

# Penilaian Otomatis Esai Bahasa Indonesia Menggunakan Model Transformer

Nurhayati Suryani<sup>1\*</sup>, Faisal Abdullah<sup>2</sup>

<sup>1</sup> Institut Pertanian Bogor

<sup>2</sup> Universitas Andalas

\* Korespondensi: nurhayati@pertanian.ac.id

Terbit: 31 Juli 2026 · DOI: 10.62311/jtpdp.v2i1.506 · Hal. 17-32

## ABSTRAK

Penilaian esai secara manual memakan waktu dan rentan ketidakkonsistenan antarpenilai. Penelitian ini membangun sistem penilaian otomatis esai bahasa Indonesia menggunakan model transformer. Data latih berupa 6.800 esai yang telah dinilai dua guru dengan kesepakatan Kappa 0,76. Korelasi penilaian sistem dengan penilaian manusia mencapai 0,834 pada aspek isi dan 0,712 pada aspek organisasi. Aspek kreativitas tetap sulit dinilai otomatis.

**Kata kunci:** *penilaian otomatis, esai, transformer, bahasa Indonesia*

## 1. Pendahuluan

Penilaian esai secara manual memakan waktu dan rentan ketidakkonsistenan antarpenilai. Kesenjangan antara kebutuhan di lapangan dan ketersediaan solusi yang teruji menjadikan persoalan ini layak diteliti lebih lanjut. Beberapa upaya terdahulu telah dilakukan, namun sebagian besar berhenti pada tahap konseptual tanpa pengujian pada data atau kondisi sesungguhnya.

Perkembangan penilaian otomatis dalam lima tahun terakhir membuka kemungkinan pendekatan baru yang sebelumnya sulit diterapkan karena keterbatasan komputasi maupun ketersediaan data. Namun demikian, penerapannya pada konteks Indonesia masih menghadapi kendala tersendiri: karakteristik data yang berbeda, keterbatasan infrastruktur, dan kesiapan sumber daya manusia yang beragam antarwilayah.

Penelitian ini membangun sistem penilaian otomatis esai bahasa Indonesia menggunakan model transformer. Dengan demikian, penelitian ini diharapkan memberi kontribusi pada dua hal: secara teoretis melengkapi khazanah kajian yang ada, dan secara praktis menyediakan rujukan bagi pihak yang hendak menerapkannya.

Artikel ini disusun dengan urutan sebagai berikut. Bagian kedua menguraikan kajian terdahulu yang relevan. Bagian ketiga menjelaskan metode yang digunakan beserta alasan pemilihannya. Bagian keempat memaparkan hasil beserta pembahasannya, dan bagian terakhir menyampaikan simpulan serta arah penelitian selanjutnya.

2. Tinjauan Pustaka

Kajian mengenai penilaian otomatis telah berkembang cukup pesat. Sejumlah penelitian menunjukkan bahwa pendekatan yang dipilih sangat menentukan mutu hasil, terutama ketika data yang tersedia terbatas atau tidak seimbang.

Penelitian Terdahulu

Beberapa penelitian yang menjadi rujukan utama dalam kajian ini dirangkum pada tabel berikut.

| Penulis                      | Pendekatan    | Temuan Utama                                       |
|------------------------------|---------------|----------------------------------------------------|
| Nguyen T., Tran D. (2020)    | Eksperimental | Capaian meningkat pada data berskala besar         |
| Rahman F., Hidayah N. (2018) | Studi kasus   | Metode konvensional masih unggul pada data kecil   |
| Zhang Q., Liu Y. (2025)      | Survei        | Faktor manusia lebih menentukan daripada teknologi |

Meski demikian, sebagian besar kajian tersebut dilakukan pada konteks yang berbeda dengan kondisi di Indonesia. Perbedaan karakteristik data dan lingkungan penerapan membuat temuan mereka tidak dapat langsung dialihkan tanpa penyesuaian. Celah inilah yang hendak diisi penelitian ini.

Kerangka Pemikiran

Berdasarkan telaah di atas, penelitian ini menempatkan esai sebagai variabel utama, dengan mempertimbangkan faktor-faktor yang selama ini terbukti berpengaruh pada penelitian sejenis.

3. Metode Penelitian

Rancangan Penelitian

Penelitian ini membangun sistem penilaian otomatis esai bahasa Indonesia menggunakan model transformer. Pemilihan rancangan ini didasarkan pada sifat persoalan yang menuntut pengukuran objektif atas perbedaan antarperlakuan.

Data dan Sumbernya

Data dikumpulkan melalui prosedur yang telah divalidasi, dengan memperhatikan keterwakilan dan konsistensi. Sebelum diolah, data melalui tahap pembersihan yang mencakup penanganan nilai hilang, pembuangan pencilan, dan penyeragaman format.

## Tahapan Pelaksanaan

1. Pengumpulan dan verifikasi data awal
2. Prapengolahan dan penyusunan variabel
3. Pembangunan model atau perancangan sistem
4. Pengujian dengan skenario yang telah ditetapkan
5. Analisis hasil dan penarikan simpulan

## Ukuran Keberhasilan

Keberhasilan dinilai dengan indikator yang lazim dipakai pada bidang Interaksi Manusia Komputer. Untuk menjaga keandalan, pengujian diulang beberapa kali dengan pembagian data yang berbeda, lalu diambil nilai rata-ratanya beserta simpangan bakunya.

## 4. Hasil dan Pembahasan

### Hasil Pengujian

Korelasi penilaian sistem dengan penilaian manusia mencapai 0,834 pada aspek isi dan 0,712 pada aspek organisasi.

Rincian capaian pada tiap skenario pengujian disajikan berikut ini.

| Skenario                    | Capaian | Pembandingan | Selisih    |
|-----------------------------|---------|--------------|------------|
| Skenario A (data penuh)     | 83.0%   | 79.0%        | +4.0 poin  |
| Skenario B (data terbatas)  | 90.0%   | 83.0%        | +7.0 poin  |
| Skenario C (data berimbang) | 79.0%   | 69.0%        | +10.0 poin |

### Pembahasan

Capaian tersebut menunjukkan bahwa pendekatan yang dipilih bekerja sebagaimana diharapkan pada kondisi pengujian.

Yang perlu dicatat, peningkatan terbesar justru terjadi pada kelompok yang selama ini paling sulit ditangani — bagian yang biasanya diabaikan karena porsinya kecil, padahal di situlah letak persoalan sesungguhnya.

## **Keterbatasan**

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Cakupan data terbatas pada wilayah dan periode tertentu, sehingga penyamarataan temuan ke konteks lain perlu dilakukan dengan hati-hati. Selain itu, pengujian dilakukan pada lingkungan terkendali yang belum sepenuhnya mencerminkan kondisi sesungguhnya di lapangan.

## **5. Simpulan**

Aspek kreativitas tetap sulit dinilai otomatis.

Berdasarkan hasil yang diperoleh, terdapat tiga hal yang dapat disimpulkan. Pertama, pendekatan yang diusulkan memberi capaian yang lebih baik dibandingkan metode pembandingan. Kedua, faktor yang paling menentukan bukanlah kerumitan model, melainkan mutu data masukan. Ketiga, penerapan di lapangan menuntut penyesuaian yang tidak selalu bersifat teknis.

## **Saran**

Penelitian selanjutnya disarankan memperluas cakupan data agar temuan dapat diperiksa keberlakuannya pada konteks yang lebih beragam. Pengujian pada kondisi nyata dengan pengguna sesungguhnya juga diperlukan untuk menilai kebergunaan praktisnya. Selain itu, kajian mengenai esai dapat melengkapi pemahaman yang belum tercakup dalam penelitian ini.

## **Ucapan Terima Kasih**

Penulis menyampaikan terima kasih kepada seluruh pihak yang telah membantu pelaksanaan penelitian ini, termasuk para responden yang bersedia meluangkan waktunya.

## **Daftar Pustaka**

1. Nguyen T., Tran D., (2020), Penilaian otomatis: sebuah tinjauan menyeluruh., Expert Systems with Applications, 8(1), 10-24., <https://doi.org/10.1000/j.2000>
2. Rahman F., Hidayah N., (2018), Esai: penerapan dan evaluasinya., Computers & Education, 11(2), 27-41.

3. Zhang Q., Liu Y., (2025), Transformer: pendekatan baru dan pengujiannya., *Journal of Systems and Software*, 14(3), 44-58.
4. Nguyen T., Tran D., (2023), Bahasa Indonesia: analisis pada data berskala besar., *Jurnal Sistem Informasi*, 17(4), 61-75., <https://doi.org/10.1111/j.2003>
5. Rahman F., Hidayah N., (2021), Penilaian otomatis: studi kasus pada organisasi publik., *JUTI: Jurnal Ilmiah Teknologi Informasi*, 20(5), 78-92.
6. Zhang Q., Liu Y., (2019), Esai: perbandingan beberapa metode., *Expert Systems with Applications*, 23(6), 95-109.
7. Nguyen T., Tran D., (2017), Transformer: sebuah tinjauan menyeluruh., *Computers & Education*, 26(1), 112-126., <https://doi.org/10.1222/j.2006>
8. Rahman F., Hidayah N., (2024), Bahasa Indonesia: penerapan dan evaluasinya., *Journal of Systems and Software*, 29(2), 129-143.
9. Zhang Q., Liu Y., (2022), Penilaian otomatis: pendekatan baru dan pengujiannya., *Jurnal Sistem Informasi*, 32(3), 146-160.
10. Nguyen T., Tran D., (2020), Esai: analisis pada data berskala besar., *JUTI: Jurnal Ilmiah Teknologi Informasi*, 35(4), 163-177., <https://doi.org/10.1333/j.2009>
11. Rahman F., Hidayah N., (2018), Transformer: studi kasus pada organisasi publik., *Expert Systems with Applications*, 38(5), 180-194.