# KLASIFIKASI BERITA PALSU BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA BERT

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika


disusun oleh

**AFFAN ARDANA**

**20.11.3636**


Kepada


**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

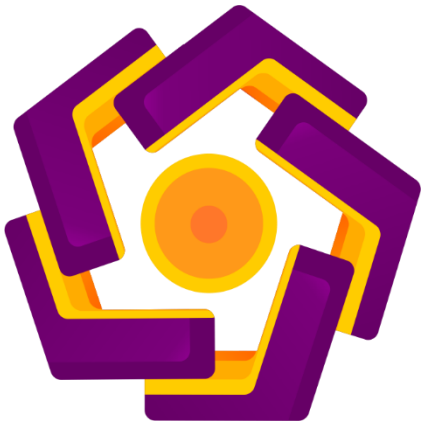

**2023**

# KLASIFIKASI BERITA PALSU BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA BERT

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika




disusun oleh

**AFFAN ARDANA**

**20.11.3636**


Kepada


**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**


**2023**

## HALAMAN PERSETUJUAN

### SKRIPSI

# KLASIFIKASI BERITA PALSU BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA BERT

yang disusun dan diajukan oleh

**Affan Ardana**

**20.11.3636**

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 November 2023

**Dosen Pembimbing,**

**Theopilus Bayu Sasongko, S.Kom, M.Eng**
**NIK. 190302375**

**HALAMAN PENGESAHAN**

**SKRIPSI**

# KLASIFIKASI BERITA PALSU BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA BERT

yang disusun dan diajukan oleh

**Affan Ardana**

**20.11.3636**

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 November 2023

**Susunan Dewan Penguji**

| Nama Penguji | Tanda Tangan |
|---|---|
| **Dr. Ferry Wahyu Wibowo, S.Si., M.Cs.**<br>**NIK. 190302235** | |
| **Ferian Fauzi Abdulloh, M.Kom**<br>**NIK. 190302276** | |
| **Theopilus Bayu Sasongko, S.Kom, M.Eng**<br>**NIK. 190302375** | |

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 November 2023

**DEKAN FAKULTAS ILMU KOMPUTER**

**Hanif Al Fatta,S.Kom., M.Kom.**
**NIK. 190302096**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

**Nama mahasiswa : Affan Ardana**
**NIM : 20.11.3636**

Menyatakan bahwa Skripsi dengan judul berikut:

**Klasifikasi Berita Palsu Berbahasa Indonesia Menggunakan Algoritma BERT**

Dosen Pembimbing : Theopilus Bayu Sasongko, S.Kom, M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 20 November 2023

Yang Menyatakan,


Affan Ardana

# KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas limpahan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan penelitian berupa skripsi ini dengan judul "**Klasifikasi Berita Palsu Berbahasa Indonesia Menggunakan Algoritma BERT**" sebagai salah satu syarat menyelesaikan Program Sarjana Informatika Universitas Amikom Yogyakarta.

Skripsi ini dapat terselesaikan dengan baik tidak lepas dari bimbingan, bantuan, dan doa dari berbagai pihak. Maka dari itu, penulis mengucapkan terima kasih kepada:

1. Allah SWT yang telah memberikan rahmat kepada penulis sehingga dapat menyelesaikan skripsi ini.
2. Bapak Theopilus Bayu Sasongko, S.Kom, M.Eng selaku dosen pembimbing yang telah memberikan arahan dan bimbingan kepada penulis dalam pengerjaan skripsi.
3. Kedua orang tua penulis yang selalu memberikan doa dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, sehingga penulis mengharapkan kritik dari pembaca. Semoga skripsi ini memberikan manfaat bagi pembaca.

Yogyakarta, 20 November 2023

Penulis

**DAFTAR ISI**

# DAFTAR TABEL

# DAFTAR GAMBAR

## DAFTAR LAMPIRAN

# INTISARI

Berita palsu menyebar dengan cepat karena akses internet yang mudah. Klasifikasi menggunakan *deep learning* dapat digunakan sebagai cara untuk mengidentifikasi sebuah berita. Arsitektur *deep learning* yang bisa digunakan untuk tugas klasifikasi adalah arsitektur *recurrent*. Namun, arsitektur *recurrent* memiliki keterbatasan yaitu pemahaman konteks yang kurang baik. Arsitektur *Transformer* dibuat untuk menangani keterbatasan arsitektur *recurrent* tersebut. Salah satu algoritma yang didasarkan pada arsitektur *Transformer* adalah BERT. Penelitian ini bertujuan untuk menggunakan BERT dalam masalah klasifikasi berita palsu berbahasa Indonesia untuk melihat hasil performanya serta perbandingan performa BERT terhadap arsitektur *recurrent*.

Dataset yang digunakan berjumlah dua dataset yang berasal dari penelitian sebelumnya. Dataset A merepresentasikan dataset kecil sedangkan Dataset B merepresentasikan dataset besar. Token yang panjangnya melebihi ketentuan dipotong menggunakan tiga jenis *truncation* yaitu *head-only*, *tail-only*, dan *head+tail*. *K-fold Cross Validation* digunakan sebagai validasi hasil performa BERT. Model *pre-trained* BERT yang digunakan untuk pelatihan adalah IndoBERT$_{\text{BASE}}$.

Hasil penelitian menunjukkan bahwa BERT menghasilkan performa terbaik pada *head-only truncation* di Dataset A serta Dataset B. BERT dengan *head-only truncation* menghasilkan akurasi sebesar 67%, *precision* sebesar 53%, *recall* sebesar 48%, dan *f-score* sebesar 45% pada Dataset A serta menghasilkan akurasi sebesar 93%, *precision* sebesar 90%, *recall* sebesar 90%, dan *f-score* sebesar 89% pada Dataset B. Perbandingan BERT dengan model *recurrent* menunjukkan bahwa LSTM dan GRU dapat mengungguli BERT dalam dataset yang kecil. Sedangkan pada dataset yang besar, BERT dapat meningkatkan akurasi sebesar 8%, *precision* 6%, dan *recall* sebesar 7% daripada LSTM.

**Kata kunci:** berita palsu, BERT, penambangan teks.

## ABSTRACT

Fake news spreads quickly due to easy internet access. Classification using deep learning can be used to identify news. The deep learning architecture that can be used for classification tasks is recurrent architecture. However, recurrent architectures have limitations, namely poor context understanding. Transformer architecture was created to address the limitations of the recurrent architecture. One of the algorithms based on Transformer architecture is BERT. This study aims to use BERT in the problem of classification of fake news in Indonesian to see the results of its performance and comparison of BERT performance against recurrent architecture.

The datasets used amounted to two datasets derived from previous research. Dataset A represents a small dataset while Dataset B represents a large dataset. Tokens that exceeded the required length were truncated using three types of truncation: head-only, tail-only, and head+tail. K-fold Cross Validation was used to validate the BERT performance results. The pre-trained BERT model used for training is IndoBERT$_{\text{BASE}}$.

The results show that BERT produces the best performance on head-only truncation on Dataset A and Dataset B. BERT with head-only truncation produces an accuracy of 67%, precision of 53%, recall of 48%, and f-score of 45% on Dataset A and produces an accuracy of 93%, precision of 90%, recall of 90%, and f-score of 89% on Dataset B. Comparison of BERT with recurrent models shows that LSTM and GRU can outperform BERT in small datasets. While on large datasets, BERT can improve accuracy by 8%, precision by 6%, and recall by 7% than LSTM.



**Keyword:** fake news, BERT, text mining.