

Klasifikasi Jenis Kejahatan berdasarkan Teks Amar Putusan Pengadilan Hukum Pidana KUHP menggunakan IndoBERT

Tirtanusa Kurnia Adhi Perdana¹, Dewi Soyusiawaty^{1,*}

¹ Program Studi Informatika, Universitas Ahmad Dahlan, Indonesia

* Correspondence: dewi.soyusiawaty@tif.uad.ac.id

Copyright: © 2025 by the authors

Received: 16 Mei 2025 | Revised: 11 Juni 2025 | Accepted: 2 Juli 2025 | Published: 12 Agustus 2025

Abstrak

Peningkatan jumlah putusan pengadilan setiap tahun menjadi salah satu tantangan bagi lembaga peradilan. Salah satu solusi strategis adalah penerapan *Artificial Intelligence* (AI). Model berbasis Bahasa Indonesia seperti IndoBERT berpotensi untuk meringankan beban kerja seperti mengelompokkan kasus hukum secara otomatis. Penelitian ini bertujuan untuk mengeksplorasi kemampuan IndoBERT dalam mengklasifikasikan teks amar putusan pidana KUHP secara otomatis untuk mempercepat identifikasi jenis kejahatan. Jenis penelitian ini adalah eksperimen klasifikasi teks berbasis *supervised learning*. Variabel yang diteliti berupa teks amar putusan berjumlah 12000 entri yang diambil dari website Mahkamah Agung diklasifikasi menggunakan IndoBERT yang di-*fine-tune* dengan berbagai konfigurasi *hyperparameter*. Hasil temuan kami menunjukkan model dengan 8 *batch* dan $5e-5$ *learning rate* mencapai akurasi 92,59%, presisi 92,93%, *recall* 92,59% dan *f1-score* 92,59% terhadap data yang belum pernah dilihat saat training. Akurasi tinggi didukung oleh penyebutan eksplisit jenis kejahatan didalam teks amar putusan. Hingga kini, belum ditemukan penelitian yang menggunakan IndoBERT atau model lain untuk klasifikasi pasal KUHP secara otomatis. Temuan ini berpotensi untuk ditambahkan kedalam sistem Direktori Putusan sebagai alat bantu klasifikasi otomatis dan penataan arsip hukum digital.

Kata kunci: analisis teks putusan pengadilan; indobert; klasifikasi pasal hukum pidana; natural language processing dalam hukum; teks hukum

Abstract

The increasing number of the court's rulings each year presents a challenge for the judiciary. One strategic solution is the application of Artificial Intelligence (AI). Indonesian-based models such as IndoBERT are potential to ease workloads by automatically classifying legal cases. This study aims to explore the capability of IndoBERT to automatically classifying the verdict of section of Indonesian KUHP rulings to accelerate crime type identification. This is an experimental study using supervised text classification. The dataset consists of 12000 verdicts collected from the Indonesian Supreme Court website, classified using IndoBERT fine-tuned with various hyperparameter configuration. Our findings show that the model with a batch size of 8 and learning rate $5e-5$ achieved accuracy of 92.59%, precision of 92.93%, recall of 92.59%, and F1-Score of 92.59% on unseen test data. The high accuracy is supported by the explicit mention of crime types within verdict texts. To date, no study has specifically utilized IndoBERT or other models for automatic classification of KUHP articles. This finding has the potential to be integrated into the Supreme Court's Directory of Decision as a support tool for automatic classification and legal document archiving.

Keywords: *judicial decision text analysis; indobert; criminal law article classification; natural language processing in law; legal text*



PENDAHULUAN

Revolusi Industri 4.0 telah membawa transformasi besar dalam berbagai bidang, termasuk sektor hukum (Yasir & Gunawan, 2024). Salah satu inovasi utama dalam revolusi ini adalah berkembangnya kecerdasan buatan (*Artificial Intelligence/AI*), khususnya dalam bidang pemrosesan bahasa alami (*Natural Language Processing/NLP*), yang memungkinkan mesin untuk memahami, menganalisis, dan menghasilkan teks secara otomatis. Kemunculan model bahasa generatif seperti ChatGPT merupakan terobosan besar dalam teknologi AI berbasis teks ditandai dengan satu juta pengguna di awal peluncuran (Deng et al., 2023). Seperti halnya kehadiran smartphone yang merevolusi akses informasi di era 2010-an, AI diproyeksikan akan mengubah cara manusia bekerja dan mengakses informasi (Fahrudin et al., 2025).

Mahkamah Agung mencatat lebih dari 10.000 putusan dikeluarkan sepanjang tahun 2024, sebagian besar merupakan perkara pidana umum yang mengacu pada Kitab Undang-Undang Hukum Pidana (KUHP). Sebagai dasar utama hukum pidana di Indonesia, KUHP mencakup berbagai jenis kejahatan dan menjadi rujukan dalam putusan pengadilan. Volume putusan yang tinggi kerap menimbulkan beban administratif bagi hakim dan beresiko meningkatkan human bias dalam pengambilan keputusan (Rohsawati et al., 2023). Selain itu, inkonsistensi dalam putusan dan kecenderungan menghindari kasus politis masih menjadi masalah sistemik. Sistem peradilan modern sejatinya merupakan industri pemrosesan teks terbesar, sehingga menjadi bidang yang potensial untuk penelitian teknologi NLP (*Natural Language Processing*). Salah satu kemampuan utama NLP adalah mengekstrak informasi penting dan fitur relevan dari dokumen hukum, termasuk putusan pengadilan yang dapat membantu praktisi hukum dalam menjalankan tugasnya (Pakpahan, 2021; Wu et al., 2025; Zhang & Lu, 2021).

Klasifikasi teks, sebagai salah satu tugas fundamental dalam NLP kerap digunakan untuk membagi dokumen kedalam kategori tertentu (Rahma & Suadaa, 2023). Metode ini dapat mendukung pengguna dalam mengambil keputusan, meminimalkan bias, dan meningkatkan konsistensi (Khoirunnisaa et al., 2024; Yoo, 2021). Perkembangan arsitektur NLP modern seperti *Transformers* telah memungkinkan munculnya model-model yang sangat kuat seperti BERT (Bidirectional Encoder Representation from Transformers). Salah satu keunggulan BERT adalah kemampuannya dalam memahami konteks secara dua arah (*bidirectional*), yaitu dengan memproses teks dari kanan ke kiri dan dari kiri ke kanan secara bersamaan, sehingga mampu menangkap makna kata dengan lebih akurat. Model ini secara signifikan mengungguli pendekatan NLP tradisional seperti *Logistic Regression* atau *Linear SVC* dalam berbagai tugas pemrosesan teks (Garrido-Merchan et al., 2023). Namun demikian, model BERT memiliki keterbatasan dalam memahami bahasa Indonesia karena umumnya dilatih menggunakan data berbahasa Inggris. Salah satu keunggulan BERT adalah kemampuannya dalam memahami konteks secara dua arah (*bidirectional*), yaitu dengan memproses teks dari kanan ke kiri dan dari kiri ke kanan secara bersamaan, sehingga mampu menangkap makna kata dengan lebih akurat.

Model IndoBERT sebagai turunan dari BERT merupakan salah satu alternatif yang menjanjikan dalam pengolahan dokumen hukum berbahasa Indonesia (Budiman et al., 2024). IndoBERT adalah model *transformer* yang secara khusus dilatih dengan korpus berbahasa Indonesia berjumlah sekitar 220 juta kata (Koto et al., 2020). Dengan pelatihan yang mengacu pada sumber data berbahasa Indonesia, IndoBERT mampu mengatasi keterbatasan BERT dalam memahami konteks linguistik Bahasa Indonesia. IndoBERT menunjukkan keunggulan dalam tugas klasifikasi teks berbahasa Indonesia dibanding model pendahulunya (Juarto & Yulianto, 2023; Nabiilah et al., 2024). Penelitian oleh (Maulana & Aditya, 2025) menunjukkan bahwa IndoBERT mencapai akurasi 90% dalam mendeteksi pelanggaran berdasarkan UU ITE. Hasil lebih baik didapat ketika IndoBERT digabungkan dengan arsitektur Bi-LSTM. Namun penelitian tersebut belum menunjukkan efektivitas IndoBERT dalam mengklasifikasikan jenis

pelanggaran dalam konteks hukum pidana KUHP. Walaupun menunjukkan performa yang baik, penelitian tersebut tidak mengeksplorasi jenis kejahatan dalam konteks hukum pidana KUHP. Struktur bahasa yang berbeda antara dataset pelanggaran UU ITE dengan teks amar putusan pengadilan menjadi tantangan tersendiri bagi model.

Penelitian kami bertujuan untuk menguji kemampuan IndoBERT dalam menangani kompleksitas bahasa hukum, serta mengevaluasi kelayakannya dalam menangani kompleksitas teks hukum sebagai baseline model untuk penelitian klasifikasi otomatis untuk teks hukum seperti amar putusan pengadilan. Penelitian ini diharapkan dapat berkontribusi dalam mempercepat proses pencarian dan pemrosesan dokumen hukum, sekaligus meningkatkan efisiensi kerja lembaga peradilan secara keseluruhan.

METODE

Model yang akan digunakan pada penelitian ini adalah IndoBERT. Sebuah model yang dilatih menggunakan arsitektur BERT untuk teks berbahasa Indonesia. Penelitian ini akan melewati beberapa tahap meliputi: *data scraping*, *data preprocessing*, tokenisasi dengan AutoTokenizer, pelatihan model menggunakan pre-trained model IndoBERT yang kemudian akan di *fine-tune* untuk mengeksplorasi potensi dari model pre-trained IndoBERT, dan evaluasi model.

Tahap pertama adalah data *scraping* dari situs resmi Direktori Putusan Mahkamah Agung (<https://putusan3.mahkamahagung.go.id>) yang tersedia secara terbuka. Data yang dikumpulkan merupakan putusan pidana umum dengan kategori kejahatan terhadap keamanan negara, pemalsuan, pembunuhan, penganiayaan, penggelapan, penghinaan, penipuan, dan perjudian dengan masing-masing kategori diambil 1000 entri, sehingga total data yang akan di proses adalah sekitar 8000 data. Kategori tersebut dipilih karena tidak semua kategori kejahatan dalam KUHP memiliki entri lebih dari 1000 putusan serta untuk menjaga keseimbangan antar kelas dan mencegah overfitting.

Data *scraping* dilakukan secara otomatis menggunakan skrip python memanfaatkan *library* Scrapy dan disimpan dalam format CSV untuk mempermudah proses data *pre-processing* menggunakan *library* Pandas. Proses *scraping* dilakukan sesuai dengan kategori yang terdapat pada website putusan sehingga mempermudah ketika pelabelan. Seluruh skrip Python yang digunakan untuk *scraping*, *preprocessing* dan pelatihan model saat ini disimpan secara lokal. Namun, kami tidak menutup kemungkinan untuk menggunggahnya ke repositori online.

Pada tahap *preprocessing*, data melalui beberapa proses seperti penghapusan data ganda (duplikat), NaN, karakter spesial atau simbol selain tanda baca, entri yang tidak sesuai dengan kategori, serta *whitespace* (Furqon & Soyusiawaty, 2025). Penggunaan IndoBERT atau model lain yang menggunakan arsitektur transformers tidak memerlukan *pre-processing* yang berat seperti penghilangan stopwords atau stemming yang berpotensi mempersulit model dalam mempelajari data karena menghilangkan konteks dari kalimat atau kata (Denk & Ramallo, 2020; Khairani et al., 2024). Proses *pre-processing* diakhiri dengan pelabelan data sesuai dengan asal kategori data tersebut.

Proses pelatihan dimulai dengan mengubah label *string* ke bentuk *integer* dengan angka (0-7), lalu dibagi dengan proporsi 80:20 menggunakan metode hold-out tanpa teknik stratifikasi tambahan dengan memastikan seluruh kategori sudah seimbang. Proses pelatihan dilakukan menggunakan model Pre-Trained IndoBERT, kemudian model tersebut di-*fine-tune* menggunakan *Transformer* dari HuggingFace dengan PyTorch. Pelatihan dilakukan selama 5 epoch dengan *batch size* sebesar 8 dan 16. Optimizer yang digunakan adalah Adafactor dengan beberapa konfigurasi *learning rate* seperti $1e-5$, $2e-5$ dan $5e-5$. *Evaluasi* dan *Logging* dilakukan setiap akhir epoch dengan metrik *eval_loss* dan model terbaik disimpan secara otomatis di akhir pelatihan.

Langkah terakhir adalah evaluasi model untuk mengetahui seberapa baik model menyelesaikan tugas klasifikasi. Set tes berisi data yang tidak pernah dilihat sebelumnya oleh model untuk menguji kemampuan model dalam klasifikasi teks hukum. Model akan dinilai menggunakan metrik penilaian seperti akurasi, presisi, recall, dan F1 score. Dalam konteks ini, presisi dan akurasi lebih diprioritaskan karena penting untuk model memberikan prediksi yang tepat dalam mengkategorikan, sementara recall tetap diperhatikan untuk memastikan tidak ada kategori yang penting terlewat. Confusion matrix juga akan disajikan sebagai pelengkap yang digunakan untuk memberi gambaran distribusi klasifikasi antar kategori dan kesalahan klasifikasi.

HASIL DAN PEMBAHASAN

Hasil

Penelitian ini menggunakan dataset yang diperoleh melalui proses *web scraping* dari situs Direktori Putusan Mahkamah Agung Republik Indonesia, dengan fokus pada delapan kategori perkara pidana umum, yaitu: kejahatan terhadap keamanan negara, pemalsuan, pembunuhan, penganiayaan, penggelapan, penghinaan, penipuan, dan perjudian. Untuk mengantisipasi kemungkinan adanya data yang tidak valid, mengandung noise atau tidak sesuai dengan kategori, proses pengambilan data dilakukan secara berlebih hingga sekitar 1500 entri per kategori. Dengan demikian, total data awal yang dikumpulkan mencapai kurang lebih 12.000 per entri .

Setelah proses *scraping*, seluruh data diberi label sesuai dengan kategori asalnya berdasarkan taksonomi yang digunakan pada situs resmi. Namun, pada tahap *preprocessing*, ditemukan ketimpangan distribusi data. Kategori penghinaan mengalami penurunan signifikan karena sebagian besar entri tidak memuat unsur penghinaan, melainkan jenis kejahatan lain. Selain itu, ditemukan pula entri kosong. Hal ini kemungkinan disebabkan oleh inkonsistensi pelabelan dokumen pada situs Mahkamah Agung.

Ketidakseimbangan data diatasi dengan teknik *oversampling* pada kategori kejahatan terhadap keamanan negara dan penghinaan, serta *downsampling* untuk kategori lainnya agar seimbang. Teknik *random sampling* digunakan untuk menjaga distribusi tetap merata. Namun pendekatan ini beresiko *overfitting* karena terjadi perulangan data ketika melakukan *oversampling*, dan kehilangan informasi penting pada *downsampling*. Teknik seperti Synthetic Minority Oversampling Technique (SMOTE) disarankan untuk penelitian mendatang agar data sintetis yang dihasilkan bervariasi dan representatif.



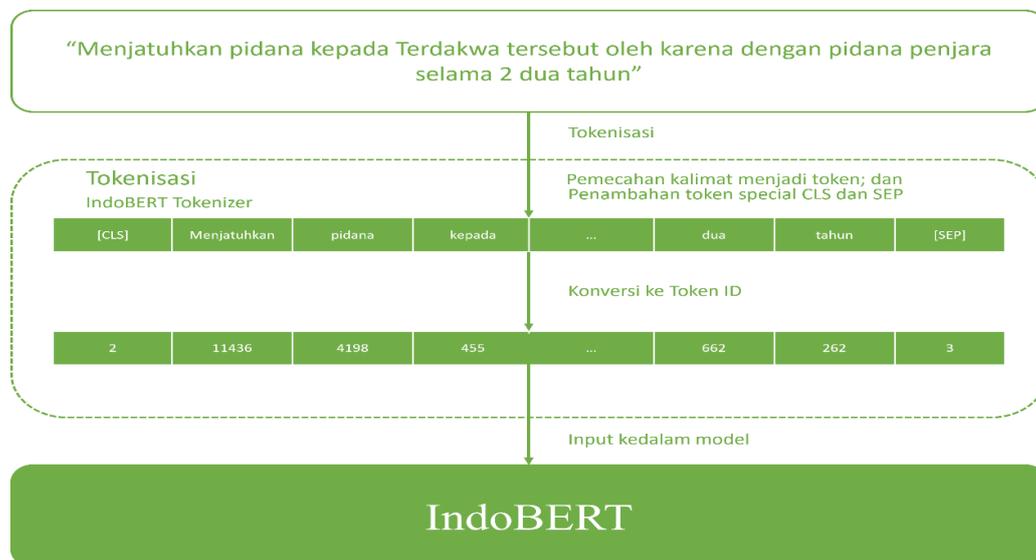
Gambar 1. Persebaran data setelah resampling

Gambar 1 menunjukkan data yang telah melalui proses resampling, dengan jumlah akhir sebanyak 1300 entri untuk masing-masing kategori, kemudian dibagi menjadi data pelatihan dan data pengujian dengan rasio 80:20. Sebelum digunakan, seluruh data diproses menggunakan tokenizer milik IndoBERT untuk mengonversi teks ke dalam format token ID yang dapat dipahami oleh model *transformer*.

Tabel 1. Hasil *preprocessing*

Sebelum	Sesudah
Menyatakan Terdakwa EDY SUSANTO Bin AGUS SUPARMAN yang identitasnya lengkap sebagaimana tersebut di atas terbukti secara sah dan meyakinkan bersalah melakukan tindak pidana Penipuan "; Menjatuhkan pidana kepada Terdakwa tersebut oleh karena dengan pidana penjara selama 2 (dua) tahun; Menyatakan masa lamanya terdakwa ditangkap dan ditahan dikurangkan seluruhnya dari pidana yang dijatuhkan;	Menyatakan Terdakwa EDY SUSANTO Bin AGUS SUPARMAN yang identitasnya lengkap sebagaimana tersebut di atas terbukti secara sah dan meyakinkan bersalah melakukan tindak pidana Penipuan; Menjatuhkan pidana kepada Terdakwa tersebut oleh karena dengan pidana penjara selama 2 dua tahun; Menyatakan masa lamanya terdakwa ditangkap dan ditahan dikurangkan seluruhnya dari pidana yang dijatuhkan;

Tabel 1 menunjukkan hasil dari pemrosesan pada teks amar putusan yang digunakan dalam penelitian ini. Tidak seperti pendekatan biasanya dalam pemrosesan teks, model IndoBERT tidak memerlukan *preprocessing* yang bersifat agresif seperti penggunaan metode *stemming* atau penghapusan *stopwords*, karena langkah tersebut berpotensi menghilangkan konteks penting dalam dokumen. Oleh karena itu, *preprocessing* untuk model IndoBERT difokuskan pada pembersihan elemen-elemen yang tidak relevan atau tidak sesuai dengan kaidah penulisan formal, tanpa mengubah struktur atau substansi teks.

**Gambar 2.** Visualisasi tokenisasi menggunakan indobert tokenizer

Selanjutnya pada gambar 2 merupakan proses tokenisasi oleh IndoBERT. Kalimat dipecah menjadi bagian kecil, kemudian ditambahkan token khusus [CLS] pada awal dokumen, dan [SEP] di akhir. Panjang token dibatasi hingga 256 token untuk menyesuaikan dengan keterbatasan sumber daya komputasi yang tersedia, sekaligus mempercepat waktu pelatihan model. Meskipun dokumen hukum pada umumnya memiliki panjang yang bervariasi dan sering kali melebihi 512 token (batas maksimum panjang *sequence* untuk IndoBERT). Penelitian di masa mendatang dapat mengeksplorasi teknik pemrosesan dokumen panjang yang lebih kompleks, seperti penggunaan model *hierarchical transformer* untuk dokumen panjang. Kemudian setiap token dikonversi menjadi ID numerik berdasarkan kamus IndoBERT.

Tokenisasi penting dilakukan agar IndoBERT dapat memahami struktur dan konteks kalimat secara sistematis.

Berdasarkan data hasil *training* dengan model *pre-trained* IndoBERT pada tabel 2, model memiliki akurasi yang sangat baik dengan akurasi tertinggi terjadi pada *epoch* 5 yaitu 89,1%. Akurasi meningkat dengan stabil walaupun model menunjukkan indikasi kecil terjadinya *overfitting*, ditandai dengan naiknya *validation loss* pada *epoch* 4 dan naik lagi pada *epoch* 5. Namun model masih efektif dalam mengklasifikasikan data dengan benar.

Selanjutnya, dilakukan *fine-tuning* menggunakan 8 dan 16 *batch*, *learning rate* 1e-5, 2e-5 dan 5e-5. *Optimizer* Adafactor digunakan untuk memperkecil *overfitting* yang terjadi dan membuat model lebih stabil selama masa pelatihan. *Optimizer* Adafactor bekerja dengan mengatur pembaruan parameter secara adaptif di selama pelatihan tanpa menyimpan histori gradien parameter, sehingga membuat pelatihan lebih efisien terhadap memori.

Tabel 2. Hasil *Training* menggunakan model pretrained indobert

<i>Epoch</i>	<i>Training Loss</i>	<i>Validation Loss</i>	Akurasi
1	0,77	0,46	85,86%
2	0,46	0,43	86,21%
3	0,36	0,40	88,12%
4	0,27	0,42	88,86%
5	0,18	0,49	89,1%

Tabel 3. *Fine tuning* indobert

<i>Batch</i>	<i>Learning Rate</i>	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
8 Batch	1e-5	88,02%	89,02%	88,02%	88,01%
	2e-5	88,46%	89,04%	88,46%	88,48%
	5e-5	90,67%	90,78%	90,67%	90,69%
16 Batch	1e-5	88,5%	88,98%	88,5%	88,4%
	2e-5	90,19%	90,03%	90,19%	90,01%
	5e-5	89,95%	90,23%	89,95%	90,03%

Tabel 3 menunjukkan bahwa kombinasi *batch size* 8 dan *learning rate* 5e-5 menghasilkan performa terbaik dengan akurasi 90,67% dan presisi 90,78%, lebih tinggi dibanding kombinasi lainnya. Model ini mampu meminimalkan kesalahan klasifikasi, yang penting dalam konteks teks hukum ditandai dengan presisi yang tinggi. Model juga mampu mengenali detail penting pada data ditandai dengan F1-Score yang tinggi.

Tabel 4. *Classification report* 8 batch learning rate 5e-5

Kategori	Presisi	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Perjudian	97%	98%	98%	271
Penggelapan	80%	85%	82%	259
Penganiayaan	94%	95%	94%	257
Penipuan	83%	78%	81%	274
Pembunuhan	94%	89%	91%	265
Pemalsuan	84%	89%	87%	251
Kejahatan Terhadap Keamanan Negara	98%	96%	97%	254
Penghinaan	96%	96%	96%	249

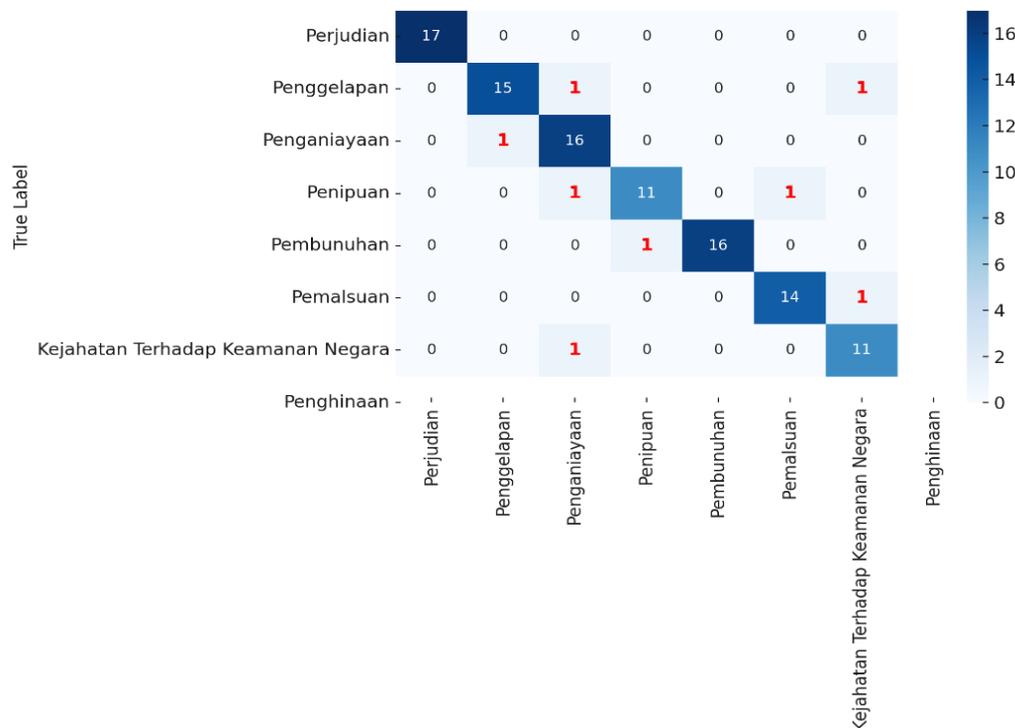
Tabel 4 menunjukkan kategori seperti perjudian dan kejahatan terhadap keamanan negara memiliki presisi tertinggi dengan 97% dan 98%. Sementara itu, kategori seperti

penggelapan dan penipuan menunjukkan presisi yang lebih rendah dengan presisi 80% dan 83%. Berdasarkan Classification report untuk *batch* 8 dengan *learning rate* $5e-5$ mencapai akurasi 91% dengan *macro average* untuk presisi recall dan f1 score mencapai 91% dan *weighted avg* juga sama-sama 91% .

Berdasarkan tabel 5, hasil evaluasi dengan data yang validasi yang belum pernah dilihat sebelumnya oleh model, menunjukkan bahwa model terbaik dicapai oleh 8 *batch* dengan kombinasi *learning rate* $5e-5$. Model tersebut mencapai akurasi 92,59% dengan F1-Score 92,59%, hal ini menunjukkan bahwa model tersebut mampu mengenali target secara tepat dan konsisten.

Tabel 5. Hasil tes dengan set yang berbeda

<i>Batch</i>	<i>Learning Rate</i>	Akurasi	Recall	Presisi	F1-Score
8	1e-5	87,96%	87,96%	98,34%	92,33%
	2e-5	87,96%	87,96%	92,93%	89,56%
	5e-5	92,59%	92,59%	92,93%	92,59%
16	1e-5	87,04%	87,04%	90,93%	88,49%
	2e-5	90,74%	90,74%	92,64%	91,54%
	5e-5	87,96%	87,96%	92,81%	89,8%



Gambar 3. Confusion matrix model batch 8 learning rate $5e-5$

Confusion matrix pada gambar 3 menunjukkan bahwa model berhasil melakukan klasifikasi sempurna terhadap data validasi terutama pada kasus perjudian dan kejahatan terhadap keamanan negara. Kinerja model untuk klasifikasi kelas penghinaan juga sangat baik, mengingat saat proses *preprocessing* kelas tersebut kekurangan data dibandingkan dengan kelas lainnya. Hasil yang baik juga didapat dari model 16 *batch* dengan kombinasi *learning rate* $2e-5$. Model tersebut mendapat akurasi sebesar 90,74% dengan F1-Score 91,54%. Kedua model tersebut menunjukkan bahwa hasil klasifikasi terdistribusi secara rata sehingga tidak terjadi bias yang signifikan terhadap satu label.

Pembahasan

Hasil penelitian ini menunjukkan model pre trained IndoBERT sebelum dilakukan *fine-tuning* dapat menyelesaikan masalah klasifikasi dengan baik. Menggunakan model *pre-trained*, mendapat akurasi 89,1% dan selama pelatihan model terindikasi terus belajar dengan meningkatnya akurasi. Setelah proses fine-tuning, performa model meningkat terutama dengan kombinasi *batch size* 8 dan *learning rate* $5e-5$.

Kombinasi *batch* 8 dengan *learning rate* $5e-5$ pada tabel 3 menghasilkan f1-score dan recall yang tinggi. Performa tersebut diperoleh berkat penggunaan *batch size* kecil yang memungkinkan model lebih sensitif terhadap detail kecil pada data yang umum ditemukan dalam dokumen hukum. Di sisi lain, *learning rate* yang lebih tinggi dapat mempercepat proses konvergensi karena model IndoBERT telah melalui fase *pre-training* sehingga penyesuaian parameter *learning size* yang lebih besar diperlukan.

Tabel 4 menunjukkan kategori seperti perjudian mendapatkan akurasi dan presisi yang tinggi, karena sering kali mengandung kata kunci yang spesifik seperti “kartu domino”, “nomor togel”, atau “sabung ayam”. Sedangkan kategori penipuan, pemalsuan dan penggelapan memiliki presisi yang lebih rendah dibanding kategori lainnya. Hal ini bisa disebabkan kategori-kategori tersebut memiliki kemiripan konteks dan kesamaan pada barang bukti yang terkandung pada teks amar putusan, sehingga membingungkan model dalam membedakan tiap kategori.

Pengujian menggunakan data uji yang sebelumnya tidak pernah dilihat oleh model pada tabel 5, menunjukkan bahwa kombinasi hyperparameter yang paling optimal terdapat pada model dengan *batch size* 8 dan *learning rate* $5e-5$ dengan akurasi mencapai 92,59%, dengan recall sebesar 92,59%, presisi 92,93% dan f1-score 92,59%. Pencapaian metrik yang tinggi tersebut dipengaruhi oleh penyebutan secara eksplisit jenis kejahatan pada teks amar putusan, seperti “penipuan”, “penggelapan”, atau “penganiayaan” yang dapat langsung dirujuk kepada pasal atau jenis kejahatan yang terdapat pada KUHP. Dengan model ini, IndoBERT dapat memanfaatkan pola bahasa dalam teks amar untuk mengaitkannya pada kategori hukum pidana atau pasal KUHP yang sesuai.

Berdasarkan perbandingan dengan studi sebelumnya oleh Maulana & Aditya (2025), untuk mendeteksi pelanggaran UU ITE menggunakan IndoBERT. Model pada penelitian kami mencapai akurasi yang lebih baik, yaitu 92,59% dengan f1-score 92,59% dibanding dengan model pada penelitian sebelumnya dengan akurasi 90% dengan f1-score 92%. Peningkatan ini menunjukkan bahwa IndoBERT tidak hanya efektif dalam domain UU ITE, namun dapat juga beradaptasi terhadap struktur data dan bahasa hukum pada teks amar putusan pidana KUHP. Model kami juga mengungguli model *baseline* seperti SVM yang hanya mendapat akurasi sebesar 75% hingga 78% (Sudianto et al., 2022).

Secara keseluruhan, hasil ini signifikan karena struktur bahasa dalam dokumen hukum cenderung berbeda dari teks biasa dengan pola kalimat yang formal dan cenderung panjang. Model berhasil mengklasifikasikan teks menunjukkan bahwa IndoBERT memiliki potensi besar sebagai awal dari pengembangan pipeline NLP hukum di Indonesia, baik untuk klasifikasi dokumen, ekstraksi informasi, maupun pencarian hukum berbasis konteks.

SIMPULAN

Penelitian ini menilai seberapa efektif IndoBERT digunakan untuk mengelompokkan teks hukum dalam 8 kategori yaitu perjudian, penggelapan, penganiayaan, penipuan, pembunuhan, pemalsuan, kejahatan terhadap keamanan negara, dan penghinaan. Model terbaik dengan kombinasi 8 *batch* dan $5e-5$ *learning rate* mencapai akurasi sebesar 92,59% dengan presisi 92,93%, *recall* 92,59% dan *f1-score* 92,59% terhadap data uji yang belum pernah dilihat oleh model pada masa *training*. Temuan ini menunjukkan bahwa IndoBERT mampu menangani kompleksitas bahasa hukum dalam konteks KUHP dan berpotensi menjadi fondasi

bagi sistem klasifikasi otomatis dokumen hukum di Indonesia, sekaligus mendukung efisiensi kerja peradilan, meminimalisir bias, dan membuka peluang integrasi ke dalam sistem informasi Mahkamah Agung.

REFERENSI

- Budiman, I., Faisal, M. R., Faridhah, A., Farmadi, A., Mazdadi, M. I., Saragih, H., & Abadi, F. (2024). Classification Performance Comparison of BERT and IndoBERT on Self-Report of COVID-19 Status on Social Media. *Journal of Computer Sciences Institute*, 30(December 2023), 61–67. <http://dx.doi.org/10.35784/jcsi.5564>
- Deng, Y., Zhao, N., & Huang, X. (2023). Early ChatGPT User Portrait through the Lens of Data. *2023 IEEE International Conference on Big Data (BigData)*, 4770–4775. Italy: IEEE. <https://doi.org/10.1109/BigData59044.2023.10386415>
- Denk, T. I., & Peleteiro Ramallo, A. (2020). Contextual {BERT}: Conditioning the Language Model Using a Global State. *Proceedings of the Graph-based Methods for Natural Language Processing (TextGraphs)*, 46–50. Barcelona, Spain: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.textgraphs-1.5>
- Fahrudin, A., Nurhaipah, T., Ikhwan, G., & Sabda, A. (2025). Peran AI dalam Transformasi Komunikasi : Peluang dan Tantangan. *Jurnal Ilmu Komunikasi Andalan*, 8(1), 1–10.
- Furqon, I. N., & Soyusiawaty, D. (2025). The Role of VADER and SentiWordNet Labeling in Naïve Bayes Accuracy for Sentiment Analysis of Rice Price Increases. *Aviation Electronics, Information Technology, Telecommunications, Electricals, and Controls (AVITEC)*, 7(1), 72–85. <https://doi.org/10.28989/avitec.v7i1.2806>
- Garrido-Merchan, E. C., Gozalo-Brizuela, R., & Gonzalez-Carvajal, S. (2023). Comparing BERT Against Traditional Machine Learning Models in Text Classification. *Journal of Computational and Cognitive Engineering*, 2(4), 352–356. <https://doi.org/10.47852/bonviewJCCE3202838>
- Indriani, F., Nugroho, R. A., Faisal, M. R., & Kartini, D. (2024). Comparative Evaluation of IndoBERT, IndoBERTtweet, and mBERT for Multilabel Student Feedback Classification. *Jurnal RESTI*, 8(6), 748–757. <https://doi.org/10.29207/resti.v8i6.6100>
- Juarta, B., & Yulianto. (2023). Indonesian News Classification Using IndoBert. *International Journal of Intelligent Systems and Applications in Engineering*, 11(2), 454–460. <http://dx.doi.org/10.59422/global.v2i02.229>
- Khairani, U., Mutiawani, V., & Ahmadian, H. (2024). Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indoberttweet Untuk Mendeteksi Emosi Pada Komentar Akun Berita Instagram. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(4), 887–894. <https://doi.org/10.25126/jtiik.1148315>
- Khoirunnisaa, N., Nabila Nastiti Kesuma, K., Setiawan, S., & Yunizar Pratama Yusuf, A. (2024). Klasifikasi Teks Ulasan Aplikasi Netflix Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Svm. *SKANIKA: Sistem Komputer Dan Teknik Informatika*, 7(1), 64–73. <https://doi.org/10.36080/skanika.v7i1.3138>
- Koto, F., Rahimi, A., Lau, J., & Baldwin, T. (2020). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. *Proceedings of the 28th International Conference on Computational Linguistics*, 757–770. Barcelona, Spain: International Comitee on Computational Linguistics <https://doi.org/10.18653/v1/2020.coling-main.66>
- Maulana, M. D., & Aditya, C. S. K. (2025). Perbandingan IndoBERT dan Bi-LSTM Dalam Mendeteksi Pelanggaran Undang-Undang ITE. *SINTECH (Science and Information Technology) Journal*, 8(1), 52–59.
- Nabiilah, G. Z., Alam, I. N., Purwanto, E. S., & Hidayat, M. F. (2024). Indonesian multilabel classification using IndoBERT embedding and MBERT classification. *International*

- Journal of Electrical and Computer Engineering*, 14(1), 1071–1078.
<https://doi.org/10.11591/jece.v14i1.pp1071-1078>
- Pakpahan, R. (2021). Analisa Pengaruh Implementasi Artificial Intelligence Dalam Kehidupan Manusia. *Journal of Information System, Informatics and Computing*, 5(2), 506-513.
<https://doi.org/10.52362/jisicom.v5i2.616>.
- Rahma, I. A., & Suadaa, L. H. (2023). Penerapan Text Augmentation untuk Mengatasi Data yang Tidak Seimbang pada Klasifikasi Teks Berbahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(6), 1329-1340.
<https://doi.org/10.25126/jtiik.2023107325>
- Rohsawati, M. (2023). Tajamnya Pedang Jabatan Bisa Memenggal Keadilan. *Seminar Nasional HUBISINTEK*, 449–452.
- Sudianto, S., Sripamuji, A. D., Ramadhanti, I. R., Amalia, R. R., Saputra, J., & Prihatnowo, B. (2022). Penerapan Algoritma Support Vector Machine dan Multi-Layer Perceptron pada Klasifikasi Topik Berita. *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, 11(2), 84-91.
- Wu, T., Wang, Y., & Quach, N. (2025). Advancements in Natural Language Processing : Exploring Transformer-Based Architectures for Text Understanding. *5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA 2025)*. <https://doi.org/10.48550/arXiv.2503.20227>. Xi'an, China: IEEE
- Yasir, A. H., & Gunawan, A. (2024). Mengungkap Dampaknya : Peran Teknologi AI dalam Revolusi Industri 4 . 0 bagi Sumber Daya Manusia Pendahuluan Metode Hasil dan Pembahasan. *GLOBAL Jurnal Lentera BITEP*, 02(02), 48–55.
<https://doi.org/10.59422/global.v2i02.229>
- Yoo, S. (2021). Comparison of Artificial Intelligence and Human Motivation. *Technium Social Sciences Journal*, 25, 345–351. <https://doi.org/10.47577/tssj.v25i1.4736>
- Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, 23, 100224.
<https://doi.org/10.1016/j.jii.2021.100224>