

# Comparison of TF-IDF and BERT in Fake News Classification Using Soft Voting Majority

**Moh. Rizky Rahmadian Makkani<sup>\*1</sup>, Fadaukas Daffa Tajuddin<sup>2</sup>, Davin Amadeo Wijaya<sup>3</sup>**  
*Institut Teknologi Sepuluh Nopember, Jl. Raya ITS, Surabaya, Jawa Timur, (031) 5994251*  
*E-mail : rizkyrahmadian817@gmail.com<sup>\*1</sup>, fadaukasdaffa04@gmail.com<sup>2</sup>,*  
*davinamadeo1110@gmail.com<sup>3</sup>*

**Abstract** - Perkembangan media digital dan media sosial membuat penyebaran berita semakin cepat, namun juga meningkatkan risiko tersebarnya informasi palsu (*fake news*). Penyebaran berita palsu dapat menyebabkan kesalahan persepsi publik, instabilitas sosial, dan menurunnya kepercayaan terhadap media. Permasalahan utama pada deteksi *fake news* terletak pada representasi teks yang kurang optimal dan performa model klasifikasi yang belum konsisten sehingga akurasi deteksi masih bervariasi. Oleh karena itu, dibutuhkan pendekatan yang mampu membandingkan efektivitas representasi teks tradisional seperti TF-IDF dan representasi berbasis konteks seperti BERT dalam meningkatkan performa model.

Penelitian ini mengusulkan tiga pendekatan utama, pendekatan pertama menggunakan TF-IDF yang digabungkan dengan empat algoritma klasifikasi yaitu *Logistic Regression*, *Naive Bayes*, *XGBoost*, dan *Random Forest* serta *ensemble Soft Voting Majority*. Pendekatan kedua menggunakan BERT *Embeddings* yang dikombinasikan dengan algoritma yang sama dan *Soft Voting Majority*. Pendekatan ketiga menggunakan model BERT yang di-*fine tune* menggunakan arsitektur BERT *for Sequence Classification*.

Rencana pengujian dilakukan melalui beberapa skenario, yaitu perbandingan performa TF-IDF dan BERT menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*, pengujian stabilitas model dengan analisis efektivitas *Soft Voting Majority* dalam meningkatkan kinerja dibanding model tunggal. *Dataset* yang digunakan berasal dari Kaggle dengan 72.134 baris data.

**Keywords** - *Fake News*, TF-IDF, BERT, *Soft Voting*, *Text Classification*.

## 1. INTRODUCTION

Penyebaran informasi melalui internet menjadi semakin masif seiring meningkatnya penggunaan media sosial dan portal berita daring. Meskipun akses informasi menjadi lebih cepat, kemunculan berita palsu (*fake news*) justru menjadi ancaman serius karena dapat memicu disinformasi, memengaruhi opini publik, hingga memicu konflik sosial. Deteksi otomatis terhadap *fake news* menjadi salah satu fokus penelitian penting dalam bidang *Natural Language Processing* (NLP). Namun, berbagai model klasifikasi masih menghadapi kendala dalam memahami konteks bahasa yang kompleks, terutama ketika menggunakan representasi teks tradisional seperti *bag-of-words* atau TF-IDF.

Dalam beberapa tahun terakhir, model berbasis *deep learning*, khususnya *contextual representation model* seperti BERT, menawarkan kemampuan pemahaman konteks yang lebih baik. Meski demikian, model klasik tetap populer karena efisiensi komputasi dan kinerjanya yang stabil. Oleh karena itu, diperlukan studi komprehensif yang membandingkan kedua pendekatan representasi teks tersebut dalam konteks deteksi *fake news*. Untuk mengatasi variabilitas performa antar model, pendekatan *ensemble* seperti *Soft Voting Majority* dapat digunakan untuk meningkatkan konsistensi dan akurasi prediksi.

Penelitian ini melibatkan dataset WELFake dari Kaggle yang berisi 72.134 berita, terdiri dari judul (*Title*), isi berita (*Text*), dan label (*fake* atau *real*). Penelitian ini bertujuan untuk menguji efektivitas TF-IDF dan BERT dalam menghasilkan representasi berita serta menilai kemampuan *ensemble Soft Voting Majority* dalam meningkatkan performa klasifikasi.

## 2. RESEARCH METHOD

Metodologi penelitian ini mencakup penjelasan mengenai karakteristik *dataset*, serta langkah-langkah dalam perancangan sistem yang meliputi pemodelan, implementasi, dan alur pemrosesan. Terakhir, skenario pengujian dijelaskan untuk memastikan model dievaluasi secara sistematis berdasarkan metrik performa, variasi data, serta prosedur eksperimen yang telah ditentukan.

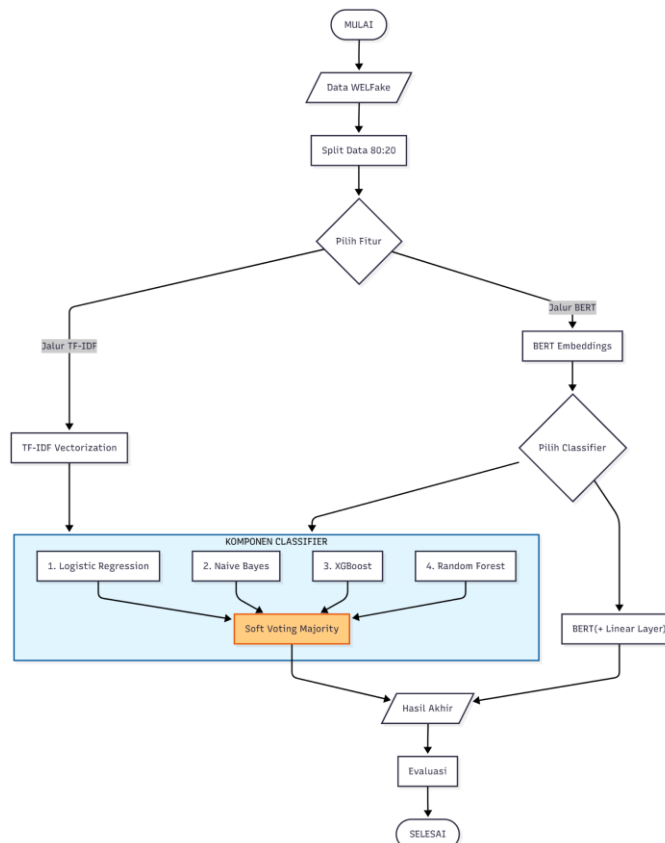
### 2.1. Dataset yang Digunakan

Dataset yang digunakan dalam penelitian ini adalah WELFake *Fake News Classification Dataset* yang diperoleh dari platform Kaggle. Dataset ini berisi 72.134 baris berita yang terdiri dari empat fitur:

1. *Index*: ID berita
2. *Title*: Judul berita
3. *Text*: Isi berita dengan panjang bervariasi
4. *Label*: Kelas berita (0 = *fake*, 1 = *real*)

Seluruh data digunakan dalam proses pelatihan dan pengujian untuk memastikan model memperoleh variasi data yang cukup luas serta menghasilkan performa yang stabil.

### 2.2. Perancangan Sistem



Gambar 1. *Flowchart* skenario pengujian.

*Flowchart* tersebut menggambarkan alur sistem klasifikasi berita palsu yang dimulai dari tahap pengambilan dataset WELFake, kemudian data dibagi menggunakan rasio 80:20 untuk proses pelatihan dan pengujian. Setelah itu, sistem memasuki tahap pemilihan fitur, di mana terdapat dua jalur pemrosesan utama, yaitu TF-IDF dan BERT. Pada jalur TF-IDF, teks dikonversi menjadi representasi vektor menggunakan *TF-IDF Vectorization* dan kemudian diproses oleh beberapa model klasifikasi, yakni *Logistic Regression*, *Naive Bayes*, *XGBoost*, dan

*Random Forest*, yang selanjutnya digabungkan menggunakan metode *Soft Voting Majority* untuk menghasilkan prediksi akhir. Sementara itu, pada jalur BERT, teks diubah menjadi BERT *embeddings* dan diteruskan ke tahap pemilihan *classifier*, baik menggunakan pendekatan BERT dengan *linear layer (fine-tuning)* maupun dikombinasikan dengan *classifier* lain. Hasil dari kedua jalur tersebut kemudian menghasilkan *output* klasifikasi yang selanjutnya dievaluasi menggunakan metrik performa, sebelum proses diakhiri pada tahap penyelesaian.

#### 2.2.1. Preprocessing Teks

Proses *preprocessing* teks bertujuan untuk membersihkan dan menormalisasi data berita. Pada tahap ini, beberapa proses dilakukan seperti *case folding* untuk mengubah seluruh huruf menjadi huruf kecil, *tokenization* untuk memecah teks menjadi unit-unit kata, serta *stopword removal* untuk menghapus kata umum yang tidak memiliki kontribusi signifikan dalam analisis. Selain itu, dilakukan pula penghapusan tanda baca, angka, dan karakter khusus sehingga teks yang dihasilkan lebih bersih dan siap digunakan untuk proses ekstraksi fitur.

#### 2.2.2. Ekstraksi Fitur

Proses ekstraksi fitur dilakukan melalui dua pendekatan berbeda. Pendekatan pertama menggunakan TF-IDF *Vectorizer* untuk menghasilkan representasi *bag-of-words* berbobot berdasarkan frekuensi kemunculan kata dalam dokumen. Pendekatan kedua memanfaatkan BERT untuk menghasilkan *contextual embeddings* dari model *pretrained*, yang mampu menangkap konteks semantik antar kata secara lebih mendalam. Kedua representasi ini akan menjadi *input* bagi model klasifikasi yang dibangun.

#### 2.2.3. Pembangunan Model Klasifikasi

Pada penelitian ini digunakan empat model dasar (*base classifier*), yaitu *Logistic Regression*, *XGBoost*, *Naive Bayes*, dan *Random Forest*. Setiap model dilatih dalam dua kondisi berbeda, yakni menggunakan representasi teks berbasis TF-IDF dan menggunakan *contextual embeddings* yang dihasilkan oleh BERT. Selain kedua pendekatan tersebut, penelitian ini juga memasukkan model ketiga berupa BERT *Classifier* yang memanfaatkan *pooled output* dari model BERT *pretrained* dengan proses *fine-tuning* yang minimal. Dengan kombinasi ini, penelitian dapat melakukan evaluasi secara komprehensif untuk melihat sejauh mana representasi fitur memengaruhi performa model klasifikasi yang sama.

#### 2.2.4. Ensemble Soft Voting Majority

Untuk meningkatkan akurasi dan konsistensi prediksi, seluruh model dasar digabungkan menggunakan metode *ensemble Soft Voting Majority*. Pada metode ini, keputusan akhir tidak diambil berdasarkan suara terbanyak dari model, tetapi berdasarkan rata-rata probabilitas prediksi yang dihasilkan oleh masing-masing model. Metode *soft voting* dipilih karena memiliki kemampuan menggabungkan keunggulan model-model dasar dan mengurangi kelemahan model tertentu, sehingga menghasilkan prediksi yang lebih stabil dan akurat.

#### 2.2.5. Evaluasi Model

Evaluasi model yang dilakukan menggunakan empat metrik utama, yaitu akurasi, presisi, *recall*, dan F1-score. Evaluasi ini bertujuan untuk membandingkan performa antara representasi TF-IDF dan BERT, mengukur dampak penggunaan *ensemble Soft Voting Majority* terhadap peningkatan performa, serta mengamati stabilitas model melalui beberapa kali *resampling* dataset. Melalui tahapan perancangan sistem ini, penelitian diharapkan mampu memberikan analisis komprehensif terhadap efektivitas berbagai representasi teks dan model *ensemble* dalam mendeteksi berita palsu.

### 2.3 Skenario Pengujian

#### 2.3.1. Komparasi Fitur

Skenario pertama, yaitu Komparasi Fitur, dirancang untuk membuktikan apakah representasi konteks yang dihasilkan oleh BERT lebih efektif dibandingkan representasi frekuensi yang dihasilkan oleh TF-IDF dalam mendeteksi berita palsu. Pada skenario ini, *dataset* dibagi secara acak dengan rasio 80:20. Tiga kelompok model kemudian diuji, yakni Kelompok A yang terdiri dari lima model berbasis TF-IDF, Kelompok B yang terdiri dari lima model berbasis BERT *Embeddings*, dan Kelompok C yang terdiri dari model BERT *Classifier*. Evaluasi pada skenario ini difokuskan pada metrik F1-Score serta *False Negative Rate* (FNR), dengan tujuan utama untuk mengetahui apakah kemampuan pemahaman konteks pada BERT dapat mengurangi kemungkinan lolosnya berita palsu sebagai berita asli.

### 2.3.2. Pengujian Granularitas Informasi

Skenario kedua adalah Pengujian Granularitas Informasi, yang bertujuan untuk mengetahui ketahanan dan sensitivitas model terhadap perbedaan kedalaman informasi pada *input* teks. Pengujian dilakukan dalam dua tahap, yaitu tahap pertama yang hanya menggunakan kolom judul (*title only*) dan tahap kedua yang menggunakan kombinasi judul dan konten (*title + text*). Hipotesis pada skenario ini menyatakan bahwa model berbasis TF-IDF akan mengalami penurunan performa yang signifikan ketika hanya diberikan teks pendek seperti judul, karena fitur frekuensi kata menjadi sangat terbatas. Sebaliknya, model berbasis BERT diperkirakan tetap stabil karena memiliki kemampuan memahami konteks semantik meskipun teks yang diberikan relatif minim. Dengan demikian, skenario ini memberikan gambaran mengenai performa model pada tingkat granularitas informasi yang berbeda.

### 2.3.3. Pengujian Efisiensi Data melalui *Learning Curve*

Skenario ketiga adalah Pengujian Efisiensi Data melalui *Learning Curve*, yang bertujuan untuk mengetahui seberapa banyak data yang dibutuhkan oleh masing-masing metode untuk mencapai performa optimal. Pada skenario ini dilakukan serangkaian eksperimen menggunakan 10%, 50%, dan 100% data *training*. Hipotesis yang diajukan adalah bahwa model berbasis BERT, berkat proses *pretraining*-nya yang sangat besar, mampu mencapai akurasi tinggi bahkan dengan hanya 10% data pelatihan, misalnya di atas 90%. Sebaliknya, model berbasis TF-IDF diperkirakan membutuhkan data yang lebih besar untuk menghasilkan performa yang setara. Melalui skenario ini, penelitian dapat memberikan pemahaman mengenai efisiensi pembelajaran pada masing-masing pendekatan representasi teks.

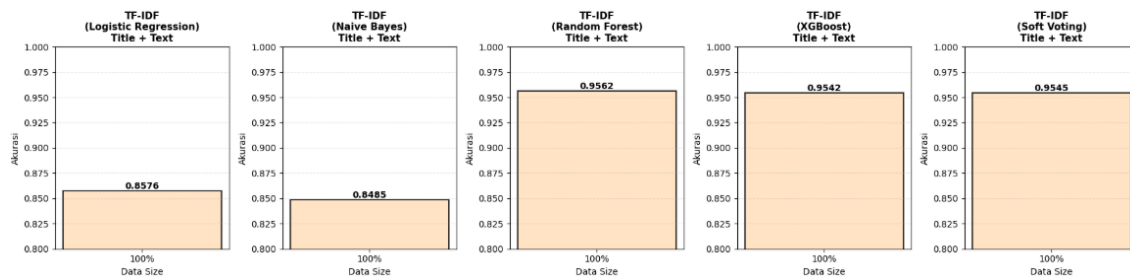
## 3. RESULTS AND DISCUSSION

### 3.1. Komparasi Fitur antara TF-IDF dan BERT

Skenario pertama bertujuan untuk melihat pengaruh representasi teks terhadap performa model klasifikasi. Dua kelompok eksperimen dilakukan: model berbasis TF-IDF dan model berbasis BERT *Embeddings* menggunakan seluruh data (100%) dengan *input* artikel lengkap (*Title + Text*). Selain itu, model ketiga berupa *Fine-Tuned BERT Classifier* diuji secara terpisah untuk melihat performa model konteks penuh.

#### 3.1.1. Performa Model berbasis TF-IDF

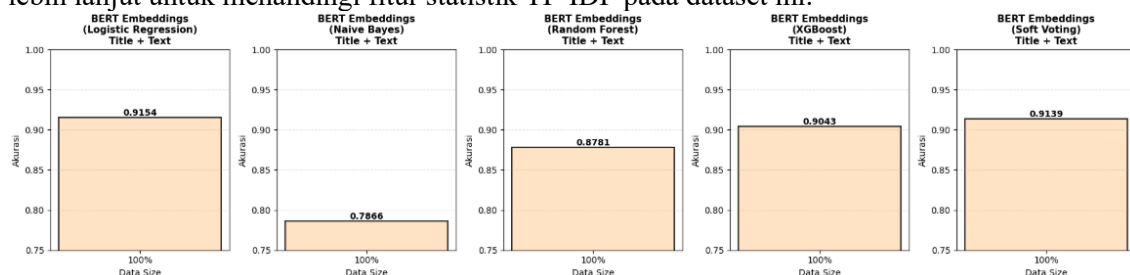
Hasil eksperimen menunjukkan bahwa TF-IDF memberikan performa yang sangat impresif, terutama pada model berbasis *tree*. *Random Forest* mencatatkan akurasi tertinggi sebesar 0.956, diikuti sangat ketat oleh *Soft Voting* (0.955) dan *XGBoost* (0.954). Hal ini mengoreksi asumsi awal bahwa *XGBoost* akan kesulitan dengan vektor sparse; faktanya, *XGBoost* mampu menangani fitur TF-IDF dengan sangat baik. Sebaliknya, model linear dan probabilistik seperti *Logistic Regression* (0.858) dan *Naive Bayes* (0.848) justru menunjukkan performa yang lebih rendah dibandingkan model *ensemble*. *Ensemble Soft Voting Majority* terbukti efektif menjaga stabilitas performa di angka 0.955, menggabungkan kekuatan dari model-model *tree-based* yang mendominasi kelompok ini.



Gambar 2. Perbandingan Performa lima *classifier* yang berbeda pada TF-IDF

### 3.1.2. Performa BERT *Embeddings*

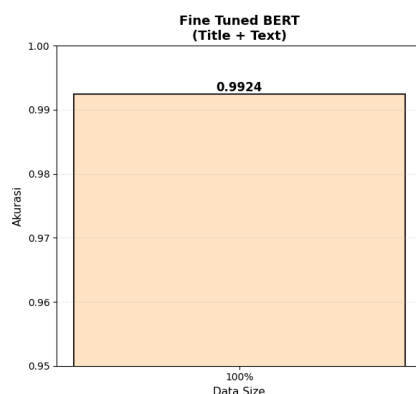
Pada eksperimen kedua, representasi *contextual embedding* dari BERT (tanpa *fine-tuning*) diuji. Hasilnya cukup mengejutkan di mana *Logistic Regression* menjadi model terbaik dengan akurasi 0.915, mengungguli *Soft Voting* (0.914) dan *XGBoost* (0.904). Hal ini menunjukkan bahwa representasi *embedding* yang padat (*dense*) lebih mudah dipisahkan secara linear dibandingkan fitur sparse TF-IDF. *Naive Bayes* kembali menunjukkan performa terlemah (0.787) karena asumsi independensi fiturnya tidak cocok dengan natur *embedding* BERT yang memiliki korelasi antar dimensi yang tinggi. Secara umum, meskipun stabil, akurasi terbaik BERT *Embeddings* (0.915) masih berada di bawah performa terbaik TF-IDF dengan *Random Forest* (0.956), mengindikasikan bahwa *embedding* mentah BERT memerlukan penyesuaian lebih lanjut untuk menandingi fitur statistik TF-IDF pada dataset ini.



Gambar 3. Perbandingan Performa lima *classifier* yang berbeda pada BERT *Embeddings*

### 3.1.3. Performa *Fine-tuned* BERT *Classifier*

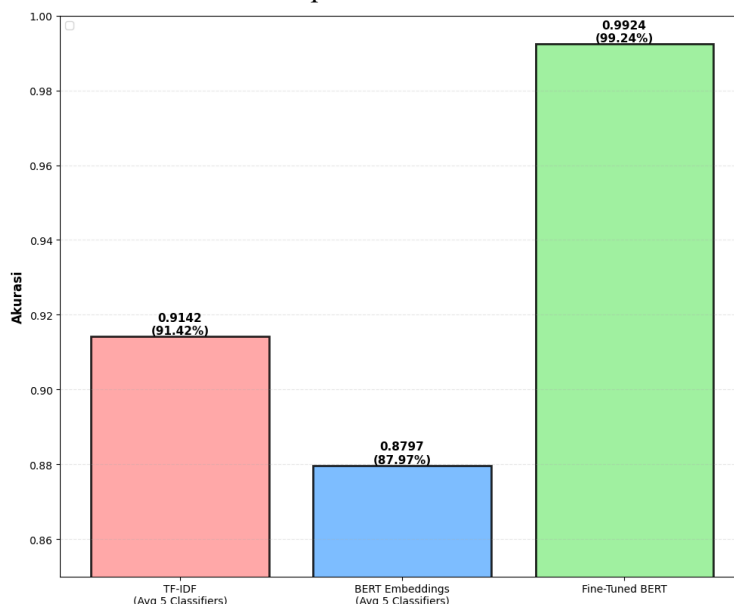
Model *Fine-Tuned* BERT, yang dilatih secara *end-to-end*, memberikan hasil mutlak terbaik pada seluruh skenario eksperimen. Model ini mencapai akurasi 0.992, jauh melampaui pendekatan ekstraksi fitur (TF-IDF maupun BERT *Embeddings*). Dengan memanfaatkan *contextual representation* penuh yang disesuaikan dengan domain dataset *WELFake*, model ini mampu meminimalkan *False Negative* secara signifikan. Hal ini membuktikan bahwa proses *fine-tuning* adalah kunci utama untuk membuka potensi penuh arsitektur *Transformer* dalam deteksi berita palsu.



Gambar 4. Performa *Fine-Tuned* BERT *Classifier*

### 3.1.4. Kesimpulan Komparasi Fitur antara TF-IDF dan BERT

Secara keseluruhan, TF-IDF terbukti sangat efektif jika dipadukan dengan model *ensemble* (*Random Forest/XGBoost*), bahkan mengungguli BERT *Embeddings* standar. Namun, *Fine-Tuned* BERT adalah juara sejati dengan performa nyaris sempurna (0.99). Jika sumber daya komputasi terbatas, TF-IDF + *Random Forest* adalah alternatif terbaik. Jika akurasi adalah prioritas utama, *Fine-Tuned* BERT adalah pilihan mutlak.



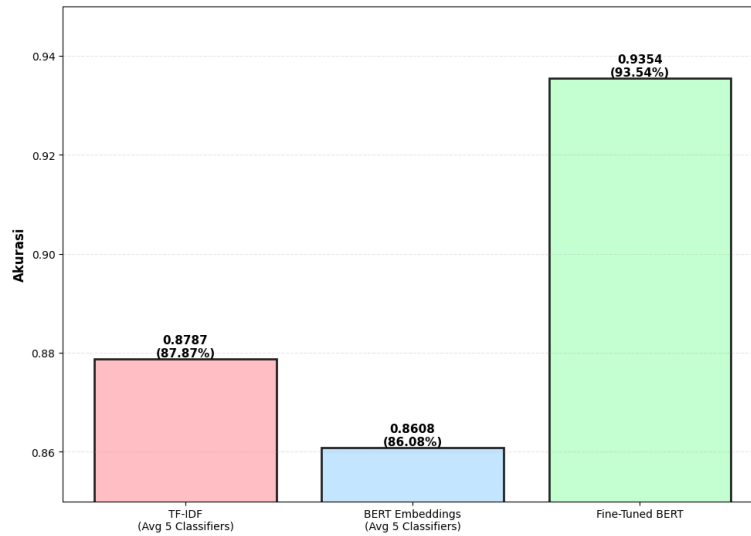
Gambar 5. Perbandingan Performa *Fine Tune* BERT vs BERT *Embedding* vs TF-IDF

## 3.2. Pengujian Granularitas Informasi

Skenario kedua berfokus pada pengaruh panjang dan kedalaman teks. Pengujian membandingkan penggunaan hanya judul (*Title Only*) versus judul dan isi (*Title + Text*).

### 3.2.1. Klasifikasi Berbasis Judul

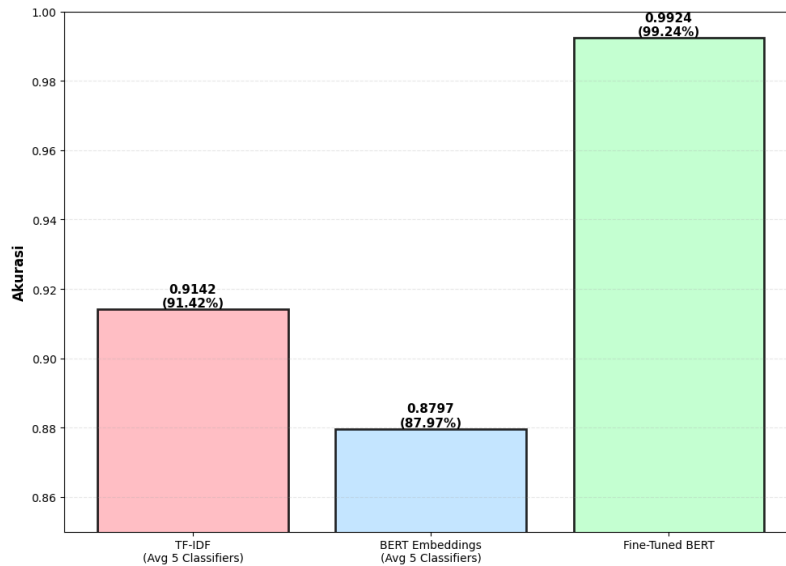
Pada kondisi informasi terbatas (hanya judul), performa kedua metode cenderung seimbang namun TF-IDF sedikit lebih unggul. TF-IDF mencapai akurasi tertinggi melalui *Soft Voting* (0.903) dan *Random Forest* (0.898). Sementara itu, BERT *Embeddings* sedikit tertinggal dengan *Soft Voting* (0.889) dan *XGBoost* (0.888). Menariknya, pada *TF-IDF*, model *XGBoost* justru anjlok ke performa terendah (0.829), sedangkan pada BERT, *Naive Bayes* yang menjadi terendah (0.793). Hal ini menunjukkan bahwa untuk teks sangat pendek, pendekatan statistik sederhana (TF-IDF) lebih mampu menangkap kata kunci indikatif *clickbait* dibandingkan *embedding* semantik yang membutuhkan konteks lebih panjang.



Gambar 6. Perbandingan Performa *Title Only* pada TF-IDF, BERT *Embedding*, dan *Fine Tune* BERT

### 3.2.2. Klasifikasi Berbasis Artikel Lengkap

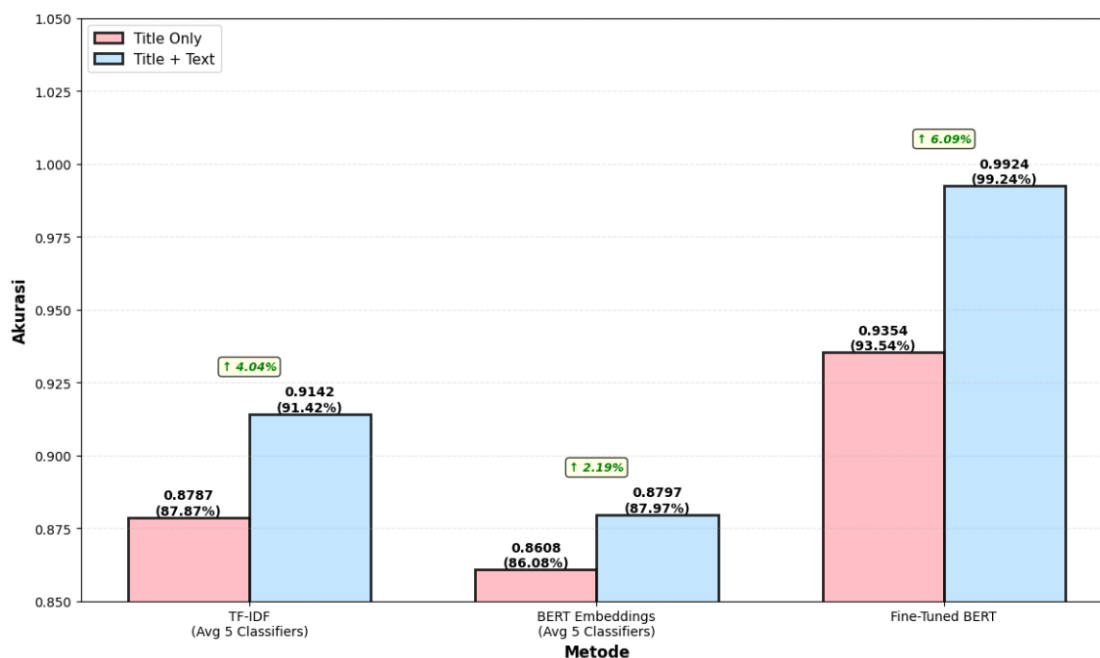
Ketika informasi lengkap tersedia, *Fine-Tuned* BERT melesat ke akurasi 0.992. Namun, untuk pendekatan berbasis fitur, TF-IDF kembali menunjukkan dominasinya atas BERT *Embeddings*. Model-model berbasis TF-IDF (*Random Forest*, *XGBoost*, *Soft Voting*) konsisten mencapai akurasi di atas 0.95, sedangkan model berbasis BERT *Embeddings* mentok di angka 0.915 (*Logistic Regression*). Temuan ini membantah asumsi bahwa BERT selalu lebih baik; tanpa *fine-tuning*, representasi statistik TF-IDF pada teks panjang justru lebih mudah dimanfaatkan oleh klasifikator *machine learning* tradisional dibandingkan *embedding* BERT yang kompleks.



Gambar 7. Perbandingan Performa *Title + Text* pada TF-IDF, BERT *Embedding*, dan *Fine Tune* BERT

### 3.2.3. Kesimpulan Pengujian Granularitas Informasi

Granularitas sangat berpengaruh. Pada level judul, TF-IDF dan BERT bersaing ketat di angka  $\sim 0.90$ . Pada level artikel lengkap, terjadi divergensi: TF-IDF sangat kuat dengan model *tree-based* ( $\sim 0.95$ ), BERT *Embeddings* unggul dengan model linear namun tertahan di  $\sim 0.91$ , sementara *Fine-Tuned* BERT mendominasi total ( $\sim 0.99$ ).

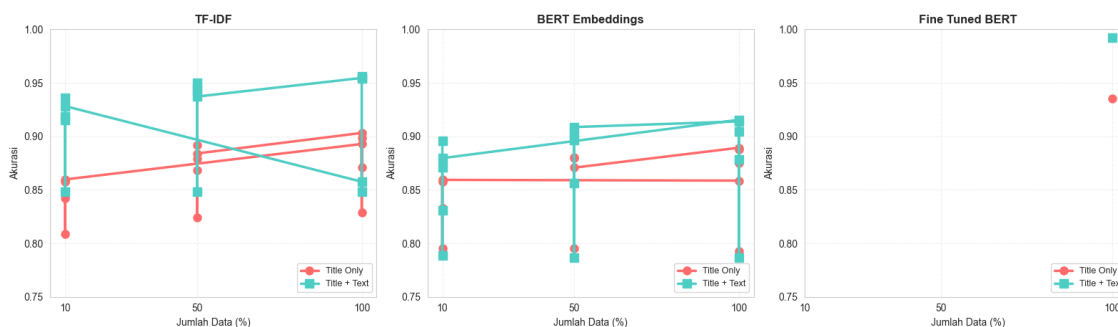


Gambar 8. Perbandingan Performa *Title Only* dan *Title + Text*

### 3.3. Pengujian Efisiensi Data melalui *Learning Curve*

Skenario ketiga menguji efisiensi model dalam memanfaatkan data. Tiga ukuran dataset diuji: 10%, 50%, dan 100% dari total *training data*.

Pengaruh Jumlah Data (10%, 50%, 100%) terhadap Akurasi Setiap Metode



Gambar 7. Perbandingan Performa Pada Setiap Persentase Penggunaan Dataset *Train*

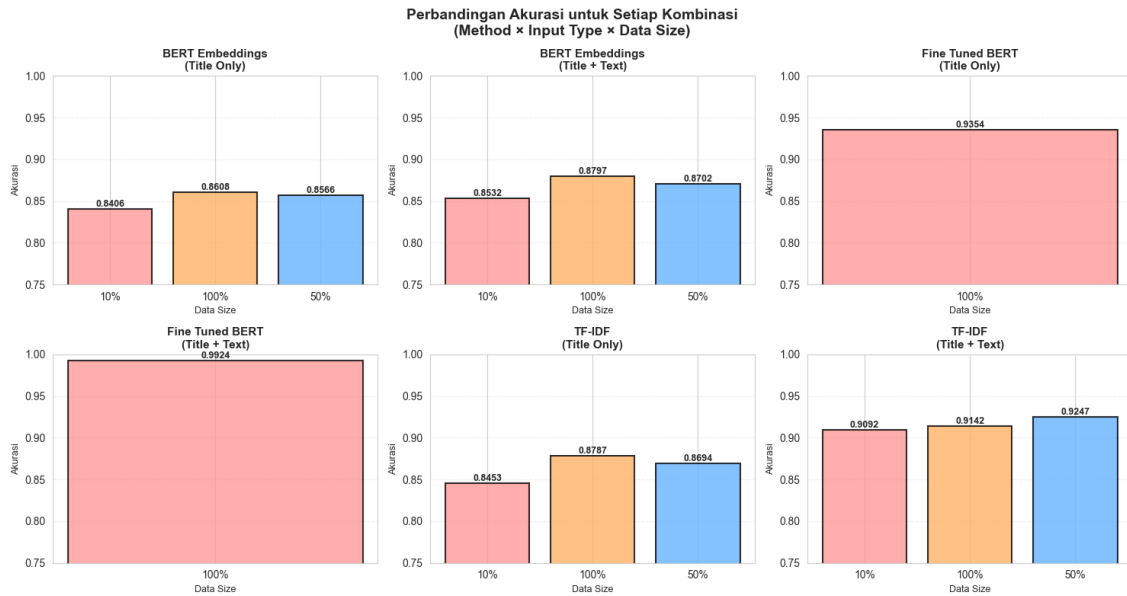
#### 3.3.1. Performa TF-IDF pada Jumlah Data yang Berbeda

TF-IDF menunjukkan perilaku unik. Model *Logistic Regression* justru mengalami penurunan performa drastis saat data mencapai 100% (turun dari 0.937 di 50% menjadi 0.858 di 100%), yang mengindikasikan masalah pada dimensi tinggi (*curse of dimensionality*) atau *noise* pada dataset penuh. Sebaliknya, model *Random Forest* dan *XGBoost* menunjukkan tren positif yang konsisten, naik dari  $\sim 0.91$  (di 10%) menjadi  $\sim 0.95$  (di 100%). Ini menegaskan bahwa model *ensemble tree* lebih *robust* menangani fitur TF-IDF berskala besar.



### 3.3.2. Performa BERT pada Jumlah Data yang Berbeda

BERT *Embeddings* menunjukkan stabilitas yang luar biasa. *Logistic Regression* meningkat stabil dari 0.896 (10%) ke 0.915 (100%). Model lain seperti *XGBoost* dan *Random Forest* juga mengalami peningkatan linear seiring bertambahnya data. Keunggulan utama BERT di sini adalah pada data sedikit (10%), model-modelnya sudah mencapai akurasi tinggi (rata-rata di atas 0.87 untuk model non-NB), menunjukkan manfaat dari *pre-trained knowledge* yang dimiliki BERT.



Gambar 8. Perbandingan Performa Setiap Kasus

### 3.3.3. Kesimpulan Pengujian Efisiensi Data melalui *Learning Curve*

BERT *Embeddings* adalah pilihan paling aman dan stabil untuk data terbatas, memberikan performa awal yang tinggi dan konsisten. Namun, jika data tersedia melimpah (100%), TF-IDF dengan *XGBoost/Random Forest* mampu mengejar dan bahkan melampaui performa BERT *Embeddings*. *Logistic Regression* pada TF-IDF justru harus diwaspadai pada data besar karena kecenderungan penurunan performa.

## 4. CONCLUSION

Penelitian ini bertujuan untuk membandingkan efektivitas dua pendekatan representasi teks, TF-IDF dan BERT, dalam mendeteksi berita palsu serta menilai kontribusi metode *ensemble Soft Voting Majority*. Berdasarkan hasil pengujian yang komprehensif, penelitian ini berhasil menjawab seluruh rumusan masalah dengan temuan empiris yang signifikan.

Pertama, komparasi fitur menunjukkan hasil yang menarik dan menantang asumsi umum. TF-IDF terbukti bukan sekadar *baseline* sederhana; ketika dikombinasikan dengan model *ensemble* seperti *Random Forest* atau *XGBoost*, representasi ini mampu menghasilkan akurasi yang sangat tinggi (~95%) pada artikel lengkap, bahkan mengungguli representasi BERT *Embeddings* standar (~91%). Superioritas BERT baru terlihat secara mutlak ketika dilakukan proses *fine-tuning* (*Fine-Tuned BERT*), yang mencapai akurasi nyaris sempurna (~99%). Hal ini mengindikasikan bahwa untuk representasi fitur statis, statistik kata (TF-IDF) masih lebih mudah dimanfaatkan oleh *classifier* tradisional dibandingkan *embedding* semantik mentah.

Kedua, pengujian granularitas informasi mengonfirmasi bahwa penambahan konteks (Isi Berita) meningkatkan performa model secara signifikan dibandingkan hanya menggunakan Judul (*Headline*). Pada skenario hanya judul, TF-IDF sedikit lebih unggul karena kemampuannya menangkap kata kunci *clickbait* secara eksplisit. Namun, *Fine-Tuned BERT* menunjukkan lonjakan performa tertinggi saat diberikan teks lengkap, membuktikan bahwa arsitektur *Transformer* membutuhkan konteks panjang untuk memaksimalkan mekanisme *self-attention*-nya.

Ketiga, analisis *learning curve* memperlihatkan karakteristik efisiensi yang berbeda. BERT *Embeddings* menunjukkan stabilitas yang luar biasa, mampu memberikan performa tinggi yang konsisten sejak penggunaan 10% data pelatihan, membuktikan efisiensi *transfer learning*. Sebaliknya, TF-IDF sangat sensitif terhadap jenis *classifier* dan jumlah data; model linear (*Logistic Regression*) justru mengalami penurunan performa pada data besar, sementara model *tree-based* terus mengalami peningkatan. Ini menegaskan bahwa TF-IDF membutuhkan algoritma *non-linear* yang kuat untuk menangani data berskala besar.

Terakhir, metode *Soft Voting Majority* terbukti berperan vital dalam menjaga konsistensi prediksi. Pada hampir semua skenario, metode ini berhasil menempati peringkat atas atau menstabilkan hasil di antara model-model tunggal, menjadikannya strategi yang efektif untuk mengurangi varians kesalahan model.

Secara keseluruhan, penelitian ini menyimpulkan:

1. *Fine-Tuned BERT* adalah solusi terbaik untuk akurasi maksimal, namun membutuhkan sumber daya komputasi pelatihan ulang.
2. TF-IDF dengan *Random Forest/XGBoost* adalah alternatif terbaik yang efisien komputasi, mampu mengungguli BERT *Embeddings* standar pada dataset ini.
3. *Soft Voting Majority* direkomendasikan sebagai standar penerapan untuk meningkatkan ketahanan sistem deteksi.

Prospek penelitian ke depan dapat diarahkan pada eksplorasi teknik efisiensi seperti *Parameter-Efficient Fine-Tuning* (PEFT/LoRA) untuk menekan biaya komputasi BERT, serta penggabungan data multimodal (teks dan gambar) mengingat berita palsu sering hadir dalam format visual. Selain itu, investigasi lebih lanjut mengenai penyebab penurunan performa model linear pada data TF-IDF berdimensi tinggi juga menjadi celah riset yang menarik.

## REFERENCES

- [1] G. Kumar and C. Dhiman, "A Comparative Study of TF-IDF, BERT and Gemma Representations for Fake News Detection," in 2024 IEEE 21st India Council International Conference (INDICON), New Delhi, India, 2024, pp. 1–6, doi: 10.1109/indicon63790.2024.10958426.
- [2] Y. B. Lasotte, E. J. Garba, Y. M. Malgwi, and M. A. Buhari, "An Ensemble Machine Learning Approach for Fake News Detection and Classification Using a Soft Voting Classifier," *European Journal of Electrical Engineering and Computer Science*, vol. 6, no. 2, pp. 1–7, Mar. 2022, doi: 10.24018/ejece.2022.6.2.409.
- [3] A. Gupta, A. Batla, C. Kumar, and G. Jain, "Comparative Analysis Of Machine Learning Models For Fake News Classification," in 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2023, pp. 1–6, doi: 10.1109/ICIEM59379.2023.10205870.
- [4] N. Capuano, G. Fenza, V. Loia, and F. D. Nota, "Deep learning for fake news detection: A comprehensive review," *Information Systems*, vol. 116, p. 102342, 2023, doi: 10.1016/j.is.2023.102342.
- [5] W. Zhang, B. C. Ooi, and Q. Yang, "Fake News Detection Based on BERT and XGBoost," in 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), Chengdu, China, 2021, pp. 1–5, doi: 10.1109/ICCCBDA51879.2021.9442566.
- [6] F. Mohr and M. Wever, "Learning curves for decision making in supervised machine learning," *Machine Learning*, vol. 113, pp. 1–45, 2024, doi: 10.1007/s10994-024-06619-7.
- [7] P. K. Verma, P. Agrawal, I. Amorim, and R. Prodan, "WELFake: Word Embedding Over Linguistic Features for Fake News Detection," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 4, pp. 881–893, Aug. 2021, doi: 10.1109/TCSS.2021.3068519.
- [8] M. Nanhamano, "Fake News Detection Using Naive Bayes and XGBoost Classifiers," in 2023 8th International Conference on Business and Industrial Research (ICBIR), Bangkok, Thailand, 2023, pp. 1–5, doi: 10.1109/ICBIR57571.2023.10197712.
- [9] I. Ahmad, "A Comparative Analysis of Machine Learning and Deep Learning Algorithms for Fake News Detection," in 2023 International Conference on Computing, Intelligence and Data Analytics (ICCIDA), Kocaeli, Turkey, 2023, pp. 1–6, doi: 10.1109/ICCIDA57710.2023.10178457.
- [10] M. Fayaz, A. Khan, and J. Rahman, "Performance Analysis of Ensemble Learning Methods for Fake News Detection," in 2023 International Conference on Digital Futures and Transformative Technologies (ICoDT2), Rawalpindi, Pakistan, 2023, pp. 1–6, doi: 10.1109/ICoDT258162.2023.10127560.
- [11] S. S. Aljameel, "Fake News Detection Using Advanced Machine Learning and Deep Learning Techniques," *IEEE Access*, vol. 12, pp. 1–12, 2024, doi: 10.1109/ACCESS.2024.10560211.
- [12] H. Zhang and B. Li, "The Shifting Landscape of Fake News Detection: A Survey," *arXiv preprint arXiv:2304.09421*, 2023. [Online]. Available: <https://arxiv.org/abs/2304.09421>(<https://arxiv.org/abs/2304.09421>).
- [13] R. K. Kaliyar, A. Goswami, and P. Narang, "Fake News Detection: A Hybrid Mechanism using BERT and CNN," in 2021 IEEE International Conference on Computing, Power and Communication Technologies (GUCON), Kuala Lumpur, Malaysia, 2021, pp. 1–5, doi: 10.1109/GUCON50781.2021.9573611.
- [14] Y. Wang and M. Zhang, "Navigating the misinformation landscape: The role of media literacy and trust in news," *Mass Communication and Society*, vol. 27, no. 2, pp. 1–22, 2024, doi: 10.1080/23808985.2024.2323736.
- [15] G. Pennycook and D. G. Rand, "The psychology of fake news," *Trends in Cognitive Sciences*, vol. 25, no. 5, pp. 388–402, May 2021, doi: 10.1016/j.tics.2021.02.007.

- [16] K. Ognyanova, D. Lazer, M. A. Baum, and R. H. Perlis, "Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power," Harvard Kennedy School Misinformation Review, vol. 1, no. 4, Jun. 2020. [Online]. Available: [<https://misinforeview.hks.harvard.edu/article/misinformation-in-action/>]
- [17] A. S. Alshahrani, "Fake News Detection Using Machine Learning Ensemble Methods," European Scientific Journal, vol. 20, no. 6, pp. 1–15, 2024, doi: 10.19044/esj.2024.v20n6p1.