

BAB I

PENDAHULUAN

1.1 Latar Belakang

Modernisasi di Indonesia terus mengalami perkembangan pesat dari berbagai sisi salah satunya teknologi [1]. Perkembangan teknologi berbasis internet telah membawa perubahan yang signifikan dalam proses penyampaian informasi, dimana masyarakat yang sebelumnya mendapatkan informasi melalui media cetak dan beralih ke media digital yang lebih mudah diakses melalui internet [2]. Namun dengan kemudahan yang didapatkan menimbulkan dampak yang negatif yaitu penyebaran berita hoaks, dikutip dari laman website komdigi penyebaran berita hoaks di Indonesia pada tahun 2018 sampai 2023 terhimpun 12.547 berita hoaks dari berbagai platform [3]. Penyebaran berita hoaks yang semakin meluas dapat menyebabkan masyarakat mengalami kesulitan dalam membedakan antara informasi bersifat faktual dan tidak faktual, dikarenakan informasi hoaks cenderung menyebar lebih cepat dan lebih luas dibandingkan informasi benar [4]. Oleh karena itu dibutuhkan sebuah teknologi atau sistem yang dapat mengklasifikasikan berita hoaks atau fakta.

Sistem yang menjadi pendekatan untuk tugas klasifikasi teks adalah Natural Language Processing (NLP). Perkembangan NLP modern mengalami kemajuan yang signifikan dengan diperkenalkannya arsitektur transformer, seperti model BERT, yang dapat merepresentasikan kata serta teks menjadi lebih kontekstual dengan mekanisme perhatian dua arah (*bidirectional attention*). Perlu diketahui bahwa model BERT memiliki varian yang secara khusus dilatih atau dikembangkan khusus menggunakan korpus bahasa Indonesia, yaitu IndoBERT [5].

Performa IndoBERT yang merupakan model deep learning sangat dipengaruhi oleh konfigurasi hyperparameter, khususnya learning rate dan batch size, karena learning rate salah satu hyperparameter paling penting dalam pelatihan neural network serta nilai yang tidak tepat dapat menyebabkan kegagalan

konvergensi, sedangkan batch size berinteraksi langsung dengan learning rate dalam menentukan dinamika pelatihan dan konvergensi model [6]. Dapat diketahui bahwa IndoBERT sangat sensitif terhadap hasil nilai hyperparameter, diperlukannya eksplorasi lebih lanjut dengan menemukan pola atau metode terbaik dengan tujuan menghasilkan model yang cukup stabil dalam prosesnya, kestabilan model dapat membuka ruang bagi penelitian lebih dalam pada pengoptimalan hyperparameter IndoBERT.

Penelitian ini dikembangkan untuk fokus pada metode dari pengoptimalan IndoBERT melalui fine-tuning menggunakan kombinasi dari nilai learning rate dan batch size, dengan meliputi pengumpulan data, preprocessing teks, tokenisasi, pembuatan data tensor, serta penerapan K-Fold Cross Validation. Tahapan ini diharapkan dapat menghasilkan model klasifikasi berita hoaks dan fakta yang lebih akurat, stabil, serta efektif untuk mendukung upaya pencegahan dari penyebaran berita hoaks di Indonesia dan menghadirkan masyarakat yang terbuka dengan fakta-fakta dari berita yang disebarakan.

Dari penelitian ini, pemanfaatan IndoBERT dalam pengklasifikasian berita hoaks dan fakta diharapkan mendapatkan akurasi yang tinggi serta kestabilan yang baik dalam mendeteksi berita hoaks. Hasil dari penelitian diharapkan dapat bermanfaat untuk pemerintah dalam mengurangi penyebaran berita hoaks dan membantu masyarakat agar meningkatkan kewaspadaan menerima informasi.

1.2 Rumusan Masalah

Mengoptimalkan model klasifikasi berita hoaks dan fakta berbahasa Indonesia menggunakan IndoBERT melalui fine-tuning berdasarkan konfigurasi hyperparameter learning rate dan batch size.

1.3 Batasan Masalah

1. Dataset penelitian berasal dari Kaggle dengan jumlah 3009 data.
2. Menggunakan IndoBERT base p2 sebagai model untuk proses fine-tuning.
3. Evaluasi pada model menggunakan K-Fold Validation (k5), confusion matrix, dan classification report.
4. Penelitian ini hanya berfokus dengan mencari kombinasi hyperparameter fine-tuning hanya learning rate dan batch size, selain itu ditetapkan sebagai parameter konstan atau tetap.
5. Learning rate yang digunakan berupa $1e-5$, $2e-5$, $3e-5$, $4e-5$, dan $5e-5$.
6. Batch size yang digunakan berupa 8, 16, dan 32.
7. Penelitian ini tidak berfokus pada pengaplikasian model ke sebuah perangkat lunak.

1.4 Tujuan Penelitian

1. Menerapkan serta mengevaluasi kinerja pada model IndoBERT dengan fine-tuning dalam klasifikasi berita hoaks dan fakta berbahasa Indonesia.
2. Menentukan kombinasi hyperparameter learning rate dan batch size yang optimal pada IndoBERT berdasarkan nilai akurasi, confusion matrix, serta *classification report*.

1.5 Manfaat Penelitian

1. Dapat memberikan kontribusi dalam penerapan model IndoBERT di bidang Natural Language Processing (NLP) dalam penerapan dan optimalisasi IndoBERT fine-tuning untuk klasifikasi berita hoaks dan fakta berbahasa Indonesia.

2. Model IndoBERT hasil fine-tuning dengan konfigurasi learning rate dan batch size yang optimal dapat digunakan sebagai dasar dalam pengembangan pada klasifikasi berita hoaks dan fakta berbahasa Indonesia dengan memiliki tingkat akurasi yang baik.
3. Hasil dari analisis terhadap konfigurasi hyperparameter learning rate dan batch size pada kinerja IndoBERT yang sudah dievaluasi menggunakan akurasi, confusion matrix dan classification report dapat dijadikan sebagai referensi bagi peneliti selanjutnya dalam mengembangkan serta mengoptimalkan model transformer khususnya IndoBERT untuk klasifikasi teks berbahasa Indonesia.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut.

1. **BAB I PENDAHULUAN**
Bab ini berisi uraian tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.
2. **BAB II TINJAUAN PUSTAKA**
Bab ini berisi literatur yang mencakup penelitian-penelitian terdahulu dan dasar teori yang digunakan dalam penelitian ini.
3. **BAB III METODE PENELITIAN**
Bab ini berisi penjelasan metode yang digunakan untuk menyelesaikan penelitian meliputi alur penelitian serta alat dan bahan.
4. **BAB IV HASIL DAN PEMBAHASAN**
Bab ini berisi penjelasan mengenai proses pengumpulan data, pra-pemrosesan, pembagian, pelatihan, hingga evaluasi.
5. **BAB V PENUTUP**
Bab ini berisi kesimpulan yang menjawab rumusan masalah dan diperoleh dari hasil dan pembahasan sebelumnya. Bab ini juga berisi saran yang dapat digunakan untuk mengembangkan penelitian ini.