

BAB I

PENDAHULUAN

1.1 Latar Belakang

Peraturan perundang-undangan merupakan instrumen hukum yang sangat penting dalam menjaga ketertiban sosial, politik, dan ekonomi di suatu negara. Klasifikasi peraturan perundang-undangan menjadi esensial untuk mendukung pengambilan keputusan yang cepat dan tepat, baik bagi pemerintah, praktisi hukum, maupun masyarakat umum (Petrakova, 2022). Namun, kompleksitas teks yang terkandung dalam dokumen peraturan, serta keragaman format dan istilah yang digunakan, sering kali mempersulit proses pengolahan data (Alexopoulos dkk., 2024). Hal ini terutama menjadi tantangan ketika volume dokumen yang dikelola sangat besar, sehingga membutuhkan teknologi yang dapat secara otomatis mengklasifikasikan atau mengekstraksi informasi dari teks-teks tersebut.

Kemajuan teknologi kecerdasan buatan, khususnya dalam pemrosesan bahasa alami (*Natural Language Processing/NLP*), menawarkan solusi yang efektif untuk mengatasi tantangan ini. Model *Bidirectional Encoder Representations from Transformers* (BERT), yang telah terbukti mampu menangkap makna kontekstual kata secara mendalam, menjadi salah satu model yang paling banyak digunakan dalam berbagai aplikasi NLP (Fang dkk., 2023). Namun, dengan ukuran model yang besar dan kebutuhan komputasi yang tinggi, penggunaan BERT pada skala besar dapat menjadi tidak efisien (Dong, 2024).

Penelitian mengenai penerapan BERT dan varian-*transformer* lainnya dalam domain hukum sudah banyak dilakukan. Misalnya, penelitian (Akca dkk., 2022)

yang mengembangkan model untuk klasifikasi teks dokumen hukum Turki menggunakan *transformers* dan domain *adaptation*. Penelitian lain (Haque dkk., 2022) mengembangkan *pre-trained language models* khusus untuk domain hukum India, termasuk kosa kata yang disesuaikan dengan teks hukum India. Selain itu, (Quevedo dkk., 2022) menggunakan model BERT untuk penjawaban pertanyaan terkait dokumen hukum dalam sistem *PolicyQA*. Penggunaan DistilBERT juga telah diuji pada berbagai domain, termasuk identifikasi konten kebencian dalam media sosial (Kumari dkk., 2022) dan analisis sentimen pada opini publik terkait COVID-19 (Jojoa dkk., 2024).

Berbagai studi sebelumnya, seperti yang dilakukan (Bambroo & Awasthi, 2021), menunjukkan bahwa penggunaan DistilBERT dengan *pre-training* khusus untuk domain hukum dapat meningkatkan akurasi dalam klasifikasi dokumen hukum. Penelitian ini mengusulkan penggunaan strategi *fine-tuning* yang lebih terfokus pada domain tertentu. Namun, belum ada yang mengeksplorasi penggunaan model ini dalam klasifikasi peraturan dalam jumlah besar dan beragam, apalagi dalam konteks hukum Indonesia.

Sebagai alternatif, DistilBERT, yang merupakan varian dari BERT dengan ukuran model lebih kecil dan efisiensi yang lebih baik, muncul sebagai pilihan yang menarik. DistilBERT mempertahankan sekitar 97% kinerja BERT dengan jumlah parameter 40% lebih sedikit dan kecepatan pelatihan 60% lebih tinggi dibandingkan dengan BERT, sehingga lebih cocok untuk aplikasi dengan keterbatasan komputasi dan kebutuhan pemrosesan cepat (Sanh dkk., 2019). Meskipun DistilBERT telah terbukti lebih efisien dan cepat dibandingkan BERT,

penelitian yang mengevaluasi kinerjanya dalam pengolahan dokumen peraturan perundang-undangan di berbagai negara masih sangat terbatas. Model DistilBERT perlu diuji lebih lanjut dalam konteks klasifikasi peraturan, yang menjadi fokus utama dalam dokumen perundang-undangan Indonesia, misalnya untuk membedakan status dan jenis peraturan.

Mengingat pentingnya peraturan yang terus berkembang dan diperbarui dalam konteks hukum, optimasi model DistilBERT untuk mengatasi kompleksitas ini dapat memberikan kontribusi signifikan dalam meningkatkan efisiensi pengolahan dokumen hukum. Sebagai tambahan, penelitian ini menerapkan teknik *fine-tuning* yang dioptimalkan melalui regularisasi, seperti *dropout* dan AdamW *optimizer*, serta pendekatan *multi-task learning* untuk meningkatkan akurasi klasifikasi dua label hukum secara simultan. Regularisasi merupakan bagian integral dari *fine-tuning* yang berfungsi mengontrol kompleksitas model dan meningkatkan generalisasi terhadap data baru. Penggunaan kombinasi AdamW, *dropout*, dan *multi-task learning* memungkinkan model belajar lebih efisien, mengurangi risiko *overfitting*, serta menghasilkan prediksi yang lebih akurat dan stabil dalam menghadapi keragaman dokumen hukum (Z. Liu dkk., 2023) (Xie dkk., 2022). Optimasi yang diusulkan memberikan gambaran tentang langkah-langkah teknis yang dapat digunakan untuk meningkatkan performa model.

1.2 Rumusan Masalah

Berdasarkan penjelasan latar belakang yang telah disampaikan, rumusan masalah yang diangkat dalam penelitian ini adalah sebagai berikut.

1. Bagaimana model DistilBERT yang di-*fine-tuning* dapat digunakan dalam klasifikasi status dan jenis peraturan perundang-undangan?
2. Bagaimana performa model *fine-tuned* DistilBERT dalam klasifikasi status dan jenis peraturan perundang-undangan?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini dirumuskan secara spesifik berdasarkan masalah yang telah ditentukan sebelumnya, sebagai berikut.

1. Mengimplementasikan model DistilBERT yang di-*fine-tuning* untuk klasifikasi status dan jenis peraturan perundang-undangan.
2. Mengukur performa model *fine-tuned* DistilBERT dalam klasifikasi status dan jenis peraturan perundang-undangan.

1.4 Manfaat Penelitian

Penelitian ini memiliki potensi untuk memberikan manfaat bagi para pengguna. Beberapa manfaat yang dapat diperoleh antara lain sebagai berikut.

1. Kontribusi terhadap area penelitian NLP, khususnya dalam klasifikasi peraturan perundang-undangan.
2. Kontribusi untuk pengembangan model DistilBERT.
3. Kontribusi terhadap penggunaan model *fine-tuned* DistilBERT untuk dataset peraturan perundang-undangan Indonesia.
4. Kontribusi terhadap pemahaman pengolahan data dalam klasifikasi status dan jenis peraturan perundang-undangan di Indonesia.

5. Kontribusi terhadap klasifikasi peraturan perundang-undangan yang dapat mendukung pengambilan keputusan lebih cepat dan tepat oleh pemerintah, praktisi hukum, dan masyarakat.

1.5 Batasan Masalah

Penelitian ini memiliki sejumlah batasan yang ditetapkan untuk memperjelas fokus dan mempermudah jalannya penelitian. Adapun batasan-batasan tersebut adalah sebagai berikut.

1. Penelitian ini hanya berfokus pada pengolahan dan klasifikasi peraturan perundang-undangan yang berlaku di Indonesia, menggunakan semua jenis peraturan.
2. Dataset yang digunakan dalam penelitian ini terbatas pada data peraturan perundang-undangan yang diambil dari sumber pemerintah resmi, yaitu <https://peraturan.go.id>.
3. Penelitian ini hanya mengoptimalkan model DistilBERT yang telah di-*fine-tune* untuk tugas klasifikasi peraturan perundang-undangan.
4. Fokus penelitian ini terbatas pada klasifikasi peraturan perundang-undangan, yang mencakup status hukum suatu peraturan dan jenis peraturan.
5. Penelitian ini hanya mengevaluasi model menggunakan *confusion matrix* seperti akurasi, *recall*, *precision*, dan *F1-score*.

1.6 Sistematika Penulisan

Struktur penulisan penelitian ini disusun berdasarkan panduan penulisan tugas akhir yang diterbitkan oleh Universitas Siliwangi, yang terbagi dalam lima BAB, sebagai berikut.

BAB I PENDAHULUAN

BAB I menguraikan tentang latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, serta struktur penulisan penelitian yang dilakukan.

BAB II LANDASAN TEORI

BAB II memuat teori-teori pendukung, kajian penelitian relevan, dan metrik penelitian terdahulu untuk membentuk dasar konseptual, menghasilkan *state of the art*, serta memperjelas fokus, metode, temuan, dan kontribusi ilmiah dari penelitian yang dilakukan

BAB III METODOLOGI

BAB III menjelaskan metodologi penelitian secara rinci, termasuk *roadmap* penelitian dan tahapan penelitian yang menggambarkan alur sistematis dari awal hingga akhir pelaksanaan penelitian.

BAB IV HASIL DAN PEMBAHASAN

BAB IV menyajikan hasil yang diperoleh dari proses penelitian serta pembahasan mengenai pencapaian yang telah diraih.

BAB V KESIMPULAN DAN SARAN

BAB V menyajikan kesimpulan dari seluruh rangkaian penelitian yang telah dilakukan, dengan mengulas keunggulan dan keterbatasan penelitian serta memberikan saran untuk peluang penelitian selanjutnya guna mengatasi kekurangan yang ada pada penelitian ini.