

## INTISARI

### MODEL PREDIKSI LAMA DAN KATEGORI HUKUMAN DALAM PUTUSAN PENGADILAN DI INDONESIA MENGGUNAKAN *TEXT REPRESENTATION* (NLP) DAN ALGORITMA RANDOM FOREST

Oleh  
NASUTION  
23/524428/PPA/06563

Sistem peradilan di Indonesia menghadapi tantangan besar akibat meningkatnya jumlah perkara pidana yang mencapai 288.472 kasus pada tahun 2023, naik 15% dibanding tahun sebelumnya. Tingginya volume putusan hukum membuat hakim dan praktisi hukum sulit meninjau dokumen secara menyeluruh, sehingga muncul kebutuhan akan teknologi prediksi berbasis kecerdasan buatan untuk meningkatkan efisiensi, transparansi, dan konsistensi. Penelitian terdahulu yang menggunakan pendekatan deep learning seperti BiLSTM dan CNN dengan mekanisme attention masih menghasilkan akurasi relatif rendah (49,14%–77,32%) serta menghadapi kendala distribusi kelas tidak seimbang dan kebutuhan komputasi tinggi.

Penelitian ini mengusulkan model prediksi putusan pengadilan berbasis *Natural Language Processing* (NLP) dan algoritma *Random Forest* (RF) sebagai alternatif yang lebih ringan secara komputasi, mudah diinterpretasikan, namun tetap kompetitif dalam akurasi. Model ini menggunakan representasi teks dengan TF-IDF, Word2Vec, dan FastText, serta dievaluasi dengan dua target: prediksi lama hukuman (regresi) dan kategori hukuman (klasifikasi). Eksperimen dilakukan dengan pendekatan agregatif dan segmentatif, dilengkapi variasi n-gram, seleksi fitur, serta *hyperparameter tuning* untuk mendapatkan konfigurasi terbaik sesuai karakteristik bahasa hukum Indonesia.

Hasil penelitian menunjukkan bahwa model terbaik untuk klasifikasi kategori hukuman mencapai akurasi 81% dengan TF-IDF. Pada regresi lama hukuman, model terbaik menghasilkan nilai  $R^2$  sebesar 79%, menandakan kemampuan prediksi yang cukup baik terhadap durasi hukuman. Temuan ini membuktikan bahwa kombinasi NLP dan *Random Forest* dapat memberikan kontribusi signifikan dalam analisis putusan pengadilan di Indonesia serta mendukung penerapan teknologi prediktif yang lebih efisien dan praktis dalam sistem peradilan.

**Kata kunci:** *Natural Language Processing (NLP)*, *Machine Learning (ML)*, *Random Forest (RF)*, Prediksi putusan pengadilan

## ABSTRACT

### ***A PREDICTIVE MODEL FOR THE LENGTH OF PUNISHMENT AND PUNISHMENT CATEGORY IN INDONESIAN COURT DECISIONS USING TEXT REPRESENTATION (NLP) AND THE RANDOM FOREST ALGORITHM***

by

NASUTION

23/524428/PPA/06563

*The Indonesian judicial system is facing significant challenges due to the growing number of criminal cases, which reached 288,472 in 2023 — a 15% increase on the previous year. The large quantity of legal documents complicates the decision-making process, highlighting the need for AI-based predictive models to improve efficiency, transparency, and consistency. Previous studies using deep learning models, such as BiLSTM and CNN with attention mechanisms, achieved relatively low accuracy (49,14%–77,32%) and encountered issues such as class imbalance and high computational demands.*

*This study, therefore, proposes an alternative predictive model that combines natural language processing (NLP) with the random forest (RF) algorithm. This approach uses TF-IDF, Word2Vec, and FastText to represent text, and is applied to a dataset of 21,630 Indonesian court decisions. Two predictive tasks were considered: regression to estimate sentence duration, and classification to determine sentence categories. Both aggregate and segment-based approaches were examined, along with experiments involving n-gram variations, feature selection, and hyperparameter tuning.*

*The results of the experiments demonstrate that the best classification model achieved an accuracy of 81% when using TF-IDF with n-grams, while the regression model obtained an  $R^2$  score of 79%, indicating strong predictive capability. These results confirm that combining NLP with the RF algorithm provides a robust, efficient, and interpretable framework for analysing Indonesian court decisions and could contribute to the development of more practical and accountable legal analytics.*

**Keywords**— *Natural Language Processing, Random Forest, Machine Learning, Court decision prediction*