

Peningkatan Efisiensi Aduan Masyarakat dengan Klasifikasi Instansi Pemerintah dan Named Entity Recognition

Nailfaaz¹, Krisna Bayu Dharma Putra², Rabbani Nur Kumoro³, Yunita Sari⁴

¹Fakultas Matematika dan Ilmu Pengetahuan Alam dan Universitas Gadjah Mada, Daerah Istimewa Yogyakarta, 55281, email: nailfaaz@mail.ugm.ac.id

²Fakultas Matematika dan Ilmu Pengetahuan Alam dan Universitas Gadjah Mada, Daerah Istimewa Yogyakarta, 55281, email: krisnabayu01@mail.ugm.ac.id

³Fakultas Matematika dan Ilmu Pengetahuan Alam dan Universitas Gadjah Mada, Daerah Istimewa Yogyakarta, 55281, email: rabbani.nur.kumoro@mail.ugm.ac.id

⁴Fakultas Matematika dan Ilmu Pengetahuan Alam dan Universitas Gadjah Mada, Daerah Istimewa Yogyakarta, 55281, email: yunita.sari@ugm.ac.id

Corresponding Author: Nailfaaz

INTISARI — Layanan aduan publik sebagai alat pengawasan masyarakat dalam penyelenggaraan pelayanan publik membutuhkan infrastruktur yang memadai. Meskipun pemerintah telah berupaya memfasilitasinya melalui SP4N - LAPOR!, masih terdapat inefisiensi, khususnya dalam pengisian formulir secara manual. Hasil pelaporan masyarakat berupa teks tidak terstruktur sehingga perlu dilakukan ekstraksi entitas yang penting. Oleh karena itu, diusulkan penambahan klasifikasi lembaga serta implementasi *Named Entity Recognition* (NER) sebagai langkah peningkatan efisiensi. Penelitian ini menggunakan dataset aduan dan nama instansi hasil *scraping* dari laman lapor.go.id untuk klasifikasi lembaga, serta dataset pengenalan entitas dengan *tag* entitas berbahasa Indonesia. Dalam klasifikasi lembaga, data akan diklasifikasikan ke tujuh kelas lembaga. Hasil pengujian menunjukkan bahwa model IndoBERTweet memberikan hasil terbaik dengan skor akurasi dan F1 sekitar 0,881 untuk kedua metrik evaluasi. Selanjutnya, *fine-tuning* dilakukan terhadap model IndoBERT Uncased untuk pengembangan model NER yang menghasilkan skor akurasi dan F1 secara berurutan sebesar 0,9494 dan 0,9163. Model IndoBERTweet dan IndoBERT Uncased terbukti efektif dalam mengklasifikasikan lembaga dan mengenali entitas dalam teks aduan sehingga dapat digunakan untuk mendukung peningkatan kualitas pelayanan publik melalui pemrosesan aduan masyarakat yang lebih efisien dan akurat.

KATA KUNCI — Layanan Publik, IndoBERT, IndoBERTweet, *Long Short-term Memory*, *Natural Language Processing*.

I. PENDAHULUAN

A. LATAR BELAKANG

Pengaduan masyarakat merupakan salah satu bentuk penting dari pengawasan terhadap penyelenggaraan pelayanan publik di Indonesia. Sebagai sumbangan pemikiran, saran, gagasan, atau keluhan yang membangun, pengaduan masyarakat memainkan peran kunci dalam menjamin akuntabilitas dan kualitas layanan publik. Untuk memfasilitasi proses ini, pemerintah Indonesia telah menginisiasi Sistem Pengelolaan Pengaduan Pelayanan Publik Nasional (SP4N) melalui Layanan Aspirasi dan Pengaduan Online Rakyat (LAPOR!). LAPOR! telah diamanatkan sebagai platform resmi untuk masyarakat menyampaikan aspirasi dan pengaduan mereka, sesuai dengan Peraturan Presiden Nomor 76 Tahun 2013 dan Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 3 Tahun 2015 [1].

Meskipun telah memiliki infrastruktur seperti SP4N – LAPOR!, masih ada kendala yang perlu diatasi. Salah satunya adalah proses yang masih manual dan tidak optimal. Saat ini, proses pengaduan di LAPOR! membutuhkan pengguna untuk mengisi banyak kolom formulir secara manual, yang tidak hanya memakan waktu tetapi juga meningkatkan risiko kesalahan. Dengan meminimalkan jumlah kolom yang harus diisi, pengguna dapat melaporkan masalah mereka dengan lebih cepat dan tanpa kesulitan.

Dengan penerapan teknologi *Natural Language Processing* (NLP) seperti *Named Entity Recognition* (NER) dan klasifikasi, pesan aduan yang dikirim melalui website ataupun media sosial dengan format pesan bebas atau tidak terstruktur dapat langsung diekstraksi, terklasifikasi, dan dikirim ke instansi yang tepat. Informasi yang diidentifikasi, seperti lokasi, tanggal, dan jenis masalah, dapat diolah secara otomatis, memastikan bahwa setiap pengaduan dikelola dengan efisien. Proses ini juga meminimalisasi kebutuhan pengisian formulir secara manual, sehingga masyarakat dapat

melaporkan masalah mereka dengan lebih cepat dan mudah. Selain itu, data *tag* yang dihasilkan dari proses ini dapat digunakan untuk analisis lebih lanjut, memberikan wawasan berharga untuk peningkatan kualitas layanan publik.

B. TUJUAN

Penelitian ini bertujuan untuk membangun model *deep learning* yang secara otomatis mengklasifikasikan instansi pemerintah dan mengekstraksi entitas dari pengaduan masyarakat. Dengan model ini, pengaduan dapat langsung dikategorikan dan disalurkan ke instansi yang relevan. Selain itu, penelitian ini juga bertujuan mempercepat dan mempermudah proses pengaduan dengan memungkinkan masyarakat melapor dalam format teks bebas, yang kemudian akan dilakukan ekstraksi informasi penting menggunakan metode *Named Entity Recognition* (NER), sehingga laporan akan lebih mudah diproses dan ditindaklanjuti.

C. MANFAAT

Penelitian ini diharapkan memberikan beberapa manfaat yang signifikan. Manfaat pertama adalah kemudahan dalam proses pelaporan, di mana pengguna tidak perlu lagi mencari secara manual instansi dan geolokasi sehingga menghemat waktu dan usaha. Ekstraksi entitas bermanfaat untuk mendapatkan informasi penting dari aduan seperti lokasi atau jenis kejadian sehingga memudahkan proses tindak lanjut. Manfaat lainnya adalah peningkatan citra dan kredibilitas pemerintah. Dengan menjadi lebih responsif dan proaktif dalam menangani masalah masyarakat, pemerintah diharapkan dapat memperkuat citra positif dan kredibilitasnya di mata publik.

D. BATASAN PENELITIAN

Berikut merupakan beberapa poin yang menjadi batasan dalam penelitian ini antara lain:

- Data aduan masyarakat yang digunakan berasal dari laman **lapor.go.id** dengan rentang waktu aduan mulai dari Mei 2020 hingga sekitar Maret 2024.
- Dataset NER diambil dari dataset publik **id_nergrit_corpus** yang diunggah oleh GRIT Indonesia pada tahun 2022 dan di-update tahun 2024 di situs **huggingface.co**.
- Instansi yang menjadi kelas dalam tahap klasifikasi lembaga berjumlah tujuh, yaitu Kepolisian Negara Republik Indonesia, Kementerian Pendidikan Kebudayaan Riset dan Teknologi, Kementerian Sosial, Divisi Pemasaran dan Pelayanan Pelanggan (PLN), Direktorat Jenderal Perhubungan Darat, Kementerian Komunikasi dan Informatika, dan Direktorat Jenderal Tenaga Kesehatan.

II. KAJIAN PUSTAKA

Penelitian sebelumnya telah menginvestigasi penggunaan metode Support Vector Machine (SVM) untuk mengklasifikasi keluhan yang ditujukan kepada lembaga atau instansi pemerintah. Metode SVM digunakan untuk mengidentifikasi nama lembaga atau instansi yang relevan dengan keluhan yang disampaikan. Evaluasi dilakukan dengan berbagai rasio pembagian data latih-uji, dan hasil menunjukkan akurasi tertinggi mencapai 85% pada rasio 80:20 dan terendah 68,8% pada rasio 50:50 [2].

Penelitian lainnya adalah pertimbangan penggunaan kombinasi metode Latent Dirichlet Allocation (LDA) dan SVM untuk mengklasifikasi laporan keluhan terhadap layanan publik berdasarkan instansi yang bersangkutan. LDA digunakan untuk mengekstrak topik-topik dari laporan keluhan, sedangkan SVM digunakan untuk mengklasifikasikan laporan tersebut berdasarkan topik yang diekstraksi oleh LDA. Hasil evaluasi model menunjukkan kinerja yang baik dengan akurasi mencapai 79,85%, presisi 79,98%, *recall* 72,37%, dan skor F1 74,67% pada teknik *split train-test* dengan rasio 70:30 [3].

Terdapat juga penelitian yang menunjukkan bahwa penggunaan BERT dalam ekstraksi alamat menggunakan NER, secara spesifik standarisasi alamat, yang menghasilkan akurasi hingga 94%. Hal ini menunjukkan potensi model berbasis BERT dalam menangani permasalahan ekstraksi representasi kontekstual kata-kata individual dalam sebuah teks ataupun kalimat [4].

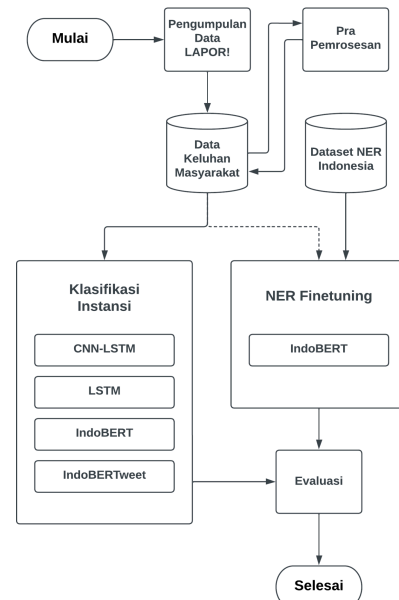
Penelitian serupa lainnya juga pernah dilakukan. Penelitian ini menggunakan teknik *self-supervision* dan *preprocessing* label diterapkan untuk meningkatkan kinerja model NER. Misalnya, RoBERTa yang dilakukan *fine-tuning* pada *dataset* CoNLL 2003 dengan hanya 10% dari total entitas berlabel mencapai kinerja setara dengan *baseline* yang dilatih dengan 50% entitas berlabel. Penelitian ini menunjukkan bahwa metode ini efektif untuk mengoptimalkan performa model NER pada *dataset* dengan pelabelan terbatas [5].

Dalam penelitian yang kami usulkan, kami menggunakan model *deep learning* untuk mengklasifikasi dan mengekstraksi informasi dari laporan keluhan di *website* **lapor.go.id**. Kami juga melakukan *fine-tuning* pada model IndoBERTweet untuk meningkatkan ekstraksi informasi penting, berbeda dengan penelitian serupa yang menggunakan BERT atau RoBERTa. Pendekatan ini diharapkan dapat meningkatkan akurasi dan efisiensi dibandingkan metode sebelumnya.

III. METODE EKSPERIMEN

A. ALUR KERJA

Seperti yang terlihat pada **Gambar 1**, penelitian ini melibatkan beberapa tahapan, termasuk pengumpulan data, pemrosesan data mentah, pemodelan klasifikasi instansi dan NER, serta evaluasi terhadap model yang diusulkan.



Gambar 1. Diagram Alur Metodologi

B. PENGUMPULAN DATA

1) Akuisisi Data Klasifikasi Instansi

Akuisisi data dilakukan dengan melakukan *scraping* menggunakan *library* Selenium. Data yang diambil adalah hasil dari aduan masyarakat yang diambil dari beberapa instansi yang tersedia di *web* **lapor.go.id**. Instansi yang kami lakukan *scraping* adalah (a) Kepolisian Negara Republik Indonesia, (b) Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, (c) Kementerian Sosial, (d) Divisi Pemasaran dan Pelayanan Pelanggan (PLN), (e) Direktorat Jenderal Perhubungan Darat, (f) Kementerian Komunikasi dan Informatika, dan (g) Direktorat Jenderal Tenaga Kesehatan.

Selanjutnya, data yang tersedia di setiap *web* instansi akan dilakukan *scraping*. *Scraping* dilakukan dengan mengambil aduan masyarakat terbaru hingga terlama. Data yang di *scraping* berupa nama pelapor, tanggal laporan, judul laporan, detail laporan, dan instansi tujuan laporan. Hasil *scraping* adalah data berformat *csv* yang berisi tujuh jenis instansi pemerintah dengan sampel *dataset* sesuai dengan **Tabel I**.

Didapatkan jumlah aduan terbanyak berasal dari Kepolisian Negara Republik Indonesia dengan 632 aduan, diikuti oleh Direktorat Jenderal Perhubungan Darat dengan 609 aduan, Divisi Pemasaran dan Pelayanan Pelanggan (PLN) dengan 607 aduan, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi dengan 606 aduan, serta Kementerian Sosial dengan 606 aduan. Sementara itu, Kementerian Komunikasi dan Informatika serta Direktorat Jenderal Tenaga Kesehatan masing-masing menerima 320 dan 310 aduan yang disebabkan oleh terbatasnya ketersediaan aduan.

TABEL I

SAMPEL HASIL SCRAPING KELUHAN MASYARAKAT WEBSITE SP4N LAPOR!

Pelapor	Tanggal Masuk	Judul Laporan	Detail Keluhan	Instansi
---------	---------------	---------------	----------------	----------

Laporan				
Yogi Wibowo	Kamis, 20:29	Skp Tidak Keluar	Skp tidak keluar ketika mengikuti	Direktorat Jenderal Tenaga Kesehatan
Gilang Hafizh Hilmy	6 Mar, 08:57	Pju mati berbulan-bulan	Pju hanya menyala 1 sisanya mati total.	Direktorat Jenderal Perhubungan Darat

2) Akuisisi Data Named Entity Recognition

Dataset Pengenalan Entitas (NER) didapat dari laman **huggingface.co**. Dataset ini terdiri dari total 17452 data berbahasa Indonesia yang terbagi menjadi data latih, data validasi, dan data uji. Data latih terbagi menjadi 12532 data, data validasi sebanyak 2399, dan data uji sebanyak 2521. Struktur dataset ini terbagi menjadi tiga buah jenis, yaitu id yang menandakan index penanda data, tokens yang merepresentasikan potongan kata dari kalimat, dan ner_tags yang merepresentasikan jenis entitas seperti pada **Tabel II**.

TABEL II
STRUKTUR DATASET NER

Struktur Data	Sampel Data
id	0
tokens	['Gubernur', 'Bank', 'Indonesia', 'menggelar', 'konferensi', 'pers']
ner_tags	[9, 28, 28, 38, 38, 38]

Pelabelan entitas pada data menggunakan awalan **B** dan **I** yang mana **B** menandakan *beginning* atau permulaan suatu entitas dan **I** yang berarti *inside* atau entitas tersebut termasuk bagian dari suatu entitas lain. Pelabelan token juga dibagi menjadi 19 entitas dan satu entitas 'O' atau *others* seperti pada **Tabel III**.

TABEL III
JENIS ENTITAS DATASET NER

'CRD' = Cardinal	'ORD' = Ordinal
'DAT' = Date	'ORG' = Organisasi
'EVT' = Event	'PER' = Orang / Tokoh
'FAC' = Fasilitas	'PRC' = Persentase
'GPE' = Entitas Geopolitik	'PRD' = Produk
'LAW' = Entitas Hukum	'QTY' = Kuantitas
'LOC' = Lokasi	'REG' = Agama
'MON' = Uang	'TIM' = Waktu
'NOR' = Organisasi Politik	'WOA' = Karya Seni
'O' = Semua entitas yang tidak termasuk 19 entitas inti	

Selanjutnya, kami juga memodifikasi dataset untuk menambahkan satu buah entitas lain bernama 'LAP' yang menandakan entitas berhubungan dengan permasalahan masyarakat, seperti perkelahian, kecelakaan, dan kekerasan. Total entitas akhir menjadi 20 entitas dan 1 buah entitas *others*. Jika digabungkan dengan awalan **B** atau **I** dalam penanda entitas, tag akhir dataset mencapai 41 tag, terdiri dari 40 tag inti dan 1 buah tag *others*[6].

C. PRA PEMROSESAN

Data keluhan yang telah diambil dibersihkan melalui serangkaian tahapan. Tahap awal adalah *case folding*, di mana semua teks diubah menjadi kata kata berhuruf kecil. Selanjutnya, semua link yang ada dalam teks dihapus, begitu juga dengan *username* dan *emoji* yang tidak relevan. Selain itu, kata-kata gaul dan kata-kata singkatan diperbaiki agar konsisten. Tagar dan spasi berlebih juga dihilangkan untuk membersihkan teks lebih lanjut. Kemudian, tanda baca dihapuskan, dan teks dibagi menjadi unit-unit yang lebih kecil melalui tokenisasi, seperti kata atau frasa. Kata-kata yang tidak memiliki makna penting, atau *stopwords*, juga dihapus, diikuti dengan *stemming* untuk mengubah token-token menjadi kata dasarnya. Hasil dari pemrosesan data pada kolom judul keluhan dapat dilihat pada **Tabel IV**.

TABEL IV
SAMPEL DATA ASLI DAN DATA HASIL PRA PEMROSESAN

Judul Keluhan Sebelum Pemrosesan	Judul Keluhan Setelah Pemrosesan
Sertifikat Vaksin Ke 3 Belum Muncul Di Satusehat Padahal Sudah Vaksin Ke 3	sertifikat vaksin belum muncul satusehat padahal vaksin
Scam Penipuan Berkedok Like Dan Subscribe	scam tipu kedok like subscribe

Pada *dataset* NER, tidak dilakukan banyak proses prapemrosesan. Satu-satunya langkah yang diambil adalah mengonversi semua huruf menjadi huruf kecil (*lowercasing*) untuk memastikan konsistensi dalam representasi kata.

D. PEMODELAN

1) Pemodelan Klasifikasi Instansi

Dataset keluhan dibagi menjadi data latih (80%) dan data uji (20%) dengan distribusi seperti pada **Tabel V**.

TABEL V
DISTRIBUSI DATASET TIAP INSTANSI PADA DATA LATIH DAN DATA UJI

Instansi	Data Latih	Data Uji
Kepolisian Negara Republik Indonesia	501	131
Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi	494	112
Kementerian Sosial	494	112
Divisi Pemasaran dan Pelayanan Pelanggan (PLN)	486	121
Direktorat Jenderal Perhubungan Darat	478	131

Kementerian Komunikasi dan Informatika	240	80
Direktorat Jenderal Tenaga Kesehatan	235	75

Beberapa model deep learning digunakan untuk melatih data, yaitu IndoBERT[7], IndoBERTweet[8], LSTM[9], dan CNN-LSTM[10]. Setiap model dilatih menggunakan 15 epochs dengan batch size sebesar 16. Dikalkulasikan juga nilai akurasi dan F1 sebagai skor untuk evaluasi model.

2) Pemodelan Named Entity Recognition

Untuk melakukan fine-tuning pada model NER, kami memilih menggunakan IndoBERT sebagai model dasar serta dataset[6] dengan proporsi 12,532 baris data untuk data training dan 2,399 baris data untuk data validasi. Dalam proses ini, kami mengatur batch size sebesar 16 dan melakukan pelatihan selama 5 epoch.

IV. HASIL DAN ANALISIS

A. HASIL EKSPERIMEN

Evaluasi dilakukan dengan membandingkan skor akurasi dan F1 pada data Uji. Selanjutnya akan dibandingkan juga total waktu latih keempat model pada kasus klasifikasi instansi untuk memahami waktu komputasi yang diperlukan setiap model dalam melakukan prediksi.

1) Hasil Evaluasi Klasifikasi Instansi

Hasil evaluasi menunjukkan bahwa keempat model memiliki skor akurasi dan F1 yang cukup tinggi. Analisis lebih dalam juga membuktikan bahwa model yang memiliki akurasi dan F1 tertinggi juga akan memiliki waktu latih yang paling lama seperti yang ditunjukkan oleh Tabel VI.

Model LSTM mencapai skor F1 sebesar 0,803 dengan waktu pelatihan 1684 detik atau 97 detik per *epoch*, menunjukkan efisiensi meskipun performanya di bawah IndoBERTweet. Model CNN-LSTM, dengan skor F1 0,791, memiliki waktu pelatihan 50% lebih cepat dari LSTM, yaitu 724 detik atau 48 detik per *epoch*, menunjukkan kombinasi yang efektif dalam mengurangi waktu pelatihan.

TABEL VI

HASIL PEMODELAN PADA DATA KELUHAN MASYARAKAT

Model	Data Latih		Data Uji		Waktu latih total(s)
	Akurasi	F1	Akurasi	F1	
Indo BERT	0,891	0,891	0,752	0,754	2415
Indo BERTweet	0,961	0,962	0,881	0,881	2450
CNN -LSTM	0,776	0,776	0,791	0,791	724
LSTM	0,815	0,815	0,803	0,803	1684

Model IndoBERT memiliki performa baik pada data latih dengan skor F1 0,891, namun *overfit* pada data uji dengan skor F1 0,754. *Overfit* pada model IndoBERT kemungkinan disebabkan oleh model *pre-trained* IndoBERT yang dilatih

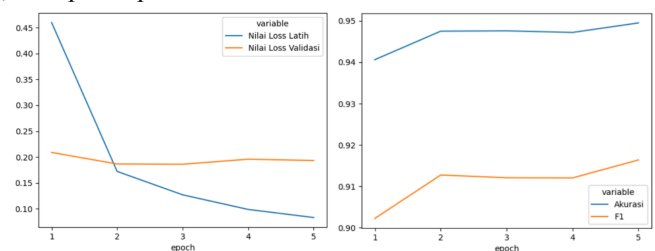
pada dataset universal, seperti wikipedia dan laman berita, biasanya menggunakan struktur kalimat yang panjang dan baku. Struktur kalimat ini kemungkinan berbeda dengan struktur kalimat yang biasanya digunakan pada pesan aduan masyarakat yang biasanya berupa pesan singkat dan kadang menggunakan bahasa yang tidak terlalu baku.

Dilain sisi, *overfit* pada model BERT tidak ditunjukkan oleh model IndoBERTweet yang mencapai skor akurasi dan F1 sebesar 0,881 untuk kedua metrik evaluasi. Skor ini merupakan skor tertinggi untuk evaluasi akurasi dan F1 jika dibandingkan dengan ketiga model lainnya. Skor yang tinggi ini kemungkinan disebabkan oleh data latih *pre-trained* IndoBERTweet yang menggunakan dataset media sosial X sehingga memiliki struktur kalimat yang cukup mirip dengan pesan aduan masyarakat. Hasil ini membuktikan bahwa dataset dari media sosial X lebih cocok digunakan sebagai dataset latih *pre-trained* model BERT untuk kasus klasifikasi instansi menggunakan pesan aduan masyarakat dibandingkan dengan dataset yang diambil dari wikipedia atau situs berita.

Analisis lebih lanjut dilakukan dengan membandingkan waktu latih untuk keempat model. CNN-LSTM memiliki waktu latih paling cepat sekitar 724 detik, namun memiliki skor akurasi dan F1 terendah dibandingkan dengan model lainnya. Posisi tercepat kedua diraih oleh LSTM dengan waktu latih 1684 detik untuk 15 *epochs*. Selanjutnya, kedua model BERT memiliki waktu latih yang cukup mirip yaitu sebesar 2415 untuk IndoBERT dan 2450 untuk IndoBERTweet. Waktu latih identik ini disebabkan karena pada dasarnya kedua model BERT memiliki arsitektur yang serupa, namun memiliki dataset latih yang berbeda. Model BERT memiliki waktu latih terlama yang menunjukkan kompleksitas model yang lebih tinggi dibandingkan model LSTM maupun CNN-LSTM..

2) Hasil Evaluasi Fine-tuning Model NER

Hasil Pelatihan menunjukkan bahwa nilai *loss* data latih terus mengalami penurunan, sementara nilai *loss* data validasi mengalami penurunan dari *epoch* 1 ke *epoch* 2, namun mengalami kenaikan pada *epoch* 3 dan 4 serta stagnan pada *epoch* 5 seperti pada Gambar 2 sebelah kiri. Val_loss berhasil diturunkan dari 0,2088 pada epoch pertama menjadi 0,1860 pada epoch ketiga, namun mengalami peningkatan menjadi 0,1932 pada epoch kelima.



Gambar 2. Nilai Loss Pelatihan NER pada Data Uji dan Validasi

Evaluasi lain dilakukan pada dataset NER dengan membandingkan hasil akurasi dan F1 pada data validasi. Hasil pelatihan menunjukkan bahwa akurasi dan F1 terus mengalami peningkatan seiring berjalannya *epoch* seperti yang terlihat pada Gambar 2 sebelah kanan. Akurasi dan nilai F1 mencapai nilai 0,9475 dan 0,9120 pada epoch ketiga dan mencapai nilai tertinggi sebesar 0,9494 dan 0,9163 pada *epoch* kelima. Hasil ini menunjukkan bahwa model sudah mencapai tingkat konvergen sehingga perubahan dari nilai akurasi dan F1 tidak terlalu signifikan.

B. ANALISIS LANJUTAN

1) Analisis Lanjutan Klasifikasi Instansi

Pada **Gambar 3** banyak kategori pada data uji berhasil mencapai akurasi lebih dari 90%. Namun, Divisi Pemasaran dan Pelayanan Pelanggan (PLN) dan Kepolisian Negara Republik Indonesia hanya mencapai sekitar 76,32% dan 77,46%. Kategori PLN sering disalahkategorikan sebagai Direktorat Jenderal Perhubungan Darat sebesar 15,79%, sedangkan Kepolisian sering disalahkategorikan sebagai Direktorat Jenderal Tenaga Kesehatan sebesar 14,08%.

Kepolisian Negara Republik Indonesia (a)	89.47	0.00	3.01	0.00	3.01	0.00	4.51
Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (b)	2.63	76.32	1.32	0.00	15.79	2.63	1.32
Kementerian Sosial (c)	4.42	0.00	92.04	0.00	2.65	0.88	0.00
Divisi Pemasaran dan Pelayanan Pelanggan (PLN) (d)	4.23	0.00	1.41	77.46	1.41	1.41	14.08
Direktorat Jenderal Perhubungan Darat (e)	1.74	0.00	0.00	0.00	97.39	0.00	0.87
Kementerian Komunikasi dan Informatika (f)	1.89	0.00	0.00	0.00	2.83	94.34	0.94
Direktorat Jenderal Tenaga Kesehatan (g)	0.85	0.00	0.00	1.69	0.85	0.00	96.61
	a	b	c	d	e	f	g

True Label

Predicted Label

Gambar 3. Confusion Matrix Hasil Prediksi Data Uji Model.

Direktorat Jenderal Perhubungan Darat juga banyak dimisklasifikasikan sebagai Kepolisian sebesar 4,51%, dan Tenaga Kesehatan sering disalahkategorikan sebagai Kementerian Pendidikan sebesar 15,79%.

Analisis *word cloud* untuk Direktorat Jenderal Perhubungan Darat dan Direktorat Jenderal Tenaga Kesehatan dapat dilihat pada **Gambar 4**. Gambar ini menunjukkan fokus keluhan yang berbeda. Namun, ada kemungkinan kesamaan tren pada beberapa kata yang bisa berkaitan dengan isu umum seperti "pelayanan" atau "tarif". Kata-kata ini dapat muncul dalam konteks yang berbeda tetapi tetap relevan dalam kedua kategori, yang kemungkinan menyebabkan misklasifikasi antara kedua kelas instansi tersebut.



Gambar 4. Word cloud (kiri) Direktorat Jenderal Perhubungan Darat dan (kanan) Direktorat Jenderal Tenaga Kesehatan

2) Analisis Lanjutan Named Entity Recognition

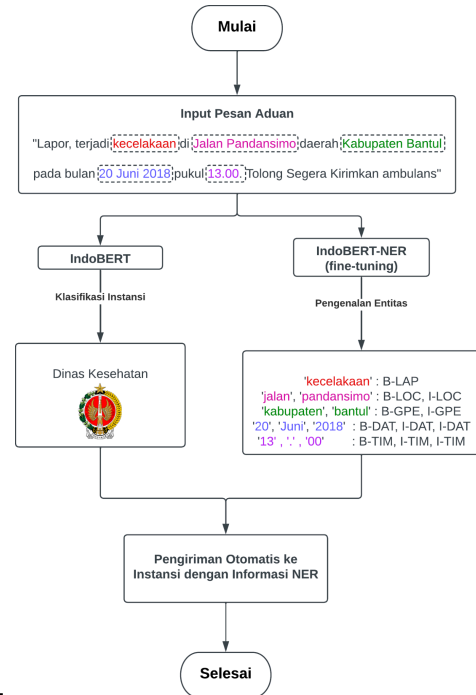
Hasil analisis pengenalan entitas menunjukkan bahwa model sudah cukup baik dalam melakukan pengenalan entitas. Pengenalan entitas berhasil mengenali entitas-entitas penting seperti jenis masalah pada laporan, waktu, tanggal, dan lokasi kejadian.

Analisis lanjutan menunjukkan kelemahan model pengenalan entitas adalah sulitnya model dalam mengenali entitas baru hasil modifikasi, yaitu LAP terutama untuk kata-kata yang berada di tengah atau akhir kalimat. Hal ini kemungkinan disebabkan oleh pelabelan tag LAP (jenis masalah) yang hanya berdasarkan suatu kata berdasarkan kriteria tertentu.

VI. USULAN IMPLEMENTASI

Implementasi dilakukan dengan input pesan laporan yang diproses menggunakan model klasifikasi dan pengenalan entitas seperti pada Gambar 5. Model klasifikasi

mengklasifikasikan isi pesan menjadi 7 instansi terlatih. Selanjutnya, isi pesan dikenali entitasnya untuk mendapatkan informasi penting seperti jenis kejadian (B-LAP), lokasi, tanggal, dan waktu kejadian. Hasil pengenalan entitas dikirim otomatis ke instansi terkait, meningkatkan efisiensi pelaporan dengan mengubah data teks tidak terstruktur menjadi terklasifikasi dan mengekstraksi informasi relevan. Sistem ini memungkinkan pengiriman laporan lebih cepat dan akurat, memastikan respon yang lebih tepat waktu dan sesuai dengan kebutuhan masing-masing instansi.



Gambar 5. Diagram Alur Implementasi Model

VI. KESIMPULAN DAN SARAN

Dari eksperimen yang telah dilakukan, dapat disimpulkan bahwa :

- Model IndoBERTtweet memberikan performa terbaik dengan skor F1 tertinggi sebesar 0,881 pada data uji. Ini menunjukkan kemampuan IndoBERTtweet dalam mengklasifikasi dan menangkap fitur-fitur relevan pada dataset keluhan berbahasa Indonesia dengan sangat baik, meskipun membutuhkan waktu pelatihan terlama.
- Pelatihan model NER dengan fine-tuning model IndoBERT berhasil mendapatkan skor akurasi dan F1 yang tinggi. Model IndoBERT mencapai tingkat konvergen, sehingga perubahan nilai akurasi cenderung kecil dan stagnan.

Evaluasi menunjukkan ada dua instansi dengan akurasi di bawah 80%, disebabkan oleh kemiripan kata-kata dan kalimat pada pelaporan. Disarankan untuk melakukan pelatihan pada *embedding* yang digunakan guna mendapatkan konteks yang lebih dalam terkait laporan sehingga perbedaan antara laporan satu dengan lainnya dapat lebih jelas.

Tag LAP pada dataset NER yang digunakan masih berupa tag yang ditambahkan oleh penulis berdasarkan suatu kriteria tertentu. Penulis menyarankan untuk membuat dataset tag LAP yang dilabeli langsung oleh seorang profesional sehingga hasil pelabelan tag kejadian akan lebih berkualitas dan konsisten.

REFERENSI

- [1] A. Aptika, "Penggunaan Sp4n-LAPOR! Tingkatkan Pelayanan Publik Berkualitas," Ditjen Aptika, <https://aptika.kominfo.go.id/2021/11/penggunaan-sp4n-lapor-tingkatkan-pelayanan-publik-berkualitas>. tanggal akses: 2-Mei-2024.
- [2] D. P. Kartini, (2022). Klasifikasi Target Keluhan Untuk Lembaga Terkait Pada Instansi Pemerintah Menggunakan Support Vector Machine. Repository UPM.
- [3] M. Alkaff, A. R. Baskara, and I. Maulani, "Klasifikasi Laporan Keluhan Pelayanan Publik Berdasarkan instansi menggunakan metode Lda-SVM," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 6, pp. 1265–1276, Nov. 2021. doi:10.25126/jtiik.2021863768.
- [4] A. Sharma, "Advancing Address Standardization through NER with bert," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Jul. 2023. doi:10.1109/icccnt56998.2023.10306547.
- [5] V. Scherbakov and V. Mayorov, "Finetuning Bert on partially annotated NER corpora," 2022 Ivannikov Ispras Open Conference (ISPRAS), Dec. 2022. doi:10.1109/ispras57371.2022.10076869.
- [6] "Grit-id/id_nergrit_corpus · datasets at hugging face," [grit-id/id_nergrit_corpus · Datasets at Hugging Face](https://huggingface.co/datasets/grit-id/id_nergrit_corpus), https://huggingface.co/datasets/grit-id/id_nergrit_corpus (accessed Jun. 26, 2024).
- [7] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "Indolem and Indobert: A benchmark dataset and pre-trained language model for Indonesian NLP," Proceedings of the 28th International Conference on Computational Linguistics, 2020. doi:10.18653/v1/2020.coling-main.66.
- [8] F. Koto, J. H. Lau, and T. Baldwin, "Indobertweet: A pretrained language model for Indonesian twitter with effective domain-specific vocabulary initialization," Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021. doi:10.18653/v1/2021.emnlp-main.833.
- [9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. doi:10.1162/neco.1997.9.8.1735.
- [10] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Apr. 2015. doi:10.1109/icassp.2015.7178838.