

BAB I

PENDAHULUAN

Bab ini membahas mengenai latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang digunakan dalam dokumen skripsi mengenai Prediksi Kategori Hukuman Tindak Pidana Berdasarkan Dokumen Riwayat Putusan Mahkamah Agung Menggunakan *Hierarchical* BERT dan LSTM.

1.1 Latar Belakang

Pengadilan tindak pidana di Indonesia menggunakan sistem hukum *civil law*, yang mengandalkan undang-undang tertulis sebagai sumber hukum utama dalam proses pengambilan keputusannya. Pengambilan keputusan pada pengadilan tersebut dilaksanakan oleh suatu kekuasaan kehakiman. Salah satu pelaksana kekuasaan kehakiman untuk tindak pidana umum di Indonesia adalah Mahkamah Agung (MA) Republik Indonesia (Naibaho & Hasibuan, 2021), yang memegang peran penting dalam menjatuhkan hukuman yang sesuai bagi pelaku pelanggaran hukum.

Dokumen riwayat putusan pengadilan, seperti hasil sidang dari Mahkamah Agung memiliki fungsi yang penting dalam pemilihan keputusan. Dokumen-dokumen tersebut dapat menjadi acuan untuk pertimbangan pada penjatuhan keputusan ketika terhadap kekosongan hukum, yaitu ketika tidak terdapat regulasi pada kasus tertentu. Kekosongan hukum tersebut menyebabkan keputusan hakim sebelumnya juga harus menjadi instrumen hukum untuk menjaga kepastian hukum (Simanjuntak, 2019). Sehingga analisis dokumen riwayat putusan pengadilan Mahkamah Agung memegang peran yang penting dalam proses pengadilan di Indonesia.

Analisis dokumen riwayat putusan Mahkamah Agung merupakan proses yang sulit jika jumlah dokumen yang terdapat berjumlah sangat banyak dan setiap bagian dalam putusan sidang harus diperhatikan secara seksama dalam menentukan kategori dalam suatu perkara pidana. Prediksi kategori merupakan sebuah metode pada *machine learning* untuk mengelompokkan data (seperti dokumen/kalimat) ke dalam suatu kategori yang dapat mempermudah analisis dokumen. Prediksi tersebut juga dapat disebutkan ke dalam pendekatan klasifikasi. Salah satu jenis klasifikasi tersebut

adalah klasifikasi *multiclass* (*multiclass classification*), yang merupakan pengelompokkan terhadap lebih dari dua kategori. Dokumen putusan pengadilan menggunakan metode tersebut karena memiliki lebih dari dua kategori (Nuranti dkk., 2022).

Berbagai metode telah digunakan dalam prediksi kategori untuk dokumen riwayat putusan pengadilan, masing-masing penelitian yang dilakukan memiliki beberapa perbedaan. Penelitian yang dilakukan oleh Nuranti dkk. (2022) memprediksi kategori serta lama hukuman pada dokumen riwayat putusan Mahkamah Agung Indonesia menggunakan metode CNN (*Convolutional Neural Network*), LSTM (*Long Short-Term Memory*), dan Bi-LSTM (*Bidirectional Long Short-Term Memory*) dengan variasinya masing-masing menggunakan mekanisme *attention* dan menggunakan metode *word embedding* Word2Vec. Hasil penelitian tersebut memiliki hasil akurasi yang berbeda-beda tiap modelnya, namun model yang memiliki nilai akurasi tertinggi adalah model yang dilatih dengan CNN dengan mekanisme *attention* yaitu sebesar 75,17%.

Penelitian selanjutnya yang dilakukan oleh Latisha dkk. (2024) memprediksi kategori pasal pelanggaran untuk dokumen riwayat putusan pengadilan yang sama, namun hanya menggunakan sebagian dari dokumen-dokumen tersebut. Penelitian ini dilakukan dengan *pre-training* dan *fine-tuning* dari berbagai model BERT (*Bidirectional Encoder Representations from Transformers*) dengan metode *hierarchical*. Hasil penelitian didapatkan bahwa model *hierarchical* BERT dengan *pre-training* dari IndoLEM menghasilkan nilai akurasi terbaik.

BERT merupakan salah satu model yang dapat digunakan untuk dapat melihat konteks dari sebuah kata dari dalam teks (Devlin dkk., 2018). BERT dapat digunakan untuk melakukan *word embedding* sama halnya dengan Word2Vec. BERT dapat dikembangkan lebih lagi dengan adanya penambahan arsitektur RNN (*Reccurent Neural Network*) seperti LSTM (*Long-Short Term Memory*) juga dapat meningkatkan kinerja model BERT dalam prediksi dengan pendekatan klasifikasi, seperti pada deteksi sarkasme di media sosial (Pandey & Singh, 2023). Maka dari itu, dapat disimpulkan bahwa kombinasi BERT dan LSTM dapat memahami konteks informasi yang ada pada teks serta dapat memprediksi kategori dokumen dengan baik.

Hierarchical BERT adalah salah satu metode yang digunakan dalam pemrosesan bahasa alami untuk menangani teks yang lebih panjang, seperti dokumen atau paragraf, dengan cara yang lebih efisien dan efektif. Model ini salah satu penanganan dari kekurangan model BERT, yaitu hanya dapat menangani masukan token pada jumlah sebanyak 512 (Sun dkk., 2019). Hal ini dilakukan dengan membagi-bagi teks panjang dibagi menjadi bagian-bagian yang lebih kecil, seperti kalimat atau paragraf. Setiap bagian kemudian diproses menggunakan model BERT untuk menghasilkan representasi yang kaya dan mendalam dari masing-masing bagian.

Berdasarkan beberapa penelitian yang dilakukan sebelumnya, perlu dilakukan penelitian untuk memprediksi kategori terhadap dokumen-dokumen putusan pengadilan Mahkamah Agung Indonesia menggunakan *Hierarchical BERT* dan LSTM. Harapannya strategi ini dapat meningkatkan kinerja model pada penelitian yang dilakukan oleh Nuranti dkk. (2022).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, dapat dirumuskan permasalahan yang dibahas pada penelitian skripsi ini, yaitu bagaimana hasil kinerja implementasi *Hierarchical BERT* dan model LSTM untuk memprediksi kategori hukuman tindak pidana berdasarkan dokumen riwayat putusan Mahkamah Agung.

1.3 Tujuan dan Manfaat

Tujuan umum dari penelitian dalam skripsi ini adalah dihasilkan model untuk memprediksi kategori hukuman tindak pidana berdasarkan dokumen riwayat putusan Mahkamah Agung menggunakan *Hierarchical BERT* dan LSTM. Sedangkan tujuan khusus dari skripsi ini meliputi:

1. Menganalisis pengaruh akurasi beberapa model *Hierarchical BERT* untuk *word embedding* terhadap kinerja prediksi kategori pada dokumen putusan pengadilan Mahkamah Agung (MA) Republik Indonesia.
2. Menemukan kombinasi *hyperparameter* model *Hierarchical BERT* dan LSTM untuk menghasilkan model prediksi terbaik.
3. Melihat kinerja metode LSTM dengan mekanisme *attention* dalam memprediksi kategori hukuman tindak pidana berdasarkan dokumen putusan pengadilan

Sementara manfaat yang dari penelitian skripsi ini adalah untuk meningkatkan wawasan untuk menjadi acuan pada prediksi kategori hukuman pada dokumen putusan pengadilan di Indonesia.

1.4 Ruang Lingkup

Penelitian dalam skripsi ini memerlukan adanya ruang lingkup agar penelitian yang dilakukan lebih terarah dan mencapai sasaran yang diharapkan. Ruang lingkup dalam skripsi ini adalah sebagai berikut:

1. Data pada penelitian ini merupakan kumpulan putusan pengadilan Indonesia yang mencakup kasus perkara pidana umum dan pidana khusus. Dataset ini terdiri atas 22.630 dokumen yang telah dianotasi berdasarkan hasil identifikasi bagian dokumennya dan hanya menggunakan sebanyak empat bagian tersebut yang merupakan riwayat tuntutan, fakta, fakta hukum, dan pertimbangan hukum. Pemilihan bagian dokumen ini dilakukan berdasarkan penelitian yang dilakukan oleh Nuranti dkk. (2022).
2. Labelisasi kategori ditentukan ke dalam empat buah kelas untuk setiap kategori hukuman tindak pidana pada putusan pengadilan Mahkamah Agung yaitu *mild*, *moderate*, *heavy*, dan *very heavy*.

1.5 Sistematika Penulisan

Gambaran mengenai pembahasan penyusunan laporan skripsi ini secara urut dan jelas, maka dibentuk sistematika sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian, dan sistematika penulisan dari penelitian dengan judul Prediksi Kategori Hukuman Tindak Pidana Berdasarkan Dokumen Riwayat Putusan Mahkamah Agung Menggunakan *Hierarchical* BERT dan LSTM.

BAB II LANDASAN TEORI

Bab ini memuat tentang tinjauan pustaka dan dasar teori yang digunakan dalam skripsi berjudul Prediksi Kategori Hukuman Tindak Pidana

Berdasarkan Dokumen Riwayat Putusan Mahkamah Agung Menggunakan *Hierarchical* BERT dan LSTM. Bab ini terdiri atas beberapa subbab antara lain *state of the art*, pembersihan *dataset*, pra-pemrosesan data, BERT, *hierarchical* BERT, LSTM, *fully connected layer*, fungsi aktivasi dan evaluasi.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai tahap-tahap penelitian yang dilakukan dalam skripsi Prediksi Kategori Hukuman Tindak Pidana Berdasarkan Dokumen Riwayat Putusan Mahkamah Agung Menggunakan *Hierarchical* BERT dan LSTM. Gambaran umum dari tahapan penelitian ini adalah pengumpulan data, pengubahan format data, pemilihan bagian dokumen, pra-pemrosesan data, pembagian data, BERT, *hierarchical* BERT, pelatihan model, *hyperparameter optimization* dan evaluasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan mengenai hasil dan analisis yang dilakukan sesuai metodologi yang digunakan dalam skripsi Prediksi Kategori Hukuman Tindak Pidana Berdasarkan Dokumen Riwayat Putusan Mahkamah Agung Menggunakan *Hierarchical* BERT dan LSTM. Bab ini membahas tentang lingkungan dan perangkat yang digunakan untuk pengujian, skenario pengujian, serta hasil dan analisis dari skenario.

BAB V PENUTUP

Bab ini berisi kesimpulan dari bab-bab yang telah dijelaskan sebelumnya dan saran untuk penelitian lebih lanjut.