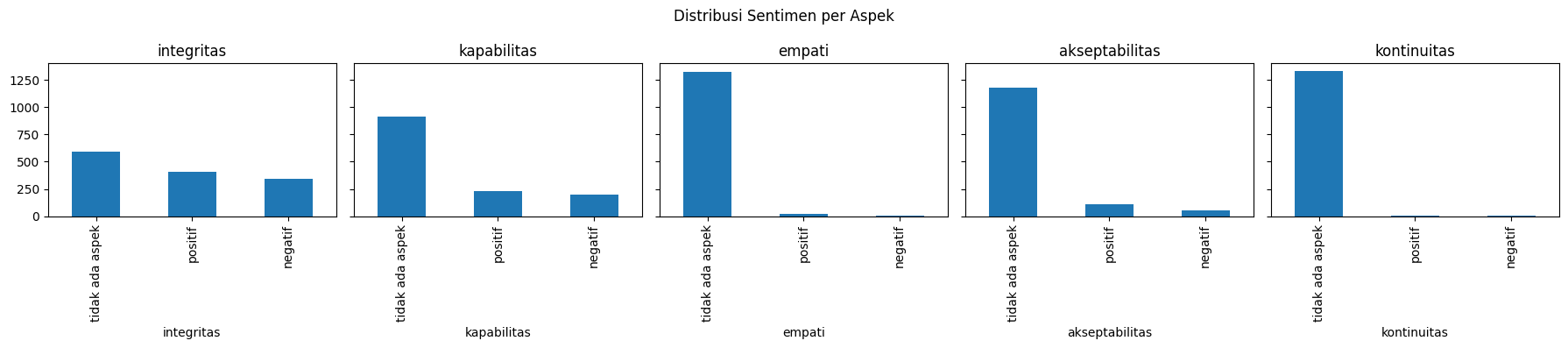
|  |  |
| --- | --- |
| **Kapabilitas** |  |
| -1 | 917 |
| 1 | 233 |
| 0 | 195 |

1. Aspek: akseptabilitas

|  |  |
| --- | --- |
| **Akseptabilitas** |  |
| -1 | 1180 |
| 1 | 110 |
| 0 | 55 |

1. Aspek: kontinuitas

|  |  |
| --- | --- |
| **Kontinuitas** |  |
| -1 | 1336 |
| 1 | 6 |
| 0 | 3 |



1. Notebook eksplorasi dataset (EDA)

Dataset Aspek: (1345, 6)

Dataset Sentimen: (1345, 6)

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

Distribusi jumlah aspek terdeteksi:

1. integritas 750
2. kapabilitas 428
3. akseptabilitas 165
4. empati 23
5. kontinuitas 9

**Kesimpulan Sementara:**

1. Jumlah total data: 1345
2. Rata-rata aspek per teks: 1.0223048327137547
3. Aspek paling sering muncul: integritas
4. Aspek paling jarang muncul: kontinuitas