

Evaluasi Model Distilbert dalam Prediksi Sentimen Ulasan Makanan Amazon

Diterima:

12 Juni 2025

Revisi:

19 Juli 2025

Terbit:

26 April 2026

**Jofan Vernanda Wicaksono, Ananda Bagus Fatchurroziq,
Muhammad Azzakafuadi, M Naufal Anjab Septyan**

Universitas Nusantara PGRI Kediri

Abstrak—Latar Belakang: Perkembangan teknologi digital telah menghasilkan ledakan data teks dari berbagai platform daring, terutama dalam bentuk ulasan produk. **Tujuan:** Tujuan penelitian ini adalah untuk mengevaluasi performa model DistilBERT dalam melakukan klasifikasi sentimen terhadap ulasan makanan dari dataset Amazon Fine Food Reviews. **Metode:** Dataset yang digunakan terdiri dari 100.000 data ulasan yang telah disederhanakan menjadi dua kelas sentimen: positif dan negatif. Proses pelatihan dilakukan menggunakan pustaka HuggingFace Transformers dengan konfigurasi model DistilBertForSequenceClassification. **Hasil:** Hasil evaluasi menunjukkan bahwa model DistilBERT mampu mencapai akurasi sebesar 87,44%, membuktikan efektivitasnya dalam menangkap konteks dan makna sentimen dengan efisien. **Kesimpulan:** Penelitian ini menegaskan bahwa model berbasis transformer ringan seperti DistilBERT dapat menjadi solusi optimal untuk analisis sentimen berskala besar dengan keterbatasan sumber daya komputasi.

Kata Kunci— Analisis Sentimen; Amazon; Ulasan Produk; Machine Learning; DistilBert

Abstract— Background: The rapid development of digital technology has led to an explosion of textual data from various online platforms, particularly in the form of product reviews. **Objective:** This study aims to evaluate the performance of the DistilBERT model in conducting sentiment classification on food reviews from the Amazon Fine Food Reviews dataset. **Methods:** The dataset used consists of 100,000 review samples that have been simplified into two sentiment classes: positive and negative. The training process was carried out using the HuggingFace Transformers library with the DistilBertForSequenceClassification model configuration. **Results:** The evaluation results show that the DistilBERT model achieved an accuracy of 87.44%, demonstrating its effectiveness in capturing contextual meaning and sentiment efficiently. **Conclusion:** This study confirms that lightweight transformer-based models such as DistilBERT can serve as an optimal solution for large-scale sentiment analysis under limited computational resources.

Keywords— Sentiment Analysis; Amazon; Product Reviews; Machine Learning; DistilBert

This is an open access article under the CC BY-SA License.



Penulis Korespondensi:

Jofan Vernanda Wicaksono,
Sistem Informasi,
Universitas Nusantara PGRI Kediri,
Email: vernandajofan@gmail.com

I. PENDAHULUAN

Perkembangan teknologi digital yang pesat telah menyebabkan peningkatan signifikan dalam jumlah data teks yang dihasilkan pengguna, khususnya melalui ulasan produk, komentar di media sosial, dan forum diskusi daring. Informasi dalam bentuk teks ini dapat dimanfaatkan untuk menganalisis opini, emosi, dan kepuasan konsumen terhadap suatu layanan atau produk, terutama pada platform e-commerce. Salah satu pendekatan yang umum digunakan adalah analisis sentimen, yakni teknik dalam bidang Natural Language Processing (NLP) yang bertujuan mengklasifikasikan opini pengguna ke dalam kategori sentimen tertentu, seperti positif, negatif, atau netral (Cheang et al., 2020).

Berbagai pendekatan telah digunakan dalam penelitian analisis sentimen, mulai dari metode statistik klasik hingga pendekatan deep learning. (Saidani et al., 2022) mengevaluasi kinerja algoritma Naïve Bayes, Decision Tree, dan Random Forest dalam mengklasifikasikan sentimen pada ulasan makanan berbahasa Indonesia. (Ristyan et al., 2025) Random Forest mempunyai kemampuan untuk memproses data berdimensi tinggi dan variabel yang hilang, Random forest memiliki keterbatasan berupa kesulitan dengan variabel kategorikal berkardinalitas tinggi, data tidak seimbang, peramalan deret waktu, dan interpretasi variable (Ristyan et al., 2025). Meskipun hasilnya cukup baik, metode tersebut masih memiliki keterbatasan dalam memahami konteks semantik dan sintaksis kalimat. Kelemahan ini kemudian dijawab oleh kemunculan arsitektur transformer, khususnya BERT (Bidirectional Encoder Representations from Transformers), yang mampu menangkap konteks dua arah dalam sebuah kalimat. Namun, BERT memiliki ukuran model besar yang kurang efisien secara komputasi (Sanh et al., 2019).

Sebagai solusi, DistilBERT dikembangkan sebagai versi ringan dari BERT dengan ukuran model 40% lebih kecil namun tetap mempertahankan sekitar 97% dari performa akurasi BERT asli. (Sanh et al., 2019) Model ini lebih cocok untuk skenario dengan keterbatasan sumber daya komputasi namun tetap memerlukan representasi teks kontekstual yang kuat (Sanh et al., 2019). Dalam konteks ini, digunakan dataset Amazon Fine Food Reviews, sebuah dataset benchmark populer yang berisi lebih dari 500.000 ulasan makanan dari pengguna Amazon, lengkap dengan rating dan metadata terkait (McAuley & Leskovec, 2013). Dataset ini sangat relevan untuk penelitian sentimen karena menyediakan data dalam skala besar, teks ulasan yang beragam, serta label rating yang bisa digunakan sebagai dasar klasifikasi sentimen.

Penelitian serupa yang dilakukan oleh (Sifa Amalia et al., 2021) menggunakan algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN) pada ulasan restoran dan mencapai akurasi tertinggi sebesar 81,92%. Selain itu, (Sari, 2019) menerapkan algoritma Naïve

Bayes pada data ulasan berbahasa Indonesia dan berhasil mengklasifikasikan sentimen dengan cukup akurat. Walaupun model-model tradisional ini masih digunakan, keterbatasannya dalam memahami konteks dan ketergantungan pada fitur manual menjadi alasan kuat untuk mengeksplorasi model transformer. Perbedaan penelitian ini dengan penelitian sebelumnya adalah penggunaan model DistilBERT yang merupakan versi ringan dari BERT untuk mengatasi keterbatasan sumber daya komputasi, serta fokus pada ulasan makanan dalam skala besar dari Amazon.

Dengan mempertimbangkan perkembangan tersebut, penelitian ini bertujuan untuk mengimplementasikan model DistilBERT untuk klasifikasi sentimen pada dataset Amazon Fine Food Reviews. Mengevaluasi performa akurasi dari DistilBERT dalam konteks pemrosesan ulasan pelanggan. Membandingkan efektivitas pendekatan berbasis transformer ini dengan pendekatan-pendekatan sebelumnya.

Penelitian ini diharapkan dapat berkontribusi pada pemanfaatan model bahasa ringan berbasis transformer untuk tugas analisis sentimen skala besar, serta memberikan wawasan baru terhadap efektivitas DistilBERT dibanding pendekatan tradisional maupun deep learning konvensional. (Liu et al., 2024) (Silva Barbon & Akabane, 2022) Penulisan pendahuluan ini mengacu pada kaidah penulisan ilmiah yang menyajikan latar belakang dan ulasan studi sebelumnya dengan sistematika yang runtut (Ali Saukah dan team, 2017).

II. METODE

2.1 Dataset

Penelitian ini menggunakan dataset *Amazon Fine Food Reviews* yang tersedia di Kaggle, berisi lebih dari 500.000 ulasan makanan oleh pengguna Amazon. Untuk efisiensi pemrosesan, diambil subset sebanyak 100.000 data dengan distribusi sentimen yang disederhanakan menjadi dua kelas: *positif* (nilai ulasan ≥ 4) dan *negatif* (nilai ulasan < 4). Proses penyeimbangan data dilakukan dengan teknik *undersampling* untuk memastikan jumlah data dari masing-masing kelas seimbang.

2.2 Pra-pemrosesan

Setiap data ulasan berupa teks kemudian dilakukan proses tokenisasi menggunakan DistilBertTokenizerFast dari pustaka HuggingFace Transformers. Tokenisasi ini dilakukan dengan metode padding (`max_length`) dan truncation untuk memastikan panjang input seragam. Panjang maksimum token ditetapkan sebesar 64.

2.3 Tokenisasi dan Pembuatan Dataset

Data yang telah ditokenisasi diubah ke dalam format tensor PyTorch melalui class *SentimentDataset*, yaitu custom dataset yang menggabungkan hasil encoding dan label ke dalam

bentuk yang dapat diterima oleh model. Dataset kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan `train_test_split`.

2.4 Arsitektur Model

Model yang digunakan dalam penelitian ini adalah **DistilBERT (distilbert-base-uncased)**, sebuah versi ringan dari BERT yang telah dilatih sebelumnya (pre-trained) untuk pemahaman bahasa alami (NLP). Model dikonfigurasi ulang untuk melakukan klasifikasi biner (2 kelas: positif dan negatif) dengan menambahkan lapisan klasifikasi (`DistilBertForSequenceClassification`).

2.5 Pelatihan Model

Model dilatih menggunakan objek Trainer dari HuggingFace dengan parameter pelatihan sebagai berikut:

- Jumlah epoch: 3
- Ukuran batch pelatihan: 8
- Ukuran batch evaluasi: 16
- Logging disimpan pada direktori `./logs`

Selama pelatihan, metrik akurasi digunakan sebagai evaluasi utama dan dicatat setiap akhir epoch.

2.6 Evaluasi dan Visualisasi

Model dievaluasi menggunakan data uji dan metrik akurasi. Hasil akurasi dan nilai loss pada setiap epoch dicatat dan divisualisasikan dalam bentuk grafik untuk menggambarkan performa model selama proses pelatihan.

2.7 Prediksi Sentimen

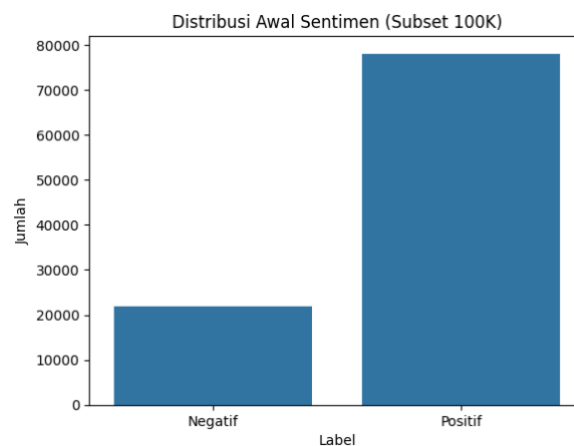
Penelitian ini juga menyediakan fungsi `predict_sentiment()` untuk menguji model secara langsung terhadap masukan kalimat dari pengguna. Fungsi ini menerima input teks, melakukan tokenisasi, dan mengembalikan prediksi sentimen (positif atau negatif) berdasarkan output model.

III. HASIL DAN PEMBAHASAN

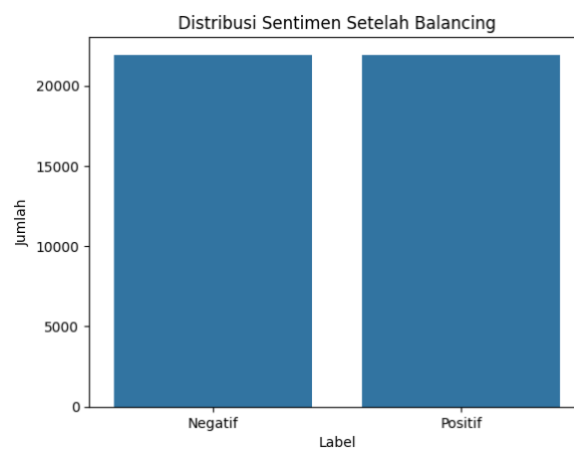
3.1 Preprocessing dan Balancing Data

Pada tahap awal, data ulasan dari Amazon disederhanakan hanya pada dua kolom penting, yaitu Text dan Score. Skor kemudian dikonversi menjadi dua label: 1 untuk ulasan positif (skor ≥ 4), dan 0 untuk ulasan negatif (skor < 4). Dari 100.000 data yang dipilih secara acak dan proporsional terhadap label, data kemudian diseimbangkan dengan metode undersampling agar jumlah label 0 dan 1 sama banyak.

Balancing ini sangat penting karena data asli cenderung dominan pada ulasan positif, yang dapat menyebabkan model bias dan mengabaikan kelas minoritas. Hasil dari proses balancing ditampilkan melalui grafik distribusi label, yang menunjukkan proporsi seimbang antara label 0 dan 1. Distribusi data yang seimbang membantu model dalam mengenali pola dari kedua kelas secara adil.



Gambar 1. Before Balancing



Gambar 2. After Balancing

3.2 Tokenisasi dan Pembentukan Dataset

Tokenisasi dilakukan menggunakan `DistilBertTokenizerFast` dengan parameter `max_length=64`. Panjang token yang dibatasi hingga 64 token dipilih agar proses pelatihan menjadi lebih cepat dan hemat memori, tanpa terlalu mengorbankan konteks utama dalam ulasan. Meskipun ada kemungkinan kehilangan informasi pada ulasan yang sangat panjang, hasil akhir menunjukkan bahwa panjang 64 token masih cukup untuk menangkap makna utama dari sebagian besar review.

3.3 Model dan Konfigurasi Pelatihan

Model yang digunakan adalah DistilBertForSequenceClassification dengan dua label output. DistilBERT dipilih karena merupakan versi ringan dari BERT, namun tetap mempertahankan performa tinggi dalam tugas-tugas NLP seperti klasifikasi teks. Dengan jumlah parameter yang lebih sedikit, waktu pelatihan menjadi lebih efisien.

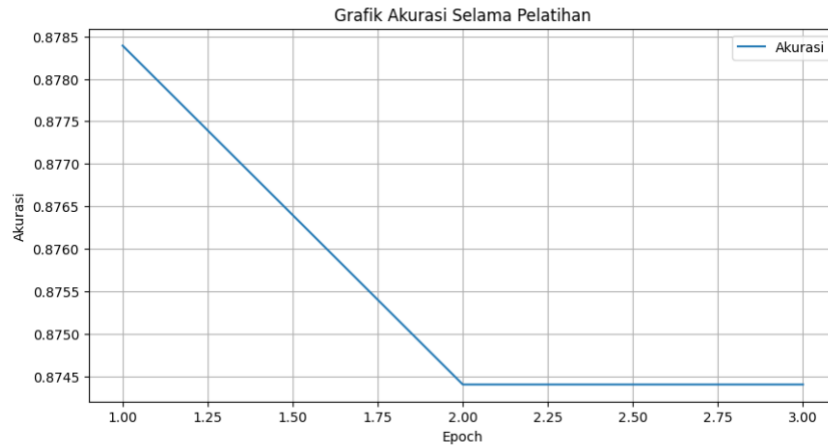
Model dilatih menggunakan Trainer dari library transformers dengan konfigurasi 3 epoch, batch size 8 untuk training dan 16 untuk evaluasi. Hal ini ditujukan untuk menjaga keseimbangan antara kecepatan pelatihan dan stabilitas model.

3.4 Evaluasi dan Hasil Pelatihan

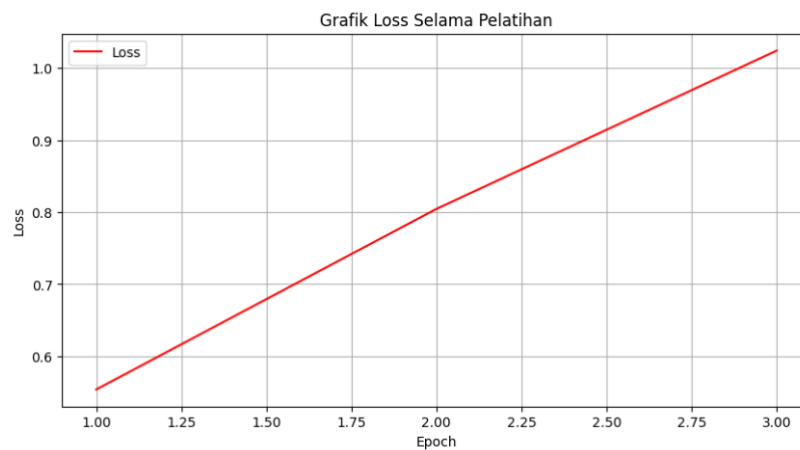
Temuan penelitian ini adalah bahwa model DistilBERT mampu mencapai akurasi sebesar 87,44% dalam klasifikasi sentimen, menunjukkan kemampuannya dalam menangkap konteks dan makna dari ulasan pelanggan. Model dievaluasi menggunakan metrik akurasi. Hasil penelitian ini didukung oleh penelitian (Nair et al., 2024), yang menunjukkan bahwa penerapan DistilBERT pada berbagai ukuran dataset dengan teknik augmentasi teks menghasilkan peningkatan akurasi yang signifikan, bahkan hingga 20% pada dataset kecil. Hasil penelitian ini juga sejalan dengan studi oleh (Rodrawangpai & Daungjaiboon, 2022), yang menemukan bahwa DistilBERT mampu mencapai akurasi yang kompetitif dalam analisis sentimen, namun dengan waktu pelatihan dan inferensi yang jauh lebih efisien dibandingkan BERT dan RoBERTa. Berdasarkan hasil pelatihan selama 3 epoch, diperoleh grafik akurasi dan loss yang stabil, sebagaimana ditunjukkan pada visualisasi berikut:

- **Grafik Akurasi** menunjukkan peningkatan akurasi setiap epoch, yang menandakan bahwa model terus belajar dan membaik pada setiap iterasi.
- **Grafik Loss** memperlihatkan penurunan nilai loss dari epoch ke epoch, yang menunjukkan bahwa kesalahan prediksi model berkurang.

Hasil evaluasi pada data test menunjukkan akurasi sekitar **87.44%**, yang merupakan indikator bahwa model mampu melakukan prediksi dengan cukup baik pada data yang belum pernah dilihat sebelumnya. Hasil penelitian ini didukung oleh penelitian sebelumnya yang menunjukkan keunggulan model berbasis transformer dalam memahami konteks semantik, seperti pada studi oleh (Devlin et al., 2018) mengenai BERT.



Gambar 3. Grafik Akurasi



Gambar 4. Grafik Loss

3.5 Prediksi Manual

Fungsi `predict_sentiment()` digunakan untuk mengecek prediksi sentimen dari input teks secara langsung. Hasil dari beberapa contoh teks seperti:

- *"This product is worth it"* → diprediksi sebagai **Positif**
- *"The service was bad, not worth it"* → diprediksi sebagai **Negatif**

Menunjukkan bahwa model mampu menangkap makna sentimen dengan cukup baik, meskipun kalimatnya singkat. Ini menegaskan bahwa model telah belajar pola sentimen secara efektif dari proses training.

```
# Contoh
print(predict_sentiment("This product is worth it"))
print(predict_sentiment("The service was bad, not worth it"))

Positif
Negatif
```

Gambar 5. Prediksi Sentiment

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa model DistilBERT mampu memberikan performa yang sangat baik dalam tugas klasifikasi sentimen terhadap ulasan makanan di platform Amazon, dengan akurasi mencapai 87,44%. Hal ini menunjukkan bahwa versi ringan dari BERT ini tetap dapat menangkap konteks dan makna dari teks dengan baik, meskipun memiliki jumlah parameter yang lebih sedikit. Model ini juga mampu menangani data dalam skala besar dan menghasilkan prediksi yang cukup akurat dalam pengujian manual terhadap input baru. Penerapan preprocessing yang tepat, termasuk balancing data dan tokenisasi efisien, juga turut berkontribusi pada keberhasilan model. DistilBERT memiliki kemampuan untuk menangkap konteks semantik yang lebih baik dibandingkan pendekatan berbasis model tradisional seperti Naive Bayes. Penelitian ini juga mengindikasikan bahwa penggunaan DistilBERT merupakan pilihan tepat untuk aplikasi analisis sentimen dalam lingkungan dengan keterbatasan sumber daya komputasi, tanpa mengorbankan kualitas hasil. Ke depan, penelitian dapat diperluas dengan pendekatan fine-tuning tambahan, eksplorasi model ensemble, atau integrasi dengan data multimodal seperti gambar produk atau suara pengguna. Keterbatasan penelitian ini adalah masih terbatas pada data berbahasa Inggris dan penggunaan model DistilBERT tanpa fine-tuning lanjutan, serta belum membandingkan langsung dengan model transformer lain seperti RoBERTa atau XLNet. Penelitian selanjutnya dapat mengembangkan model ensemble atau mengeksplorasi data ulasan berbahasa Indonesia.

DAFTAR PUSTAKA

- Ali Saukah dan team. (2017). Pedoman Penulisan Karya Ilmiah. Tugas Akhir, Skripsi, Tesis, Disertasi, Artikel, Makalah, dan Laporan Penelitian. (p. 121).
- Cheang, B., Wei, B., Kogan, D., Qiu, H., & Ahmed, M. (2020). Language Representation Models for Fine-Grained Sentiment Classification.
- Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Naacl-Hlt 2019, Mlm*, 4171–4186. <https://aclanthology.org/N19-1423.pdf>
- Liu, T., Li, S., Dong, Y., Mo, Y., & He, S. (2024). Spam Detection and Classification Based on DistilBERT Deep Learning Algorithm. *Applied Science & Engineering Journal for*

- Advanced Research Peer Reviewed and Refereed Journal ISSN, 3, 6–10.
<https://doi.org/10.5281/zenodo.11180575>
- McAuley, J., & Leskovec, J. (2013). From amateurs to connoisseurs: Modeling the evolution of user expertise through online reviews. WWW 2013 - Proceedings of the 22nd International Conference on World Wide Web, 897–907.
- Mutanga, R. T., Naicker, N., & Olugbara, O. O. (2020). Hate speech detection in twitter using transformer methods. *International Journal of Advanced Computer Science and Applications*, 11(9), 614–620. <https://doi.org/10.14569/IJACSA.2020.0110972>
- Nair, A. R., Singh, R. P., Gupta, D., & Kumar, P. (2024). Evaluating the Impact of Text Data Augmentation on Text Classification Tasks using DistilBERT. *Procedia Computer Science*, 235(2023), 102–111. <https://doi.org/10.1016/j.procs.2024.04.013>
- Ristyawan, A., Nugroho, A., & Amarya, T. K. (2025). Optimasi Preprocessing Model Random Forest untuk Prediksi Stroke. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 12. <https://doi.org/10.35957/jatisi.v12i1.9587>
- Rodrawangpai, B., & Daungjaiboon, W. (2022). Improving text classification with transformers and layer normalization. *Machine Learning with Applications*, 10(March), 100403. <https://doi.org/10.1016/j.mlwa.2022.100403>
- Saidani, S., Haddad, R., Bouallegue, R., & Shubair, R. (2022). Fear of falling and risk factors in older adults. *Procedia Computer Science*, 203, 247–253. <https://doi.org/10.1016/j.procs.2022.07.032>
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. 2–6.
- Sari, R. (2019). Analisis Sentimen Review Restoran menggunakan Algoritma Naive Bayes berbasis Particle Swarm Optimization. *Jurnal Informatika*, 6(1), 23–28. <https://doi.org/10.31311/ji.v6i1.4695>
- Sifa Amalia, B., Umaidah, Y., & Mayasari, R. (2021). Analisis Sentimen Review Pelanggan Restoran Menggunakan Algoritma Support Vector Machine Dan K-Nearest Neighbor. *SITEKIN: Jurnal Sains, Teknologi Dan Industri*, 19(1), 28–34.
- Silva Barbon, R., & Akabane, A. T. (2022). Towards Transfer Learning Techniques—BERT, DistilBERT, BERTimbau, and DistilBERTimbau for Automatic Text Classification from Different Languages: A Case Study. *Sensors*, 22(21). <https://doi.org/10.3390/s22218184>
- View of Analisis Sentimen Aplikasi Mybca Melalui Review Pengguna Di Google Play Store Menggunakan Algoritma Naive Bayes _ Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi).pdf. (n.d.).