

DAFTAR ISI

LEMBAR PERNYATAAN	i
LEMBAR PERSEMBAHAN	ii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR.....	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xi
DAFTAR ISTILAH	xii
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian	5
1.5 Batasan Masalah.....	6
1.5.1 Ruang Lingkup Masalah	6
1.5.2 Keterbatasan Penelitian.....	7
1.6 Kerangka Berpikir.....	8
1.7 Sistematika Penulisan	9
BAB II TINJAUAN PUSAKA	11
2.1 Studi Terkait	11
2.1.1 The State of the Art (SotA)	11
2.1.2 Research Gap	14
2.2 Natural Language Processing dalam Domain Hukum (Legal NLP).....	15
2.3 Objek Penelitian: Domain Hukum Indonesia	16
2.3.1 Kitab Undang-Undang Hukum Pidana (KUHP Baru)	16
2.3.2 Undang-Undang Informasi dan Transaksi Elektronik (UU ITE).....	16
2.3.3 Undang-Undang Cipta Kerja (UU Omnibus Law)	17
2.3.4 Undang-Undang Pelindungan Data Pribadi (UU PDP)	17

2.4 Karakteristik Linguistik Hukum Indonesia.....	17
2.5 Automatic Text Simplification (ATS)	18
2.5.1 Definisi dan Pendekatan Neural.....	18
2.5.2 Operasi Linguistik Penyederhanaan.....	18
2.6 Arsitektur Transformer dan Model T5	19
2.6.1 Mekanisme Self-Attention	19
2.6.2 IndoT5: Adaptasi Monolingual untuk Bahasa Indonesia	20
2.6.3 Justifikasi Pemilihan Arsitektur IndoT5.....	21
2.7 Kerangka Kerja Pengembangan (CRISP-DM)	23
2.8 Metodologi Evaluasi	25
2.8.1 Metrik Otomatis: <i>SARI</i> dan <i>BLEU</i>	25
2.8.2 Evaluasi Pakar (Expert Judgment).....	27
BAB III METODOLOGI PENELITIAN.....	29
3.1 Metode Penelitian.....	29
3.2 Alur Penelitian.....	31
3.3 Lingkungan Pengembangan	34
3.4 Koleksi dan Pemahaman Data (<i>Data Collection</i>)	35
3.5 Persiapan Data dan Pra-Pemrosesan (<i>Data Preparation</i>).....	35
3.6 Pemodelan dan Pelatihan (<i>Fine-tuning</i>).....	35
3.6.1 Strategi Tokenisasi dan Penanganan <i>Padding</i>	36
3.6.2 Konfigurasi Pelatihan (<i>Hyperparameters</i>).....	36
3.7 Skenario Parameter Inferensi (<i>Decoding Strategy</i>).....	38
3.8 Skenario Evaluasi Sistem.....	39
3.8.1 Evaluasi Kuantitatif Berbasis Metrik	39
3.8.2 Evaluasi Kualitatif (Validasi Pakar)	39
3.9 Implementasi Model dan Antarmuka (Deployment).....	40
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	41
4.1 Implementasi, Prapemrosesan Data, dan <i>Fine-tuning</i> Model IndoT5 (Menjawab Rumusan Masalah 1)	41
4.1.1 Temuan Implementasi	41
4.1.2 Pembahasan Implementasi	51
4.2 Evaluasi Kinerja Model Berdasarkan Metrik Otomatis (Menjawab Rumusan Masalah 2).....	53

4.2.1 Temuan Evaluasi Otomatis.....	53
4.2.2 Pembahasan Evaluasi Otomatis	54
4.3 Validasi Kualitatif Pakar terhadap Hasil Penyederhanaan (Menjawab Rumusan Masalah 3).....	56
4.3.1 Validasi Pakar Hukum.....	57
4.3.2 Validasi Pakar Linguistik	61
4.3.3 Kesimpulan Akhir Evaluasi Pakar dan Batasan Sistem	64
4.4 Iterasi Perbaikan Model: Implementasi Dynamic Threshold.....	65
4.4.1 Temuan Iterasi Perbaikan	65
4.4.2 Pembahasan Iterasi Perbaikan.....	70
BAB V PENUTUP.....	73
5.1 Kesimpulan	73
5.2 Saran.....	74
DAFTAR PUSTAKA	75
DAFTAR LAMPIRAN	82



DAFTAR GAMBAR

Gambar 1 1 Kerangka Berpikir	8
Gambar 2 1 Arsitektur Model T5	21
Gambar 2 2 CRISP-DM	23
Gambar 3 1 Diagram Aktivitas Penelitian	29
Gambar 3 2 Alur Tahapan Penelitian	32
Gambar 4 1 Distribusi Jumlah Kata	43
Gambar 4 2 Perbandingan Sebaran Panjang Teks.....	44
Gambar 4 3 Distribusi Rasio Kompresi	45
Gambar 4 4 Kategorisasi Efektivitas Kompresi.....	46
Gambar 4 5 Korelasi Panjang Teks Asli dan Penyederhanaan.....	47
Gambar 4 6 Grafik Training dan Validation Loss	50
Gambar 4 7 Grafik Training dan Validation Loss (Setelah Perbaikan).....	67



DAFTAR TABEL

Tabel 2 1 The State of the Art	11
Tabel 3 1 Tabel Hyperparameters.....	36
Tabel 4 1 Tabel Statistik Deskriptif Dataset.....	41
Tabel 4 2 Log Penurunan Nilai Training Loss dan Validation Loss	48
Tabel 4 3 Hasil Evaluasi Model Berdasarkan Metrik Otomatis.....	54
Tabel 4 4 Matriks Hasil Penilaian Pakar Hukum	58
Tabel 4 5 Hasil Penilaian Pakar Linguistik.....	61
Tabel 4 6 Log Penurunan Nilai Training Loss dan Validation Loss (Setelah Perbaikan)	66
Tabel 4 7 Perbandingan Luaran Model Sebelum dan Sesudah Perbaikan	68



DAFTAR ISTILAH

1. **Asas Legalitas:** Prinsip dasar dalam hukum pidana yang menyatakan bahwa tidak ada satu perbuatan pun yang dapat dikenai sanksi pidana kecuali atas kekuatan aturan perundang-undangan yang telah ada sebelum perbuatan tersebut dilakukan.
2. **Automatic Text Simplification (ATS):** Proses rekayasa komputasional untuk menyederhanakan teks berstruktur kalimat kompleks menjadi teks yang lebih mudah dipahami tanpa mengubah esensi makna aslinya.
3. **Beam Search:** Algoritma pencarian heuristik dalam pemrosesan bahasa alami yang bertugas merangkai probabilitas urutan kata terbaik saat model sedang memproduksi atau membangkitkan teks.
4. **Bilingual Evaluation Understudy (BLEU):** Metrik perhitungan otomatis yang berfungsi untuk mengevaluasi kualitas generasi teks mesin dengan cara menghitung persentase kecocokan susunan kata secara persis terhadap teks draf referensi yang dibuat oleh manusia.
5. **Data Cleaning:** Tahapan awal dalam prapemrosesan data komputasi yang bertujuan untuk membersihkan dataset dari deretan baris kosong atau anomali teks agar tidak mengganggu proses pembaruan bobot saat pelatihan.
6. **Eksplisitasi:** Teknik operasional penyederhanaan bahasa yang bekerja dengan cara menambahkan frasa penjelas atau memberikan contoh kasus ke dalam teks asli guna memecahkan ambiguitas bagi pembaca dari kalangan awam.
7. **Epoch:** Satu siklus putaran waktu secara penuh di dalam pelatihan model pembelajaran mesin, di mana seluruh kumpulan dataset dipelajari secara tuntas satu kali oleh algoritma.
8. **Exploratory Data Analysis (EDA):** Proses analisis statistik secara visual terhadap dataset untuk memahami karakteristik linguistik dari teks sumber, seperti mengukur sebaran distribusi panjang kalimat dan nilai rasio kompresi.
9. **Fine-tuning:** Proses pelatihan pembaruan bobot pada model kecerdasan buatan bahasa berskala besar agar kemampuan dasar yang dimilikinya dapat beradaptasi untuk mengerjakan tugas pada domain data yang lebih spesifik.

10. **Human-in-the-Loop:** Konsep kerangka perancangan sistem kecerdasan buatan yang mewajibkan adanya intervensi kurasi atau pengawasan dari manusia sebagai penentu keputusan akhir sebelum luaran sistem digunakan secara nyata.
11. **Ignorantia juris non excusat:** Asas fiksi hukum yang menetapkan bahwa ketidaktahuan seseorang terhadap suatu undang-undang tidak dapat dijadikan alasan pemaaf atau pembenar atas pelanggaran yang dilakukannya, karena setiap warga negara dianggap telah mengetahui hukum yang berlaku.
12. **IndoT5:** Sebuah model kecerdasan buatan berbasis arsitektur Transformer yang dilatih dan dikembangkan secara spesifik untuk mengerjakan beragam tugas pemrosesan bahasa alami di dalam tata bahasa Indonesia.
13. **JDIH (Jaringan Dokumentasi dan Informasi Hukum):** Wadah pendayagunaan bersama atas dokumen hukum secara tertib, terpadu, dan berkesinambungan yang berfungsi sebagai sarana pemberian pelayanan informasi hukum kepada publik.
14. **Korpus:** Kumpulan teks dalam jumlah besar yang telah dikurasi dan distrukturkan secara sistematis untuk digunakan sebagai bahan baku pelatihan dalam model pembelajaran mesin.
15. **Loss:** Matriks matematis yang mengukur besaran tingkat kesalahan prediksi komputasi dari sebuah model kecerdasan buatan selama tahapan pelatihan berlangsung.
16. **Natural Language Generation (NLG):** Cabang ilmu dari kecerdasan buatan yang secara khusus berfokus pada perancangan algoritma agar komputer mampu memproduksi atau membangkitkan susunan teks baru yang fasih dibaca oleh manusia.
17. **Natural Language Processing (NLP):** Disiplin ilmu komputer yang berfokus pada interaksi antara perangkat komputasi dan bahasa alami manusia untuk memproses teks dalam jumlah besar secara matematis.
18. **Overfitting:** Kondisi kegagalan sistematis saat model kecerdasan buatan terlalu menghafal detail dari data latih secara kaku, sehingga sistem gagal mengenali atau memproses data baru yang belum pernah dilihat sebelumnya.

19. System output Against References and against the Input sentence (SARI):

Metrik perhitungan otomatis yang secara khusus dirancang untuk mengevaluasi tugas penyederhanaan teks melalui kalkulasi keseimbangan tiga operasi, yakni operasi penambahan kata, penghapusan kata rumit, dan pemertahanan subjek dasar.

20. Threshold: Parameter pengaturan ambang batas yang disuntikkan secara dinamis ke dalam komputasi sistem guna mengendalikan persentase jumlah potongan karakter minimal pada teks luaran.

21. Tokenisasi: Proses pemecahan kalimat utuh menjadi unit karakter atau kosakata yang lebih kecil agar dapat dikonversi menjadi representasi angka matriks untuk diproses oleh jaringan komputasi.

22. Transformer: Arsitektur jaringan saraf tiruan mutakhir yang sangat andal dalam memproses data berurutan seperti teks, yang saat ini menjadi fondasi dasar dari pengembangan berbagai model bahasa skala besar.

23. Yurisprudensi: Serangkaian putusan hakim terdahulu yang telah berkekuatan hukum tetap dan dijadikan sebagai pedoman atau rujukan oleh hakim lain dalam memutus perkara yang memiliki kesamaan konteks.



DAFTAR LAMPIRAN

Lampiran 1 Validasi Pakar 1 Ibu Lena Ishelmiany Ziaharah, S.H., M.H.	82
Lampiran 2 Bapak Dr. Taufiq Alamsyah, S.H., M.H.	85
Lampiran 3 Ibu Dewi Mayaningsih, S.H., M.H., CPM., CPAMB., CLA., CWC.	88
Lampiran 4 Dr. Yeti Heryati, S.Ag., M.Pd.	91

