

ABSTRAK

Analisis dokumen putusan pengadilan menjadi tantangan besar ketika jumlah dokumen sangat banyak dan setiap bagian dari putusan sidang harus diperhatikan secara detail untuk menentukan lama hukuman dalam suatu perkara pidana. Penelitian ini mengusulkan pendekatan regresi sebagai solusi. Meskipun regresi untuk prediksi lama hukuman dalam dokumen pengadilan belum banyak diteliti, terdapat penelitian sebelumnya memprediksi kategori dan lama hukuman berdasarkan riwayat sidang yang telah diarsipkan dalam dokumen MA dengan menggunakan metode *Word2Vec* dan LSTM. Namun, penelitian ini menghadapi beberapa keterbatasan. Salah satunya adalah ketidakcocokan hasil *word embedding Word2Vec* dengan teks yang memiliki banyak kemiripan, yang disebabkan oleh penggunaan frasa dan format berulang dalam dokumen persidangan di Indonesia yang menyebabkan LSTM tidak efektif saat dikombinasikan dengan *Word2Vec*, karena *Word2Vec* menggunakan *word embedding* yang sama untuk kata-kata yang serupa dalam konteks yang berbeda. Berdasarkan kelemahan tersebut, penelitian ini mengajukan pendekatan regresi dengan menggunakan model BERT yang dikombinasikan dengan LSTM. Karena model BERT menggunakan teks riwayat sidang yang panjang, penelitian ini menerapkan *Hierarchical BERT*. *Dataset* yang digunakan mencakup 22.630 dokumen putusan pengadilan dari kasus pidana umum dan khusus. Yang memuat bagian hukum berisi riwayat tuntutan, fakta, fakta hukum, dan pertimbangan hukum. Penelitian ini menggunakan *hyperparameter* IndoBERT-base-uncased dengan *batch size* 32; *learning rate* 0,0001; *hidden size* 128; *num layers* 1; dan *epoch* 20 serta Google-BERT/bert-base-multilingual-uncased dengan *batch size* 32; *learning rate* 0,0001; *hidden size* 256; *num layers* 2; dan *epoch* 20. Hasilnya menunjukkan bahwa model BERT yang dikombinasikan dengan LSTM tanpa mekanisme *Attention* menghasilkan performa terbaik dengan *R2 Score* sebesar 0,5006. Sementara itu, kombinasi BERT dengan LSTM menggunakan mekanisme *Attention* memberikan *R2 Score* sebesar 0,4927. Meski sedikit lebih rendah, penelitian ini memberikan kontribusi dalam pengembangan regresi dokumen hukum, khususnya memprediksi lama hukuman pada dokumen putusan pengadilan Mahkamah Agung.

Kata kunci: Dokumen Hukum, Prediksi Lama Hukuman, *Hyperparameter*, BERT (*Bidirectional Encoder Representations from Transformers*), LSTM (*Long Short-Term Memory*)