

## ANALISIS KINERJA ALGORITMA KLASIFIKASI TEKS BERT DALAM MENDETEKSI BERITA HOAKS

Assyfa Rasida Hanum<sup>\*1</sup>, Ivykaeyla Adriana Zetha<sup>2</sup>, Julia Nur Fajrina<sup>3</sup>, Rafifah Ayu Wulandari<sup>4</sup>, Salwa Cahyani Putri<sup>5</sup>, Sherla Puspa Andina<sup>6</sup>, Novanto Yudistira<sup>7</sup>

<sup>1,2,3,4,5,6,7</sup>Universitas Brawijaya, Malang

Email: <sup>1</sup>assyfarasida81@student.ub.ac.id, <sup>2</sup>ivykaeylaaa@student.ub.ac.id, <sup>3</sup>juliarina@student.ub.ac.id,  
<sup>4</sup>rafiwulan29@student.ub.ac.id, <sup>5</sup>salwacputri@student.ub.ac.id, <sup>6</sup>sherlapuspa0@student.ub.ac.id,  
<sup>7</sup>yudistira@ub.ac.id

(Naskah masuk: 19 Desember 2023, diterima untuk diterbitkan: 11 Juni 2024)

### Abstrak

Teknologi yang mengalami perkembangan tentu memiliki dampak positif dan negatif. Internet yang menjadi hasil perkembangan teknologi tidak luput dari dampak negatif, yaitu munculnya hoaks. Hoaks merupakan berita palsu yang saat ini menjadi isu penting di masyarakat. Hoaks sengaja disebar untuk menimbulkan keresahan dan kegaduhan dalam berbagai bidang, seperti sosial budaya, politik, hingga ekonomi. Penelitian yang dilakukan berikut sangat penting karena dapat memberikan kontribusi penting dalam memerangi penyebaran berita hoaks di dunia digital yang semakin kompleks dengan melakukan klasifikasi berita benar dan berita hoaks. Dalam penelitian ini, data yang digunakan berasal dari tiga dataset publik yang diunduh dari website bernama "onlineacademiccommunity.uvic.ca", "HuggingFace.co" dan "Kaggle.com". Data dibagi menjadi tiga bagian: pelatihan (70%), validasi (15%), dan pengujian (15%). Model BERT diinisialisasi, optimizer AdamW digunakan dengan NLLLoss, dan model dilatih selama beberapa epoch. Model dievaluasi menggunakan data validasi untuk menghindari overfitting. Model yang digunakan untuk mengklasifikasikan data uji. Hasil evaluasi menunjukkan bahwa model klasifikasi BERT memiliki akurasi sebesar 76% pada data validasi dalam mengklasifikasikan berita hoaks, yang menunjukkan performa atau kinerja model Machine Learning dalam melakukan klasifikasi berita hoaks. Sedangkan pada model klasifikasi BERT Multilingual memiliki akurasi lebih rendah, yakni 63%. Penelitian ini berpotensi memberikan kontribusi penting dalam memerangi penyebaran berita hoaks di dunia digital yang semakin kompleks. Dengan menggunakan BERT sebagai pendekatan, model ini memungkinkan pengidentifikasian berita hoaks yang lebih akurat, serta membantu masyarakat dalam menghindari konsumsi informasi yang salah. Dengan hasil yang positif ini, penelitian ini menunjukkan bagaimana teknologi machine learning dapat digunakan untuk melawan disinformasi dan menjadikan dunia maya menjadi tempat yang lebih terpercaya.

**Kata kunci:** penyebaran berita hoaks, BERT, akurasi, F1-Score, presisi, recall

## PERFORMANCE ANALYSIS OF THE BERT TEXT CLASSIFICATION ALGORITHM IN DETECTING HOAX NEWS

### Abstract

The development of technology certainly has positive and negative impacts. The internet, which is the result of technological development, does not escape the negative impact, namely the emergence of hoaxes. Hoaxes are fake news that is currently an important issue in society. Hoaxes are deliberately spread to cause unrest and uproar in various fields, such as socio-culture, politics, and economics. The following research is very important because it can make an important contribution in combating the spread of hoax news in an increasingly complex digital world by classifying true news and hoax news. In this research, the data used comes from three public datasets downloaded from websites named "onlineacademiccommunity.uvic.ca", "HuggingFace.co" and "Kaggle.com". The data is divided into three parts: training (70%), validation (15%), and testing (15%). The BERT model was initialized, the AdamW optimizer was used with NLLLoss, and the model was trained for several epochs. The model was evaluated using validation data to avoid overfitting. The model was used to classify the test data. The evaluation results show that the BERT classification model has an accuracy of 76% on validation data in classifying hoax news, which shows the performance of the Machine Learning model in classifying hoax news. Meanwhile, the BERT Multilingual classification model has a lower accuracy of 63%. This research has the potential to make an important contribution in combating the spread of hoax news in an increasingly complex digital world. By using BERT as an approach, this model allows for more accurate identification of hoax news, as

*well as assisting the public in avoiding the consumption of misinformation. With these positive results, this research shows how machine learning technology can be used to fight disinformation and make cyberspace a more trusted place.*

**Keywords:** *hoax news, BERT, accuracy, F1-Score, precision, recall*

## 1. PENDAHULUAN

Dalam era informasi digital yang berkembang pesat saat ini, sumber informasi menjadi dua sisi pedang yang dapat digunakan untuk kebaikan atau kejahatan (Tambini, 2018). Internet telah memberikan akses mudah kepada masyarakat umum untuk mendapatkan informasi, namun juga memberikan platform yang memungkinkan berita palsu atau hoaks menyebar dengan cepat dan luas (Vosoughi et al., 2018). Hoaks merupakan informasi yang direkayasa untuk menutupi informasi sebenarnya. Dengan kata lain hoaks juga bisa diartikan sebagai upaya pemutarbalikan fakta menggunakan informasi yang seolah-olah meyakinkan tetapi tidak dapat diverifikasi kebenarannya. Hoaks juga bisa diartikan sebagai tindakan mengaburkan informasi yang sebenarnya, dengan cara membanjiri suatu media dengan pesan yang salah agar bisa menutupi pesan yang benar (Gumilar 2017).

Berita hoaks dapat memiliki dampak serius terhadap masyarakat, seperti mempengaruhi pemilihan umum (Guess et al., 2020), menciptakan ketegangan sosial, dan merusak reputasi individu atau kelompok tertentu (Pennycook & Rand, 2019). Indonesia, sebagai salah satu negara dengan populasi pengguna internet yang besar di Asia Tenggara, tidak terkecuali dari dampak negatif berita hoaks. (We Are Social & Hootsuite, 2021).

Adanya berita bohong dapat mengganggu ketertiban demokrasi, stabilitas kehidupan sosial, budaya, politik, dan ekonomi. Sudah banyak perusahaan yang sahamnya anjlok karena hadirnya berita hoaks di kalangan masyarakat tentang perusahaan tersebut. Menurut hasil penelitian Masyarakat Telematika Indonesia, ditemukan bahwa 44,3% dari responden mengakui sering menerima berita bohong atau misinformasi setiap harinya. Faktanya, menurut Newman et al., menunjukkan bahwa lebih dari separuh berita online dari sampel global yang mereka teliti (55%) mengungkapkan kekhawatiran. Salah satu teknik untuk mengendalikan dan mengurangi berita hoaks adalah dengan membuat sistem yang dapat melakukan klasifikasi berita secara otomatis. Dengan klasifikasi sebuah berita, berita tersebut akan diberi label bahwa berita tersebut benar atau hoaks.

Klasifikasi teks adalah proses di mana dokumen atau teks diberikan label kategori berdasarkan kontennya. Ini melibatkan pembuatan model pembelajaran mesin yang dapat memahami pola dalam teks untuk mengambil keputusan kategoris. Klasifikasi adalah suatu proses pengklasifikasian

dokumen-dokumen kedalam satu atau lebih kategori yang telah didefinisikan sebelumnya atau kedalam kelas-kelas dari dokumen-dokumen yang sama (Y. Data, 2016). Dalam permasalahan ini, klasifikasi teks akan digunakan untuk membedakan antara berita yang berisi informasi yang sesuai dengan fakta (berita non-hoaks) dengan berita yang berisi informasi yang salah atau menyesatkan (berita hoaks). Algoritma klasifikasi teks BERT akan menjadi landasan utama dalam proses ini.

Algoritma BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018) adalah varian dua arah dari jaringan Transformer (Vaswani et al., 2017) dilatih untuk memprediksi masked word dari konteksnya dan untuk mengklasifikasi apakah dua kalimat itu berurutan atau tidak. Cara kerja algoritma BERT ini tidak seperti algoritma pemrosesan bahasa lainnya. BERT akan memproses sebuah kata dengan mempelajari konteks daripada kata tersebut berdasarkan kata-kata yang ada. Algoritma BERT memproses konteks penuh dengan cara melihat pola yang muncul pada sebelum atau sesudah kata (Nayak 2019). Algoritma BERT merupakan pre-trained model yang artinya dapat dilatih pada data teks dalam jumlah besar sebelum disesuaikan dengan spesifikasi tugasnya.

Berdasarkan hasil penelitian terbaru dan kebutuhan untuk mengecek kebenaran berita di Indonesia, kami melakukan pendeteksian berita hoaks menggunakan metode *fine-tuning* BERT. Fokus utama dari studi ini adalah untuk menilai efektivitas metode BERT dalam mendeteksi berita hoaks. Dalam menjalankan proses analisis, kami menggunakan BERT-Base sebagai model, diikuti dengan melakukan penyetelan hiperparameter untuk mencapai hasil terbaik. Evaluasi kami berfokus pada akurasi dan waktu komputasi untuk setiap model yang digunakan.

Dalam penelitian ini, kami memperluas penelitian dengan menggunakan model klasifikasi lainnya yaitu BERT Multilingual (Devlin et al., 2018). BERT Multilingual adalah perluasan dari model BERT yang dirancang untuk menangani beberapa bahasa secara bersamaan. BERT Multilingual merupakan model populer untuk fine-tuning karena mampu mencapai kinerja tingkat tinggi pada berbagai tugas pemrosesan bahasa alami. Model ini telah dilatih menggunakan data dari 102 bahasa yang berbeda dari setiap sampel pelatihan. Penelitian lanjutan menunjukkan bahwa model BERT multibahasa memiliki representasi internal yang independen dan memungkinkannya mengkodekan

informasi dari bahasa yang tidak pernah dilihat selama tahap pelatihan (Pires et al., 2019, Wu et al., 2019). Namun, model BERT Multilingual biasanya dilatih dengan jumlah data yang lebih besar dibandingkan dengan model berbahasa tunggal, sehingga lebih membutuhkan waktu komputasi untuk dilatih dan digunakan.

Feng et al. (2020) baru-baru ini merilis versi yang lebih canggih dari model BERT multibahasa dan mencapai kinerja yang lebih baik pada tugas pemrosesan bahasa alami standar dengan menggunakan 109 bahasa pelatihan. Hal ini menunjukkan evolusi model BERT multibahasa dalam menghadapi tuntutan semakin kompleks dalam pemrosesan bahasa alami. Peningkatan ini mendorong penggunaan model ini dalam konteks penelitian kami untuk mendeteksi berita hoaks, dengan harapan dapat meningkatkan ketepatan dan keandalan deteksi lintas bahasa.

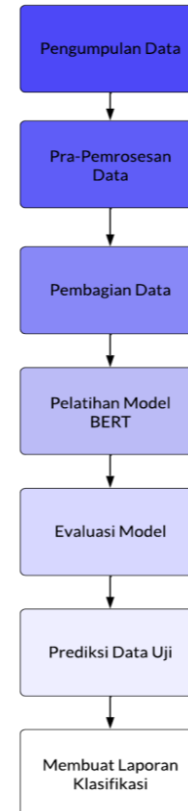
Tujuan dari penelitian ini adalah untuk menguji kinerja algoritma BERT dalam mendeteksi berita hoaks menggunakan model klasifikasi BERT dan model klasifikasi BERT Multilingual. Penelitian ini diharapkan dapat menambah wawasan serta memberikan solusi yang efektif untuk mengatasi penyebaran berita hoaks di media-media *online*. Penelitian ini memiliki relevansi yang signifikan dalam konteks penyebaran berita di Indonesia, mengingat tingginya konsumsi berita *online* dan berbagai tantangan yang dihadapi oleh masyarakat dalam membedakan berita asli dari berita hoaks (Digital 2021: Indonesia, *We Are Social & Hootsuite*). Dengan memanfaatkan teknik-teknik analisis teks dan pembelajaran mesin, penelitian ini bertujuan untuk memberikan kontribusi pada upaya mengatasi permasalahan berita hoaks di Indonesia.

## 2. METODE PENELITIAN

Metode penelitian ini mencakup sejumlah langkah yang penting dalam proses pengembangan model klasifikasi menggunakan pendekatan BERT (Bidirectional Encoder Representations from Transformers). Metode penelitian yang digunakan dalam melakukan analisis ini adalah pengumpulan data, pra-pemrosesan data, pembagian data, pelatihan model BERT, evaluasi model, prediksi data uji, dan terakhir membuat laporan klasifikasi.

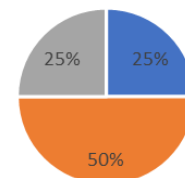
Berikut adalah penjelasan rinci dari setiap tahapan metode penelitian pada gambar 1.

1. Pengumpulan data: Data dikumpulkan dari 3 dataset publik yang terdiri dari artikel berita hoaks dan asli dengan komposisi data "onlineacademiccommunity.uvic.ca" (50%) , "HuggingFace.co" (25%) dan "Kaggle.com" (25%)



Gambar 1. Tahapan Metode Penelitian

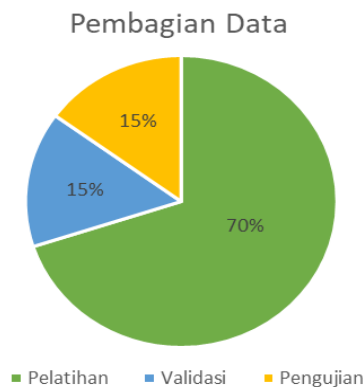
### Komposisi Penggabungan Data



- <https://www.kaggle.com/datasets/linkish/indonesian-fact-and-hoax-political-news>
- <https://onlineacademiccommunity.uvic.ca/isot/2022/11/27/fake-news-detection-datasets/>
- <https://huggingface.co/datasets/Rifky/indonesian-hoax-news/tree/main/data>

Gambar 2. Komposisi Penggabungan Data

2. Pra-pemrosesan data: Data digabungkan dari dua file (True.csv dan Fake.csv), diberi label, dan di-tokenisasi dengan BERT.
3. Pembagian data: Data dibagi menjadi tiga bagian: pelatihan (70%), validasi (15%), dan pengujian (15%).



Gambar 3. Pembagian Data

4. Pelatihan model bert: Model BERT diinisialisasi, optimizer AdamW digunakan dengan NLLLoss, dan model dilatih selama beberapa epoch.
5. Evaluasi model: Model dievaluasi menggunakan data validasi untuk menghindari overfitting.
6. Prediksi data uji: Model digunakan untuk mengklasifikasikan data uji.
7. Membuat laporan klasifikasi: Hasil prediksi dibandingkan dengan label asli, dan metrik klasifikasi seperti presisi, recall, F1-score, dan akurasi dilaporkan.

## 2.1 Pengumpulan Data

Pada penelitian ini, data yang digunakan tidak dikumpulkan secara independen, melainkan berasal dari tiga *dataset* publik yang diunduh dari *website* bernama “*onlineacademiccommunity.uvic.ca*”, “*HuggingFace.co*” dan “*Kaggle.com*”. *Dataset* pertama berjudul “ISOT Fake News Dataset” (Ahmed et al., 2018,). *Dataset* ini berisi dua jenis artikel berita hoaks dan asli dalam bahasa Inggris. Kumpulan data ini didapatkan dari sumber dunia nyata; artikel yang asli diperoleh dengan menelusuri artikel dari Reuters.com (situs web berita). Adapun artikel berita hoaks dikumpulkan dari berbagai sumber. Berita hoaks dikumpulkan dari situs web yang tidak dapat diandalkan yang ditandai oleh Politifact (pemeriksaan fakta organisasi di AS) dan Wikipedia. Kumpulan data berisi berbagai jenis artikel yang berbeda topik, namun sebagian besar artikel berfokus pada topik politik dan berita dunia.

*Dataset* “ISOT Fake News Dataset” (Ahmed et al., 2018,) terdiri atas dua bagian. Bagian pertama berisi lebih dari 12.600 artikel berita faktual dan terverifikasi dari reuter.com. Bagian kedua berisi lebih dari 12.600 artikel dari sumber outlet berita hoaks yang berbeda. *Dataset* ini memiliki 8 kolom yang terdiri atas judul artikel, teks, jenis dan tanggal artikel diterbitkan.

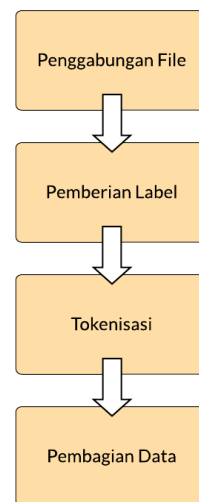
*Dataset* kedua berjudul Indonesian Hoax News yang diunduh dari *website* HuggingFace.co. *Dataset* ini berisi lebih dari 16.000 artikel berita hoaks dalam bahasa Indonesia yang berasal dari TurnBackHoax

yang merupakan komunitas online untuk memerangi berita hoaks di Indonesia.

*Dataset* ketiga dengan judul *Indonesian Fact and Hoax Political News* merupakan *dataset* yang dikumpulkan menggunakan metode Web Scraping dari CNN, Kompas, Tempo, dan Turnbackhoax. *Dataset* ini menyajikan lebih dari 30.000 artikel berita dalam 2 bentuk data yaitu raw data and cleaned data.

## 2.2 Pra-pemrosesan Data

Dalam tahap pra-pemrosesan data, terdapat beberapa langkah penting yang dilakukan untuk mempersiapkan data sebelum digunakan dalam pelatihan model klasifikasi BERT. Tahapan ini mencakup proses penggabungan file, pemberian label, tokenisasi dengan BERT, dan pembagian data menjadi tiga bagian yang melibatkan data pelatihan, data validasi, dan data pengujian.



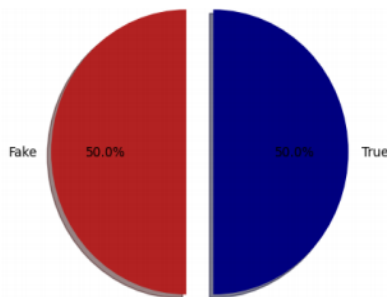
Gambar 4. Tahapan Pra-pemrosesan Data

Berikut adalah penjelasan rinci dari gambar 4 tentang setiap tahap pra-pemrosesan data.

1. Penggabungan file: Pada *dataset* pertama terdapat dua file, yaitu True.csv dan Fake.csv. Penggabungan dari *dataset* pertama kemudian digabungkan dengan *dataset* kedua dan *dataset* ketiga. *Dataset* ketiga terdiri dari 8 file yang terbagi menjadi 4 file cleaned data dan 4 file raw data, sedangkan *dataset* kedua hanya terdiri dari 1 file.
2. Pemberian label: Penggabungan dari ketiga *dataset* ini diberi tambahan satu kolom lagi yakni kolom target yang mana pada file True.csv target bernilai 1, dan pada file Fake.csv target bernilai 0. Kemudian, kedua file tersebut digabungkan dan urutan barisnya diacak.

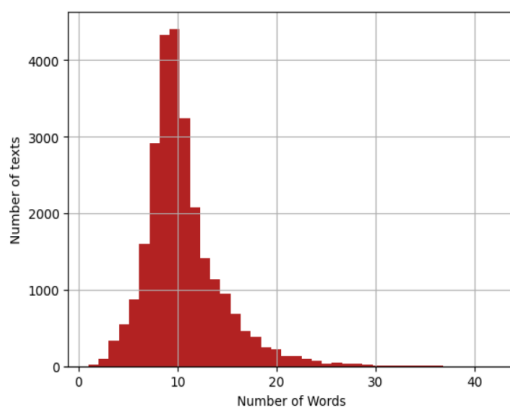
Bagian dari visualisasi grafik pada gambar 5 yang diwarnai dengan merah tua mewakili proporsi data yang memiliki label "Fake". Sedangkan bagian dari grafik yang diwarnai dengan biru tua mewakili proporsi data yang memiliki label "True". Angka-angka di sekitar grafik menunjukkan persentase relatif dari masing-masing kategori terhadap keseluruhan

dataset. Angka ini memberikan gambaran tentang sejauh mana keseimbangan data. Dengan melihat grafik tersebut, sektor "Fake" dan "True" memiliki proporsi yang sama. Maka dapat disimpulkan bahwa data antara kategori "Fake" dan "True" sudah seimbang. Seimbang data dapat bermanfaat dalam konteks analisis dan pembelajaran mesin, karena model yang dilatih pada dataset yang seimbang cenderung memiliki performa yang lebih baik dalam mengenali dan mengklasifikasikan kedua kategori.



Gambar 5. Keseimbangan Data

3. Persiapan data masukan: Selanjutnya dilakukan persiapan data masukan dengan melakukan penghitungan panjang kata untuk setiap judul dalam data latih. Selanjutnya, dilakukan visualisasi distribusi panjang kata menggunakan histogram.



Gambar 6. Visualisasi distribusi data input

Grafik histogram memberikan gambaran tentang sebaran panjang kata dalam dataset, dengan sumbu x menunjukkan jumlah kata dan sumbu y menunjukkan jumlah teks. Dari grafik di atas, didapatkan bahwa jumlah kata dalam judul paling banyak adalah dengan 10 kata dengan frekuensi lebih dari 4000.

4. Tokenisasi dengan BERT: Tokenisasi ini digunakan untuk memecah teks menjadi token-token yang lebih kecil sesuai dengan aturan model BERT. Teks diurai menjadi token, kemudian di-padding jika terlalu pendek atau dipotong jika terlalu panjang. Sehingga teks mencapai panjang maksimum yang ditentukan.

Setiap token diberi mask dan di-encode sebagai representasi angka sesuai dengan kamus kata BERT. Hasil tokenisasi adalah representasi numerik dari teks yang dapat digunakan sebagai input untuk model BERT, memungkinkan model untuk memproses dan memahami teks dengan baik dalam konteks pemrosesan bahasa alami. Selanjutnya mengubah token menjadi input IDs dan attention masks yang dibutuhkan oleh model BERT.

5. Membagi data menjadi 3 bagian, yaitu data pelatihan sebesar 70%, data validasi sebesar 15%, dan data pengujian sebesar 15%. Data yang telah melalui tahap pra-pemrosesan dan embedding kemudian dibagi menjadi dua subset, yaitu data pelatihan dan data validasi, yang digunakan untuk melatih model dan menguji performanya dalam tugas klasifikasi.

### 2.3 Pelatihan Model BERT

Pelatihan model klasifikasi BERT dilakukan dalam bentuk model transformers yang menerapkan pelatihan berarah ganda (bidirectional training) dan menggabungkan konteks dari lapisan kiri dan kanan. model pelatihan BERT berfokus pada Attention.

1. Inisialisasi model: Model BERT diinisialisasi dan struktur yang digunakan adalah model yang telah ditentukan sebelumnya.
2. Optimizer dan loss function: Optimizer AdamW digunakan dengan learning rate sebesar  $1e-5$ , dan loss function yang digunakan adalah NLLoss.
3. Data loader: Data pelatihan dan validasi dimuat menggunakan DataLoader.
4. Melatih model klasifikasi: Dalam penelitian ini, kami menggunakan 2 pemodelan yaitu menggunakan model klasifikasi BERT dan model klasifikasi BERT Multilingual. Selanjutnya. Selama pelatihan, model menghitung prediksi untuk data pelatihan dan membandingkannya dengan label yang sebenarnya untuk menghitung nilai loss. Proses pelatihan berlangsung selama beberapa epoch.
5. Menghitung embedding: Setiap batch data teks yang telah melalui tokenisasi dikirimkan ke model BERT untuk menghitung embedding. Proses ini terjadi pada setiap iterasi pelatihan dan pengambilan embedding dilakukan dari state tersembunyi terakhir token [CLS] pada setiap teks dalam batch tersebut.
6. Melakukan prediksi pada data validasi: Selama pelatihan, juga ada tahap evaluasi menggunakan data validasi untuk memonitor kinerja model dan menghindari overfitting. Hasil model pada data validasi, seperti nilai loss atau kinerja sesuai metrik yang ditentukan, akan dibandingkan dengan hasil sebelumnya. Ketika model menghasilkan hasil yang lebih baik, model tersebut akan disimpan sebagai "model terbaik" dan dapat disimpulkan model tersebut

adalah versi yang paling optimal dan akurat yang telah ditemukan selama proses pelatihan.

7. Setelah melatih model, langkah selanjutnya adalah menggunakan model yang telah dilatih untuk melakukan prediksi pada data uji yang belum pernah dilihat sebelumnya. Ini dilakukan dengan menggunakan model untuk mengklasifikasikan data uji.
8. Menampilkan laporan klasifikasi: Hasil prediksi dibandingkan dengan label asli dari set validasi, dan kemudian dilaporkan metrik klasifikasi seperti presisi, recall, F-1 score, dan akurasi untuk mengevaluasi kinerja model. Evaluasi ini memberikan pemahaman tentang sejauh mana model berhasil dalam memprediksi kategori data uji dengan benar.

## 2.4 Persamaan Aritmatika

Dataset yang digunakan untuk model ini tersedia secara publik di portal “Kaggle.com” dan berasal dari karya Ahmed, H., Traore, I. & Saad, S. Artikel berita aslinya diambil dari situs web berita ‘Reuters.com’, sementara yang palsu dikumpulkan dari dataset lain di portal ‘Kaggle.com’. Dataset ini dibagi awalnya menjadi dua file CSV. Satu untuk berita yang terverifikasi, dengan 21.417 sampel, dan yang lain untuk berita hoaks, dengan 23.481 sampel. kedua file tersebut harus digabungkan dengan diacak. Empat atribut mendeskripsikan setiap sampel: judul, teks artikel, subjek artikel, dan tanggal publikasi. Karena tujuan dari penelitian ini adalah untuk mensimulasikan konten yang ada di platform media sosial seperti Twitter, dari empat atribut tersebut, hanya ‘judul’ yang digunakan. Variable dependent harus ditambahkan secara manual ke dataset. Dataset ini dibagi menjadi bagian pelatihan dan pengujian, dengan 80% dari semua sampel termasuk dalam bagian pelatihan. Setelah model menyelesaikan tugas klasifikasi, data pengujian diperluas dengan kolom tambahan yang berisi prediksi model. Untuk tujuan evaluasi model, digunakan confusion matrix untuk mengukur kinerja model dalam melakukan klasifikasi berita hoaks. Confusion Matrix adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Keempat istilah tersebut adalah True Positive (TP), True Negative (TN), False Positive (FP) dan False Negative (FN). Confusion matrix memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh model dengan hasil klasifikasi sebenarnya. Di bawah ini menampilkan sebuah confusion matrix untuk pengklasifikasian ke dalam dua kelas.

Tabel 1. Visualisasi Confusion Matrix

		Aktual	
		Positif	Negatif
Prediksi	Positif	True Positive (TP)	False Positif (FP)
	Negatif	False Negative (FN)	True Negative (TN)

Tujuan utama dari confusion matrix adalah untuk memvisualisasikan dan menganalisis hasil

prediksi yang dibuat oleh model, sehingga memudahkan untuk memahami kelebihan dan kekurangan model dalam mengklasifikasikan berita hoaks. Confusion matrix digunakan untuk menghitung berbagai performance metrik untuk mengukur kinerja model yang telah dibuat. Berikut beberapa performance metrics yang digunakan:

### 1. Precision

Rasio presisi adalah ukuran yang menunjukkan sejauh mana model dapat mengidentifikasi berita hoaks dengan benar. Dalam konteks ini, berita hoaks yang benar-benar dideteksi oleh model adalah True Positive (TP), sedangkan berita yang diklasifikasikan sebagai berita hoaks tetapi sebenarnya bukan, adalah False Positive (FP).

Presisi (P) dapat dirumuskan dengan persamaan berikut:

$$P = \frac{TP}{TP + FP}$$

Keterangan:

- *p (Precision)*: perbandingan antara TP (*True Positive*) dengan banyaknya data yang diprediksi positif.
- TP (*True Positive*): Model memprediksi data ada di kelas Positif dan yang sebenarnya data memang ada di kelas Positif.
- FP (*False Positive*): Model memprediksi data ada di kelas Positif, namun yang sebenarnya data ada di kelas Negatif.

Dengan kata lain, presisi adalah rasio antara berita hoaks yang benar-benar dideteksi oleh (TP) dan semua berita yang diklasifikasikan oleh model sebagai berita hoaks (Jumlah TP dan FP). Nilai presisi yang tinggi menunjukkan bahwa model lebih akurat dalam mengidentifikasi berita hoaks dan menghasilkan sedikit kesalahan positif hoaks. Presisi tidak cukup untuk mengevaluasi kinerja model, karena tidak terdeteksi oleh model (*False Negative*). Presisi sering digunakan bersama dengan matriks lain untuk memberikan gambaran tentang kinerja model.

### 2. Recall

Recall adalah ukuran yang menunjukkan sejauh mana model dapat mendeteksi semua berita hoaks yang ada dalam dataset. Dalam konteks ini, berita hoaks yang benar-benar dideteksi oleh model adalah True Positives (TP), sedangkan berita hoaks yang tidak terdeteksi oleh model adalah False Negatives (FN).

Recall dapat dirumuskan dengan persamaan berikut:

$$r = \frac{TP}{TP + FN}$$

Keterangan:

- $r$  (Recall): perbandingan antara TP (*True Positive*) dengan banyaknya data yang sebenarnya positif
- TP (*True Positive*): Model memprediksi data ada di kelas Positif dan yang sebenarnya data memang ada di kelas Positif.
- FN (*False Negative*): Model memprediksi data ada di kelas Negatif, namun yang sebenarnya data ada di kelas Positif.

Dengan kata lain, recall adalah rasio antara berita hoaks yang benar-benar dideteksi oleh model (TP) dan semua berita hoaks yang ada dalam dataset (jumlah TP dan FN). Nilai recall yang tinggi menunjukkan bahwa model lebih baik dalam mendeteksi berita hoaks dan menghasilkan sedikit kesalahan *negative* hoaks.

### 3. F1-Score

F1-score adalah matriks yang menggabungkan presisi dan recall menjadi satu angka tunggal, yang merupakan rata-rata harmonik dari kedua skor tersebut. F1-score memberikan gambaran yang lebih baik tentang kinerja model secara keseluruhan dibandingkan dengan hanya melihat presisi atau recall saja.

F1-score (f1) dapat dirumuskan dengan persamaan berikut:

$$F1 = 2 * \frac{r * p}{r + p}$$

Keterangan:

- F1 (F1-Score): perbandingan rata-rata precision dan recall yang dibobotkan.
- $r$  (Recall): perbandingan antara TP (*True Positive*) dengan banyaknya data yang sebenarnya positif
- $p$  (Precision): perbandingan antara TP (*True Positive*) dengan banyaknya data yang diprediksi positif.

Dengan kata lain, F1-score adalah rata-rata harmonik dari presisi ( $p$ ) dan recall. Nilai F1-score sangat berguna dalam situasi dimana kita ingin mencari keseimbangan antara presisi dan recall dan tidak ada lebih penting daripada yang lain. Dalam kasus ini deteksi berita hoaks, kita mungkin ingin model memiliki presisi yang baik (minimalkan berita hoaks yang diklasifikasikan sebagai benar) dan recall yang baik (minimalkan berita benar yang diklasifikasikan sebagai palsu). Oleh karena itu, F1-score adalah metrik yang baik untuk mengevaluasi kinerja model dalam kasus ini.

### 4. Support

Support digunakan untuk mewakili sampel yang ada di setiap kelas set pengujian.

### 5. Macro avg

Macro avg (average) adalah Macro average merupakan teknik dalam menghitung nilai rata-rata dari berbagai metrik evaluasi seperti presisi, recall, atau f1-score untuk semua kelas tanpa mempertimbangkan

ketidakseimbangan dalam jumlah sampel antar kelas. Dalam proses ini, metrik evaluasi dihitung terlebih dahulu untuk setiap kelas, kemudian nilai-nilai tersebut dijadikan sebagai dasar untuk menghitung rata-rata macro. Pendekatan ini memberikan bobot yang sama pada setiap kelas, yang membuatnya ideal digunakan ketika distribusi sampel antar kelas cenderung seimbang.

Macro avg dapat dirumuskan dengan persamaan berikut:

$$Macro\ avg = \frac{\sum \text{Nilai matrik evaluasi semua kelas}}{\text{Jumlah kelas}}$$

Keterangan:

- Macro avg: Rata-rata sederhana dari matrik evaluasi
- $\sum$  Nilai matrik evaluasi semua kelas: Jumlah total nilai matrik evaluasi semua kelas
- Jumlah kelas: Banyak kelas yang dimiliki tiap matrik evaluasi

### 6. Weighted avg

Weighted avg adalah metode untuk menghitung rata-rata dari berbagai metrik evaluasi, seperti akurasi, presisi, recall, dan F1-Score, dengan memperhitungkan bobot yang berbeda untuk setiap kelas atau kategori dalam matriks evaluasi. Bobot untuk setiap kelas dapat ditentukan berdasarkan beberapa faktor, seperti distribusi kelas dalam dataset atau kepentingan relatif dari setiap kelas dalam evaluasi yang sedang dilakukan. Weighted avg menghitung rata-rata yang memperhitungkan perbedaan dalam jumlah sampel antara kelas-kelas yang berbeda, sehingga memberikan gambaran yang lebih akurat tentang kinerja keseluruhan dari model klasifikasi. Weighted avg dapat menilai kinerja model dengan mempertimbangkan pentingnya masing-masing kelas dalam evaluasi. Rumus untuk menghitung weighted average bergantung pada konteks penggunaannya.

Weighted avg dapat dirumuskan dengan persamaan berikut:

$$Weighted\ avg = \frac{\sum_{i=1}^n w_i \times m_i}{\sum_{i=1}^n w_i}$$

Keterangan:

- Weighted avg: Rata-rata tertimbang dari matrik evaluasi.
- $w_i$ : Bobot untuk kelas
- $m_i$ : nilai metrik (seperti presisi, recall, F1-Score) untuk kelas  $i$

## 3. HASIL DAN PEMBAHASAN

Analisis dilakukan dengan menggunakan set pengujian yang terpisah untuk mengevaluasi kinerja model BERT dengan metrik-metrik berikut:

Akurasi : Jumlah prediksi yang benar dibagi dengan total prediksi.



Presisi : Tingkat berapa banyak dari prediksi hoaks yang benar positif.  
 Recall : Tingkat berapa banyak dari berita hoaks yang terdeteksi dengan benar.  
 F1-Score : Ukuran gabungan presisi dan recall.  
 Support : Jumlah data yang termasuk dalam masing-masing kelas.  
 Macro avg : Rata-rata sederhana dari matrik evaluasi.  
 Weighted avg : Rata-rata tertimbang dari matrik evaluasi.

Dalam penelitian ini, hasil dari proses pelatihan menggunakan model BERT dapat dilihat pada tabel 1, menunjukkan hasil akurasi yang cukup tinggi, yakni sebesar 76%. Baik nilai precision dan recall juga cukup tinggi yang berarti model memprediksi dengan cukup baik. Nilai recall *class* 1 menunjukkan bahwa model mampu mengklasifikasikan dengan benar 62% berita hoaks sebagai palsu. Presisinya adalah 0,87, yang berarti 87% klasifikasi berita hoaks berdasarkan model, sebenarnya adalah berita hoaks. Sedangkan hasil dari proses menggunakan model BERT multilingual yang dapat dilihat pada Tabel 2, menunjukkan akurasi lebih rendah dari model BERT, yakni sebesar 63%. Namun, terdapat kenaikan dalam recall 1, yang menunjukkan kemampuan model mengklasifikasikan dengan benar 72% berita hoaks sebagai palsu.

Tabel 2. Akurasi Model Bert

	Precision	Recall	F1-Score	Support
0	0.71	0.91	0.79	2850
1	0.87	0.62	0.72	2850
Accuracy			0.76	5700
Macro avg	0.79	0.76	0.76	5700
Weighted avg	0.79	0.76	0.76	5700

Tabel 3. Akurasi Model Multilingual Bert

	Precision	Recall	F1-Score	Support
0	0.66	0.54	0.59	2850
1	0.61	0.72	0.66	2850
Accuracy			0.63	5700
Macro avg	0.63	0.63	0.63	5700
Weighted avg	0.63	0.63	0.63	5700

#### 4. KESIMPULAN

Hasil evaluasi menunjukkan bahwa model klasifikasi BERT memiliki akurasi sebesar 76% pada data validasi dalam mengklasifikasikan berita hoaks, yang menunjukkan performa atau kinerja model Machine Learning dalam melakukan klasifikasi berita hoaks. F1-Score untuk label positif adalah 0,72 dengan Presisi 0,87 untuk label positif, Recall untuk label positif adalah 0,62. Sedangkan, hasil evaluasi dari model klasifikasi BERT Multilingual memiliki akurasi yang lebih rendah dari model klasifikasi BERT, yakni akurasi sebesar 63% pada data validasi dalam mengklasifikasikan berita hoaks, yang menunjukkan performa atau kinerja model Machine Learning dalam melakukan klasifikasi berita hoaks. F1-Score untuk label positif adalah 0,66 dengan

Presisi 0,61 untuk label positif, Recall untuk label positif adalah 0,72. Model Multilingual BERT memiliki akurasi yang lebih rendah dibandingkan dengan Model BERT. Namun, dalam model BERT Multilingual terdapat peningkatan dalam recall untuk label positif. Dengan menggunakan BERT sebagai pendekatan, model ini memungkinkan pengidentifikasian berita hoaks yang lebih akurat, serta membantu masyarakat dalam menghindari konsumsi informasi yang salah. Penelitian ini berpotensi memberikan kontribusi penting dalam memerangi penyebaran berita hoaks di dunia digital yang semakin kompleks.

#### DAFTAR PUSTAKA

- AHMED H, TRAORE I, SAAD S. 2017 "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. In: Traore I., Woungang I., Awad A. (eds) Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. ISDDC 2017. Lecture Notes in Computer Science, vol 10618. Springer, Cham (pp. 127-128).
- AHMED, H., TRAORE, I., & SAAD, S. 2018. Detecting opinion spams and fake news using text classification. *Journal of Security and Privacy*, 1(1). 10.1002/spy2.9.
- BISRI R. B. 2022. Indonesian Hoax news. Hugging Face, <https://huggingface.co/datasets/Rifky/indonesia-n-hoax-news>.
- DEVLIN, J., CHANG, M., LEE, K., & TOUTANOVA, K. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Computing Research Repository*. arXiv:1810.04805v2.
- GHAIDA A. H., RIZQULLAH M. R., & HADI S. I. 2023. Indonesian Fact and Hoax Political News. *Kaggle*. <https://www.kaggle.com/datasets/linkgish/indonesian-fact-and-hoax-political-news>.
- GUESS, A. M., NAGLER, J., & TUCKER, J. 2020. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 6(14).
- GUMILAR, G., JUSTITO, A., & NUNIK, M. n.d.. Literasi Media: Cerdas Menggunakan Media Sosial Dalam Menanggulangi Berita Palsu (Hoax) Oleh Siswa SMA. *Jurnal Pengabdian Kepada Masyarakat*, 1(1), 35-40. <http://jurnal.unpad.ac.id/pkm/article/view/16275/7939>.
- JURAFSKY, D., & MARTIN, J. H. 2020. *Speech and Language Processing* (3rd ed.). Pearson.
- NAYAK, P. 2019. Understanding searches better than ever before. *The Keyword*. Retrieved October 4, 2023, from <https://blog.google/products/search/search-language-understanding-bert/>.



- NEWMAN, N., FLETCHER, R., KALOGEROPOULOS, A., & NIELSEN, R. K. 2019. Reuters institute digital news report 2019. Reuters Institute for the Study of Journalism, 19.
- PENNYCOOK, G., & RAND, D. G. 2019. The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Stories Increases Perceived Accuracy of Stories Without Warnings. *Management Science*, 66(11), 4944-4957.
- RAHUTOMO, F., YANUAR, I., & ASMARA, R. A. 2018. *INDONESIAN HOAX NEWS DETECTION DATASET (1)*. Mendeley Data.
- RAHUTOMO, F., YANUAR, I., & ASMARA, R. A. 2018. *INDONESIAN HOAX NEWS DETECTION DATASET*. Mendeley Data. Retrieved September 26, 2023, from <https://data.mendeley.com/datasets/p3hfgr5j3m/1>.
- SHOLIKHAH, I. I., JAKA, A. T., & LATIFAH, K. 2023. Machine Learning Untuk Deteksi Berita Hoax Menggunakan BERT. *IN-FEST 2023 Seminar Nasional Informatika – FTI UPGRIS*, 524-531.
- TAMBINI, D. 2018. From Digital Divides to Digital Inequality: The Politics of Online Inequalities. *International Journal of Communication*, 12, 10.
- VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, L., & POLOSUKHIN, I. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, 6000–6010. arXiv:1706.03762.
- VOSOUGHI, S., ROY, D., & ARAL, S. 2018. The spread of true and false news online. *Science*, 359(6380), 1146-1151.
- WE ARE SPECIAL & HOOTSUITE. 2021. Digital 2021: Indonesia. <https://wearesocial.com/digital-2021-indonesia>.
- YOUR DATA & SARKAR, D. 2016. *Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from your Data*. apress.

*Halaman ini sengaja dikosongkan*