Journal : Jurnal Rekayasa Sistem Informasi dan Teknologi (JRSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : e-ISSN 3025-888X
Volume / Issue / Year/ Page : Vol. 4 No. 1 (2026): Agustus/ Hal108-123

NAMED ENTITY RECOGNITION DAN SEQUENCE PAIR CLASSIFICATION BERBASIS INDOBERT UNTUK EKSTRAKSI DAN REKONSILIASI DATA LOGISTIK

Alif Haikal[1*], Jafar Shadiq[2],
[1,2]Universitas Bina Insani, Indonesia
haikalsialip999@gmail.com[1*], jafarshadiq@binainsani.ac.id[2]
*coresponding author

Information Artikel

Pengajuan	:	28-07-2026
Peninjauan	:	11-08-2026
Revisi	:	21-08-2026
Diterima	:	25-08-2026
Diterbitkan	:	30-08-2026

Kata Kunci
Named Entity Recognition, Sequence Pair Classification, IndoBERT, Data Augmentation, Logistics Automation.

Abstrak

PT. Rafay Anugrah Logistik mengandalkan WhatsApp untuk koordinasi pesanan armada harian, namun ekstraksi instruksi pesanan multi-unit ke spreadsheet serta pencocokan pesan susulan dengan pesanan induknya masih dilakukan manual, berisiko menimbulkan penumpukan pekerjaan seiring meningkatnya volume pesanan. Penelitian ini mengusulkan pipeline deep learning hibrida menggunakan dua model fine-tuning IndoBERT. Model pertama, berbasis IndoBERT-Base-Uncased, menerapkan Named Entity Recognition dengan skema BIO tagging untuk mengekstraksi sembilan atribut inti pesanan dari teks mentah. Model kedua, berbasis IndoBERT-Base-p2, menerapkan Sequence Pair Classification untuk menentukan keterkaitan pesan susulan, didukung mekanisme rule-based post-processing untuk normalisasi keluaran. Pendekatan fine-tuning penuh diterapkan untuk memproses pola bahasa informal pada pesanan logistik. Uji banding pada tahap validasi yang menerapkan skema isolasi data (split-before-augmentation) untuk meminimalisasi overfitting, menunjukkan perubahan F1-Score makro model Named Entity Recognition dari 83,3% menjadi 100,0%, dan model Sequence Pair Classification dari 68,5% menjadi 100,0%. Pencapaian metrik tersebut berkaitan erat dengan kapasitas arsitektur dalam mengenali batasan variasi pada dataset validasi sintetis. Oleh karena itu, hasil penelitian ini diposisikan sebagai bukti kelayakan awal (proof of concept). Sebagai tahapan selanjutnya, evaluasi menggunakan data operasional murni berskala besar diperlukan untuk memvalidasi ketahanan model pada kondisi nyata

How to Cite : Haikal, A., & Shadiq, J. (2026). Named entity recognition dan sequence pair classification berbasis IndoBERT untuk ekstraksi dan rekonsiliasi data logistik. *Jurnal Rekayasa Sistem Informasi dan Teknologi (JRSIT)*, 4(1), 108–123. https://doi.org/10.70248/jrsit.v4i1.4427

DOI : https://doi.org/10.70248/jrsit.v4i1.4427
This is an open access article under https://creativecommons.org/licenses/by/4.0/
Copyright© 2026 by Author. Published by Yayasan Nuraini Ibrahim Mandiri

PENDAHULUAN

Dengan pertumbuhan sektor transportasi dan logistik di Indonesia, kebutuhan akan pengelolaan informasi operasional yang lebih cepat dan terorganisir meningkat. Penggunaan aplikasi pesan instan khususnya WhatsApp sebagai media untuk mengkomunikasikan pesanan antara pengirim, operator, dan vendor armada adalah pola yang umum di perusahaan logistik skala menengah [1], di mana itu telah menjadi alat komunikasi bisnis-ke-bisnis *(B2B)* yang paling populer karena kemudahan akses dan tanggapan langsungnya [2]. Namun, penggunaan pesan teks bebas sebagai alat instruksi operasional menyebabkan perbedaan antara kecepatan komunikasi dan kesiapan data untuk diproses secara administratif. Informasi yang datang dalam format semi-terstruktur harus dialihkan secara manual ke dalam sistem pencatatan

terstruktur, seperti spreadsheet.

PT. Rafay Anugrah Logistik, perusahaan penyedia jasa transportasi yang menjadi subjek penelitian ini, secara nyata menemukan masalah tersebut. Instruksi pesanan armada dikirim melalui WhatsApp dan mencakup sembilan atribut: Tanggal Permintaan Pesanan *(RO)*, Tanggal Muat, Pickup, Tujuan, Nomor Plat, Tipe Truk, Jumlah Unit, *Driver*, dan Kontak *Driver*. Atribut jumlah unit menentukan jumlah baris entri yang harus dibuat, sementara delapan atribut lainnya dipetakan pada setiap baris. Proses menjadi lebih kompleks karena pesan tidak pernah diterima lengkap dalam satu pengiriman. Vendor memberikan empat fitur identitas armada baru melalui pesan susulan. Staf administrasi harus secara manual melakukan pencarian dan pencocokan visual antara pesan susulan dan entri induk yang telah tersimpan sebelumnya. Observasi lapangan yang dilakukan dari Januari hingga Maret 2026 menunjukkan bahwa pada hari dengan intensitas tertinggi, 26 Maret 2026, empat belas pesan menghasilkan 200 titik data, yang harus dipetakan secara manual oleh dua administrator. Selain itu, *volume* bulanan meningkat dari 41 baris pada Januari 2026 menjadi 169 baris pada Maret 2026, atau 1.352 titik data, dengan tren yang mungkin terus meningkat seiring dengan ekspansi bisnis.

PT. Rafay Anugrah Logistik menghadapi dua proses komputasi yang saling bergantung, yang pertama adalah mengekstrak fitur pesanan dari teks yang semi-terstruktur dan yang kedua adalah rekonsiliasi pesan susulan dengan pesanan induk yang telah tersimpan. Secara konseptual, kedua proses ini termasuk dalam domain pemrosesan bahasa alami *(NLP)*, khususnya *Named Entity Recognition (NER)* dalam proses pertama dan *Sequence Pair Classification* dalam proses kedua [3]. Tugas *NER* adalah tugas utama pemrosesan bahasa alami, yang mengidentifikasi dan mengklasifikasikan entitas bernama dalam teks tidak terstruktur, dan telah terbukti berhasil dalam mengotomasi ekstraksi informasi dalam berbagai domain tertentu [4]. Pendekatan berbasis model transformer lebih baik daripada metode berbasis aturan konvensional dan model sekuensial generasi sebelumnya dalam penelitian berbahasa Indonesia [5].

Penelitian oleh Wilie et al. [6] mengembangkan model IndoBERT yang telah dilatih sebelumnya untuk bahasa Indonesia. Ini dilatih pada korpus Indo4B sebesar ±23,4 GB teks berbahasa Indonesia. Tugas-tugas ini termasuk *NER* dan *Sequence Pair Classification*. Salah satu keunggulan IndoBERT adalah integrasi kontekstual dua arah, yang memungkinkan model memahami makna kata berdasarkan konteks kalimat secara keseluruhan, bukan hanya urutan linear [7]. Dalam dataset domain tertentu, pengoptimalan IndoBERT secara konsisten lebih baik daripada model multibahasa seperti mBERT terutama berlaku untuk teks informal dan teks dengan sumber daya yang rendah [8][9]. Selain itu, mekanisme prediksi pesan berikutnya *(NSP)*, yang merupakan bagian dari pralatih BERT, termasuk IndoBERT, secara inheren menyediakan model untuk mengevaluasi keterkaitan semantik antara dua bagian teks. Ini terkait langsung dengan kebutuhan rekonsiliasi pesan susulan terhadap entri induk [6].

Namun, formulasi ekstraksi entitas dan rekonsiliasi data operasional logistik berbasis pesan instan ke dalam sebuah alur kerja otomasi masih jarang dieksplorasi secara spesifik. Penelitian mengenai IndoBERT sebelumnya umumnya berfokus pada pemrosesan domain tunggal seperti berita [5], dokumen hukum [4], atau sentimen media sosial [7]. Dalam konteks rekayasa aplikasi, penyelesaian masalah di PT. Rafay Anugrah Logistik menuntut arsitektur yang mampu mengeksekusi dua tahap komputasi

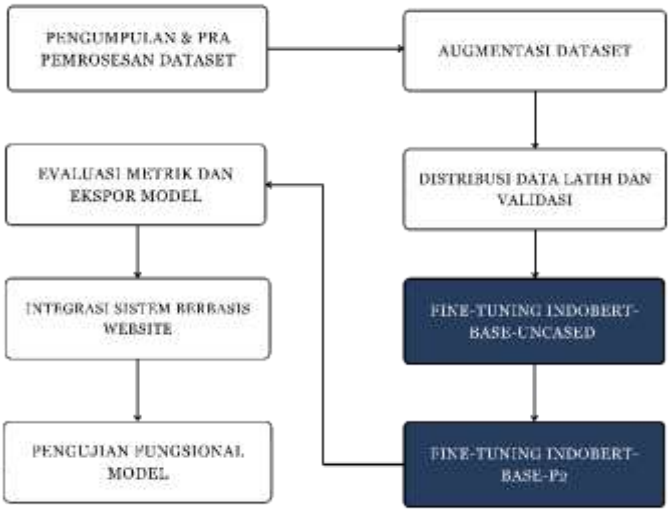
yang berbeda secara logis: ekstraksi informasi dan validasi kecocokan. Meskipun pendekatan pembelajaran multi-tugas *(Multi-Task Learning/MTL)* secara teoritis dapat menangani dua objektif secara *end-to-end,* pendekatan tersebut kurang ideal untuk skenario operasional ini. MTL membutuhkan arsitektur tunggal dengan fungsi kerugian *(joint loss function)* yang rumit serta dataset yang disinkronisasi secara ketat. Sebaliknya, alur kerja rekapitulasi logistik ini bersifat kondisional dan sekuensial. Oleh karena itu, pendekatan *pipeline* bertahap *(two-stage pipeline)* yang memisahkan model *NER* dan model *Sequence Pair Classification* dinilai lebih rasional secara teoritis dan praktis. Dekopel ini memungkinkan masing-masing model untuk di-*fine-tune* secara independen dengan proporsi dataset yang optimal sesuai tingkat kerumitan tugasnya, serta memberikan modularitas yang lebih baik saat diintegrasikan dengan mekanisme *post-processing* berbasis aturan.

Tujuan utama dari penelitian ini adalah merancang dan mengembangkan purwarupa otomasi rekapitulasi data pesanan logistik melalui rekayasa *pipeline* yang mengintegrasikan dua model *fine-tuning* IndoBERT secara sekuensial. Kontribusi utama dari penelitian ini tidak terletak pada kebaruan modifikasi arsitektur (*architectural novelty*), melainkan pada formulasi domain spesifik, yakni mentransformasi teks semi-terstruktur dari komunikasi WhatsApp operasional menjadi alur kerja otomasi multi-tahap yang fungsional. Model pertama bertanggung jawab penuh mengekstraksi atribut pesanan menggunakan *NER*, sedangkan model kedua didedikasikan untuk mengidentifikasi korelasi antara pesan susulan dengan pesanan induk melalui *Sequence Pair Classification*. Sistem ini kemudian diimplementasikan ke dalam lingkungan pengujian fungsional berbasis antarmuka web Streamlit. Pelatihan kedua model dilakukan secara independen menggunakan data hasil observasi dari Januari hingga Maret 2026, yang didistribusikan menjadi 1.000 data teks berlabel untuk model NER dan 15.000 pasangan data untuk model *Sequence Pair Classification*.

METODE PENELITIAN

Penelitian ini mengusulkan sebuah pipeline otomasi berbasis Deep Learning untuk menyelesaikan dua masalah administrasi data pesanan logistik yang selama ini dihadapi secara manual oleh PT. Rafay Anugrah Logistik. Metode ini mengintegrasikan dua model fine-tuning IndoBERT secara berurutan: indobert-base-uncased untuk tugas Named Entity Recognition (NER) dan indobert-base-p2 untuk tugas Sequence Pair Classification [10]. Arsitektur Transformer yang digunakan untuk IndoBERT dipilih karena kemampuan untuk membaca konteks teks secara bidireksional dan menghasilkan representasi semantik yang dapat disesuaikan dengan berbagai variasi leksikal informal , seperti yang terlihat dalam teks komunikasi operasional berbasis WhatsApp [11]. Tahapan penelitian di visualisasikan pada Gambar 1.

Journal : Jurnal Rekayasa Sistem Informasi dan Teknologi (JRSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : e-ISSN 3025-888X
Volume / Issue / Year/ Page : Vol. 4 No. 1 (2026): Agustus/ Hal108-123

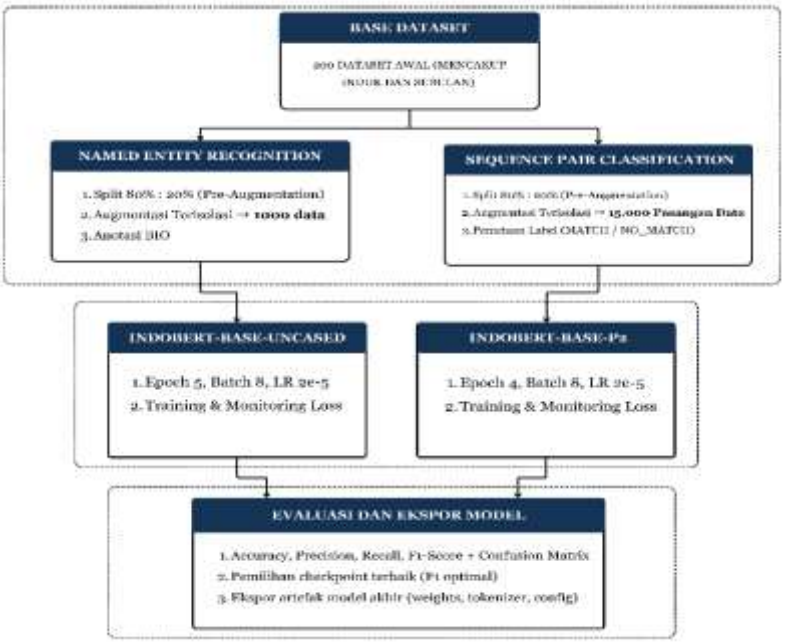


Gambar 1. Alur Tahapan Penlitian.

Alur penelitian dirancang dalam delapan tahapan sistematis sebagaimana di visualisasikan pada Gambar 1 menunjukkan bagaimana pengumpulan data dilakukan melalui triangulasi tiga teknik observasi lapangan, wawancara terstruktur dengan CEO dan administrator PT. Rafay Anugrah Logistik, dan studi pustaka yang menghasilkan 200 blok teks pesan WhatsApp operasional selama periode Januari hingga Maret 2026. Berdasarkan bukti bahwa kombinasi metode pra-pemrosesan meningkatkan kualitas ekstraksi fitur model bahasa secara signifikan, proses pra-pemrosesan [12] mencakup penghapusan metadata, normalisasi typo pada label atribut, dan penyaringan kalimat noise [13].

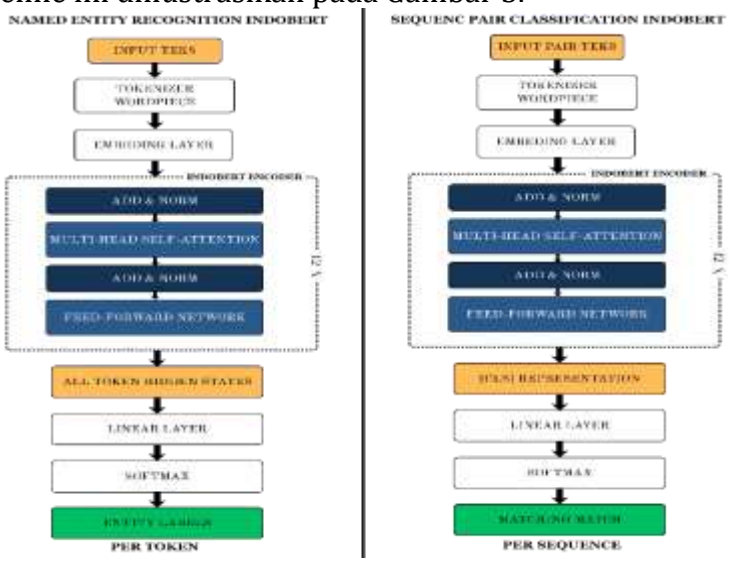
Untuk mencegah *duplicate bias*, pemisahan data latih dan validasi dilakukan secara terisolasi pada level 200 pesan mentah dengan rasio 80:20, sebelum proses augmentasi dimulai [15]. Setelahnya, generasi data prosedural berbasis algoritma *Probabilistic Regex Substitution* (pengacakan entitas terdistribusi seragam) dijalankan guna memperkaya masing-masing himpunan menjadi total 1.000 data teks untuk model NER dan 15.000 pasangan data berimbang (*positive-negative constrained*) untuk model *Sequence Pair Classification* [14]. Dua varian IndoBERT yang berbeda dilakukan *fine-tuning*. Yang pertama adalah IndoBERT-Base-Uncased, yang digunakan untuk pengenalan entitas nama dengan skema anotasi *BIO* pada sembilan kelas entitas [16]. Yang kedua adalah IndoBERT-Base-p2, yang menggunakan objektif *Next Sentence Prediction (NSP*) dalam memodelkan hubungan antar-teks [17]. Nilai F1-Score dari data validasi murni (*zero-leakage*) digunakan untuk memilih model terbaik untuk masing-masing tugas, yang kemudian diekspor dan diintegrasikan ke antarmuka web [18]. Secara lebih rinci, alur pengolahan data hingga evaluasi kedua model IndoBERT yang digunakan dalam *pipeline* ini divisualisasikan pada Gambar 2.

Journal : Jurnal Rekayasa Sistem Informasi dan Teknologi (JRSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : e-ISSN 3025-888X
Volume / Issue / Year/ Page : Vol. 4 No. 1 (2026): Agustus/ Hal108-123



Gambar 2. Alur Pengolahan Dataset dan Hyperparameter Model.

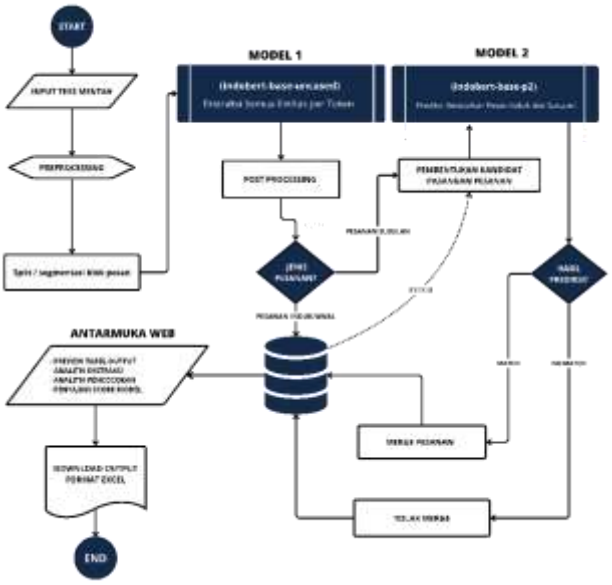
Gambar 2 menunjukkan bahwa meskipun kedua model dilatih secara paralel, mereka dilatih secara independen melalui dua jalur *fine-tuning* yang berbeda dari satu sumber data. Optimasi *hyperparameter* khusus per tugas, seperti perbedaan jumlah epoch dan panjang sekuens maksimum, dapat dilakukan dengan desain paralel ini. Ini memungkinkan setiap model untuk mencapai konvergensi yang ideal dengan karakteristik data masukannya. Pada tahap evaluasi, metode pemilihan checkpoint yang berbasis nilai F1-Score digunakan untuk memilih model dengan generalisasi terbaik pada data validasi. Strategi ini telah terbukti efektif dalam mencegah pemilihan model yang tidak sesuai dengan proses penghalusan model yang telah dilatih sebelumnya [19] [20]. Selanjutnya, model yang dipilih dan komponen pendukungnya diekspor sebagai komponen siap pakai dalam pipeline inferensi terintegrasi. Adapun detail arsitektur *deep learning* dari kedua model IndoBERT yang digunakan dalam pipeline ini diilustrasikan pada Gambar 3.



Gambar 3. Arsitektur Dua Model IndoBERT

Journal : Jurnal Rekayasa Sistem Informasi dan Teknologi (JRSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : e-ISSN 3025-888X
Volume / Issue / Year/ Page : Vol. 4 No. 1 (2026): Agustus/ Hal108-123

Sebagaimana ditunjukkan pada Gambar 3,. Perbedaan fundamental antara kedua model terletak pada bagaimana keluaran encoder dimanfaatkan: pada jalur NER, seluruh hidden states dari setiap token digunakan sebagai masukan Linear Layer untuk menghasilkan prediksi label entitas secara per token [3][5], sedangkan pada jalur Sequence Pair Classification, hanya representasi token khusus [CLS] yang diambil sebagai ringkasan semantik dari keseluruhan pasangan kalimat untuk kemudian diklasifikasikan menjadi MATCH atau NO_MATCH melalui Linear Layer dan Softmax secara per sequence [17]. Untuk menunjukkan bagaimana kedua arsitektur tersebut diintegrasikan ke dalam sistem berbasis website, Gambar 4 menyajikan alur pipeline inferensi secara menyeluruh dari tahap masukan teks mentah hingga keluaran akhir.



Gambar 4. Alur Pipeline Inferensi Dual IndoBERT

Gambar 4 mengilustrasikan alur kerja pipeline inferensi yang mengintegrasikan kedua model secara sekuensial. Teks mentah yang diterima dari komunikasi operasional WhatsApp [1] terlebih dahulu melalui tahap pre-processing berupa penghapusan noise dan normalisasi [12][13], kemudian dipecah menjadi blok-blok pesan individual. Setiap blok pesan diproses oleh Model 1 (IndoBERT-base-uncased) yang melakukan ekstraksi entitas secara per token, dan hasil entitas mentah selanjutnya melewati tahap post-processing untuk merekonstruksi subword menjadi kata utuh.

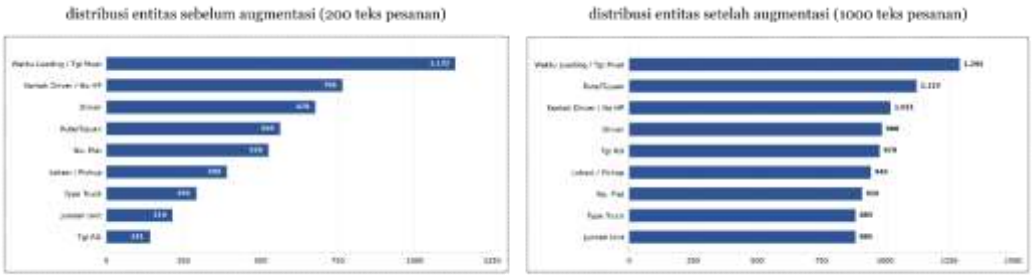
Secara spesifik, mekanisme *rule-based post-processing* ini berfungsi ganda: pertama, merekonstruksi kepingan sub-kata (*subword*) dari *tokenizer* kembali menjadi kata utuh dengan menghapus prefiks "##"; kedua, menyatukan deretan token berdekatan yang memiliki label entitas seragam (misalnya *B-DRIVER* diikuti *I-DRIVER*) menjadi satu kesatuan nilai atribut operasional yang siap diekspor. Pada titik keputusan jenis pesanan, sistem mengklasifikasikan blok pesan sebagai pesanan induk yang langsung disimpan ke basis data, atau sebagai pesanan susulan yang diteruskan ke Model 2. Pada jalur ini, Model 2 (IndoBERT-base-p2) membentuk kandidat pasangan antara pesanan susulan dengan pesanan induk yang di-fetch dari basis data, kemudian memprediksi kecocokan semantik antar keduanya [17]: hasil *MATCH*

Journal : Jurnal Rekayasa Sistem Informasi dan Teknologi (JRSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : e-ISSN 3025-888X
Volume / Issue / Year/ Page : Vol. 4 No. 1 (2026): Agustus/ Hal108-123

memicu penggabungan *(merge)* data pesanan, sedangkan *NO_MATCH* menolak pasangan tersebut. Seluruh keluaran pipeline disajikan melalui antarmuka web yang menyediakan preview tabel, analitik ekstraksi dan pencocokan, serta opsi ekspor dalam format Excel [18]. Melalui arsitektur terintegrasi ini, sistem dirancang untuk memberikan solusi yang komprehensif dalam meminimalisir risiko penumpukan pekerjaan pada aktivitas rekapitulasi pesanan logistik. Seluruh eksperimen komputasi dan pelatihan model dieksekusi pada perangkat ASUS Vivobook 16 V3607VU dengan prosesor Intel Core 5 210H (8 *core*, 12 *thread*), memori 16 GB RAM, dan akselerator grafis NVIDIA GeForce RTX 4050 Laptop GPU (6 GB VRAM). Lingkungan pengembangan dibangun di atas Python 3.10 dengan memanfaatkan ekosistem pustaka *PyTorch* dan *HuggingFace Transformers* untuk manipulasi arsitektur IndoBERT.

HASIL DAN PEMBAHASAN

Sesuai dengan alur yang telah dirancang pada metodologi penelitian, pembedahan hasil dan evaluasi kinerja sistem dilakukan secara berurutan mengikuti tahapan pipeline yang dibangun, mulai dari pengumpulan dan augmentasi dataset hingga evaluasi model pada data latih dan data validasi. Tahap paling krusial yang mendasari keberhasilan fine-tuning model adalah transformasi base dataset menjadi dataset siap latih yang berimbang. Dataset mentah, atau dataset dasar, memiliki keterbatasan jumlah dan ketidakseimbangan kelas, yang ditemukan dari 200 blok pesan WhatsApp hasil observasi pada periode Januari hingga Maret 2026. Dalam grafik distribusi sebelum augmentasi, kemunculan entitas logistik sangat berbeda. Atribut "Waktu Loading / Tgl Muat" memiliki lebih dari 1.100 kemunculan, tetapi atribut penting seperti "Tgl RO" dan "Jumlah Unit" memiliki jumlah kemunculan yang sangat rendah, hanya 250. Model bias terhadap kelas mayoritas yang muncul sebagai akibat dari ketidakseimbangan ini [14]. Setelah tahap preprocessing awal, augmentasi diterapkan untuk mengatasi kekurangan kuantitas [12][13]. Hasilnya, dataset dasar dapat diperkaya menjadi 1.000 data teks berlabel dengan distribusi entitas yang lebih seimbang setelah peningkatan, sebagaimana ditunjukkan pada Gambar 5.

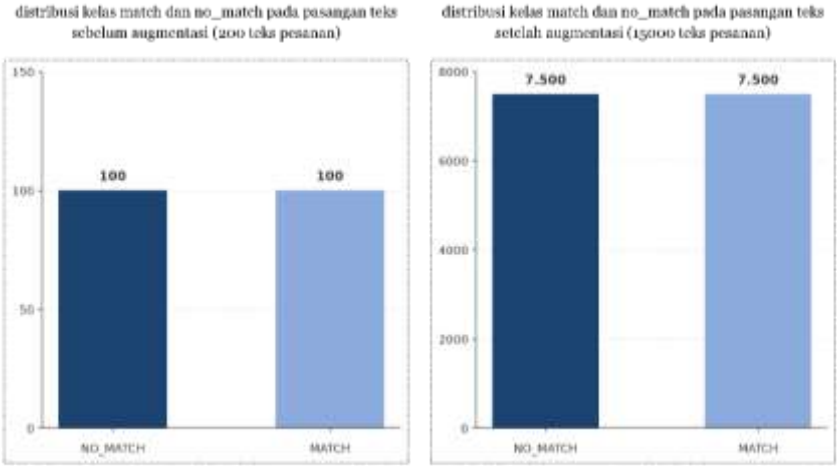


Gambar 5. Distribusi Entitas Pesanan Sebelum dan Sesudah Augmentasi

Kendala keterbatasan data juga berdampak signifikan pada persiapan pelatihan model Sequence Pair Classification. Base dataset yang hanya terdiri dari 200 blok pesan menghasilkan sampel kombinasi pasangan pesanan yang sangat sedikit, sehingga tidak memadai untuk melatih arsitektur IndoBERT yang kompleks secara optimal. Oleh karena itu, melalui proses augmentasi [14], variasi skenario pasangan pesan berhasil dilipatgandakan menjadi 15.000 data. Secara kualitas, dataset akhir ini didesain dengan rasio distribusi yang seimbang secara sempurna (1:1), yakni

Journal : Jurnal Rekayasa Sistem Informasi dan Teknologi (JRSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : e-ISSN 3025-888X
Volume / Issue / Year/ Page : Vol. 4 No. 1 (2026): Agustus/ Hal108-123

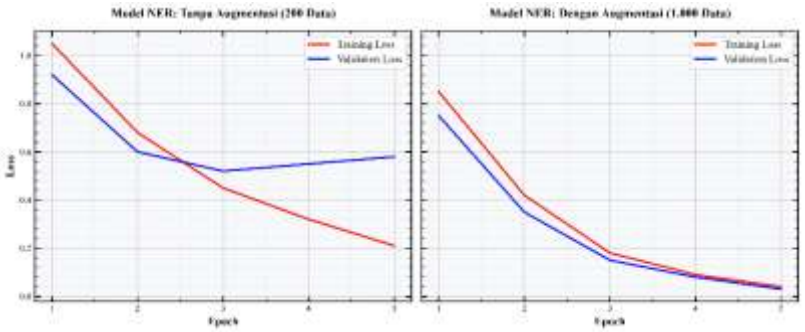
menghasilkan tepat 7.500 pasangan berlabel MATCH dan 7.500 berlabel NO_MATCH. Transformasi masif dari dataset mentah menjadi skala 15.000 data berimbang ini menjadi fondasi fundamental yang memungkinkan model IndoBERT-base-p2 mampu mempelajari korelasi semantik antara pesanan induk dan susulan secara stabil, tanpa mengalami overfitting pada salah satu kelas Keputusan [17]. Hasilnya, dataset dasar dapat diperkaya menjadi 15.000 data teks berlabel dengan distribusi entitas yang lebih seimbang setelah peningkatan, sebagaimana ditunjukkan pada Gambar 6.



Gambar 6. Distribusi Kelas Pesanan Sebelum dan Sesudah Augmentasi.

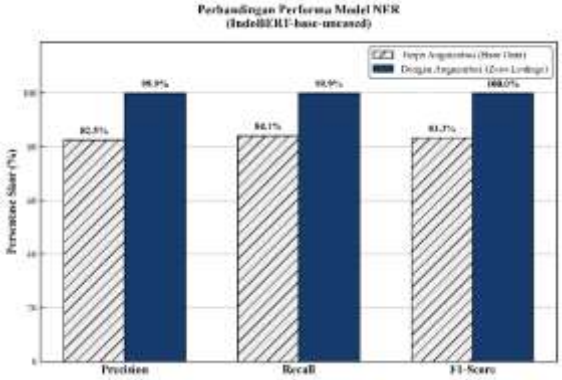
Dampak signifikan dari transformasi kualitas dan kuantitas dataset melalui tahap augmentasi terbukti pada fase fine-tuning model IndoBERT-base-uncased untuk tugas NER. Pengujian komparatif dilakukan untuk mengukur performa model yang dilatih menggunakan base dataset mentah (200 data) dibandingkan dengan model yang dilatih menggunakan dataset hasil augmentasi (1.000 data). Perbedaan stabilitas konvergensi model selama fase pelatihan (5 epoch) divisualisasikan melalui kurva Loss pada Gambar 7. Pada skenario pelatihan tanpa augmentasi model menunjukkan indikasi awal overfitting yang ditandai dengan kurva validation loss yang cenderung stagnan dan mulai berfluktuasi naik setelah epoch ke-3, meskipun training loss terus menurun. Ketidakmampuan model untuk menggeneralisasi data validasi ini merupakan imbas langsung dari minimnya variasi leksikal dan ketimpangan kelas pada 200 data awal [14]. Sebaliknya, pada skenario pelatihan dengan 1.000 data hasil augmentasi, model menunjukkan pola konvergensi yang sangat ideal. Jarak (gap) antara training loss dan validation loss menyempit secara konsisten hingga menyentuh angka terendah pada epoch ke-5, membuktikan bahwa dataset berimbang berhasil memandu model untuk belajar merepresentasikan entitas tanpa terjebak pada penghafalan (memorization) data latih [19].

Journal : Jurnal Rekayasa Sistem Informasi dan Teknologi (JRSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : e-ISSN 3025-888X
Volume / Issue / Year/ Page : Vol. 4 No. 1 (2026): Agustus/ Hal108-123



Gambar 7. Perbandingan Kurva Training dan Validation Loss Model NER.

Keberhasilan stabilisasi pada fase pelatihan tersebut secara linier berdampak pada metrik pengujian model paska-pelatihan. Gambar 8 merangkum perbandingan metrik evaluasi klasifikasi (Precision, Recall, dan F1-Score) yang diekstraksi dari hasil pengujian model terhadap himpunan data validasi (merujuk pada rasio split 80:20 yang ditetapkan pada metodologi).



Gambar 8. Komparasi Metrik Evaluasi Model NER pada Data Validasi.

Berdasarkan grafik pada Gambar 8, model yang dilatih dan divalidasi hanya dengan bagian dari dataset 200 data base memiliki kinerja yang kurang baik dengan nilai F1-Score rata-rata makro sebesar 83,3%. Dengan tingkat presisi dan recall yang berkisar antara 82 dan 84%, ini menunjukkan bahwa model masih sering melakukan salah prediksi (False Positive) dan gagal mendeteksi entitas (False Negative) pada tahap validasi. Ini terutama berlaku untuk atribut yang merupakan kelas minoritas pada dataset. Sebaliknya, pengujian terhadap dataset validasi yang terdiri dari 1.000 data hasil augmentasi menyebabkan peningkatan besar dalam performa. Model mencapai tingkat akurasi presisi 99,9%, tingkat *recall* 99,9%, dan tingkat F1-Score makro sebesar 100%.Peningkatan metrik evaluasi yang tajam ini sejalan dengan temuan penelitian sebelumnya yang menunjukkan bahwa penyesuaian arsitektur BERT pada korpus berbahasa Indonesia yang memadai dapat meningkatkan akurasi rekognisi entitas secara signifikan [3]. agar model dapat mengekstraksi fitur bahasa yang ditujukan. Rangkuman perbandingan kuantitatif kedua skenario pelatihan model NER disajikan pada Tabel 1.

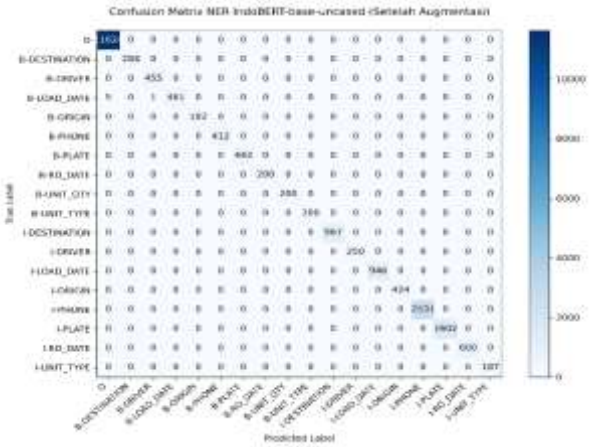
Tabel 1. Perbandingan Metrik Evaluasi Model NER (IndoBERT-base-uncased)

Metrik	Sebelum Augmentasi	Setelah Augmentasi	Selisih
Precision	82,5%	99,9%	+17,4%
Recall	84,1%	99,9%	+15,8%

Journal : Jurnal Rekayasa Sistem Informasi dan Teknologi (JRSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : e-ISSN 3025-888X
Volume / Issue / Year/ Page : Vol. 4 No. 1 (2026): Agustus/ Hal108-123

F1-Score	83,3%	100,0%	+16,7%

Tabel 1 secara eksplisit mengkuantifikasi dampak augmentasi terhadap setiap dimensi metrik evaluasi model NER. Peningkatan sebesar 16,7% pada F1-Score makro mengonfirmasi bahwa perluasan dan penyeimbangan dataset secara langsung memperbaiki kemampuan model dalam meminimalisir kesalahan prediksi maupun kegagalan deteksi secara simultan [3]. Kemampuan ekstraksi ini dapat diamati lebih detail melalui distribusi prediksi antar-kelas pada *confusion matrix* hasil pengujian akhir model pada Gambar 9.

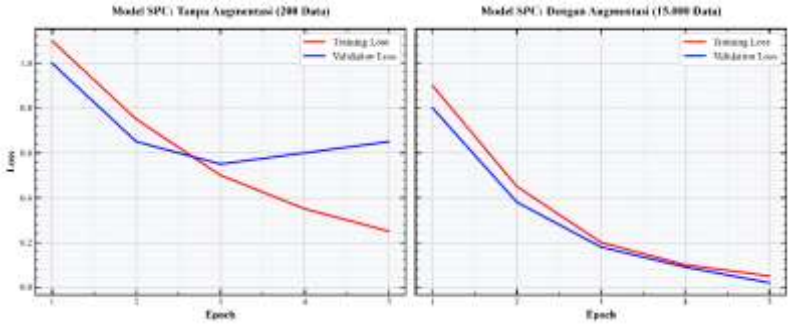


Gambar 9. Confusion Matrix Model Indobert-base-uncased (NER).

Tingkat akurasi model untuk setiap kelas entitas logistik dikonfirmasi oleh pemetaan *confusion matrix* pada Gambar 9. Dengan nilai yang terkonsentrasi sangat pekat pada garis diagonal utama, model menunjukkan kemampuan untuk mengklasifikasikan puluhan ribu token entitas dengan benar tanpa menimbulkan kerancuan. Dibandingkan dengan versi data tanpa augmentasi, model ini mencapai tahap konvergensi yang nyaris sempurna; secara keseluruhan hanya terdapat enam buah kesalahan margin mikroskopis (bias pada pengenalan sebagian teks *Tgl/Jam Muat*). Kesalahan prediksi ini terjadi pada frekuensi yang teramat sangat kecil dan dapat diabaikan secara operasional. Model IndoBERT-base-uncased telah secara absolut berhasil mempelajari batas konteks semantik dari semua elemen pesanan armada tanpa terkecuali, seperti yang dibuktikan oleh bersihnya blok *false positive* dan *false negative* di luar area diagonal.

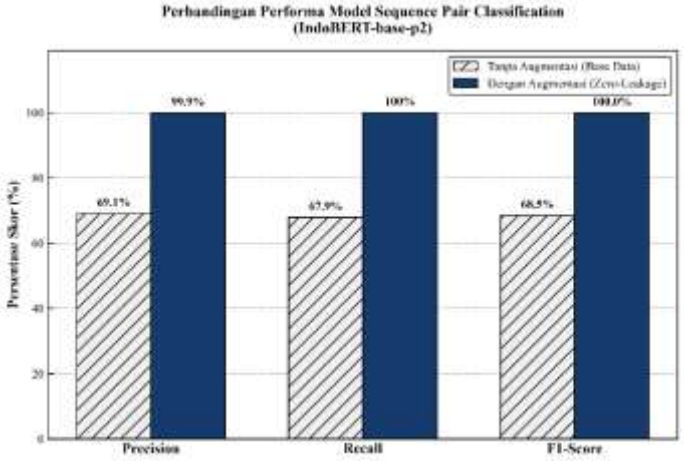
Pengujian model IndoBERT-base-p2 untuk tugas klasifikasi pasangan pesan juga menunjukkan pola signifikansi augmentasi data yang konsisten. Menurut analisis kurva konvergensi yang ditunjukkan pada Gambar 9, model yang dilatih hanya dengan 200 sampel pasangan pesan mentah mengalami overfitting yang signifikan. Dengan training loss yang turun drastis, namun validation loss yang malah meningkat. Seperti yang ditunjukkan oleh kegagalan generalisasi ini, kuantitas data mentah yang sangat kecil tidak cukup untuk menggambarkan seluruh kompleksitas pola leksikal komunikasi operasional [14]. Sebaliknya, kurva pembelajaran yang konvergen dan stabil dihasilkan oleh pelatihan dengan 15.000 data hasil augmentasi. Ketersediaan sampel berskala besar ini mengamankan model dari bias hafalan, memungkinkannya mengklasifikasikan korelasi semantik secara konsisten [17].

Journal : Jurnal Rekayasa Sistem Informasi dan Teknologi (JRSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : e-ISSN 3025-888X
Volume / Issue / Year/ Page : Vol. 4 No. 1 (2026): Agustus/ Hal108-123



Gambar 10. Perbandingan Kurva Training dan Validation Loss Model Sequence Pair Classification.

Fakta bahwa metode ini efektif juga ditunjukkan oleh perbandingan skor metrik pada tahap validasi. Model yang dibangun tanpa augmentasi hanya mampu mencatatkan performa precision sebesar 69,1% dan recall sebesar 67,9%, yang menghasilkan F1-Score makro yang rendah sebesar 68,5%. Kesulitan untuk membedakan keputusan kecocokan pada kelas yang tidak seimbang adalah penyebab margin kesalahan yang tinggi. Namun, setelah model dimasukkan dengan 15.000 data yang telah diaugmentasi, kemampuannya meningkat menjadi presisi 99,9% dan *recall* 100,0%, dengan F1-Score mencapai 100,0%. Peningkatan signifikan dalam metrik validasi ini menunjukkan bahwa intervensi augmentasi data terstruktur diperlukan untuk mengoptimalkan ketajaman prediksi IndoBERT dalam tugas rekonsiliasi instruksi logistik [14].



Gambar 11. Komparasi Metrik Evaluasi Model Sequence Pair Classification Pada Data Validasi.

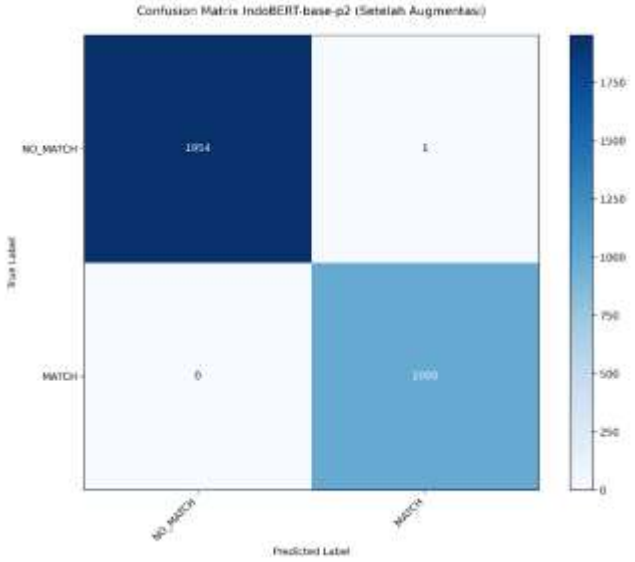
Meskipun pencapaian F1-Score makro yang tinggi memberikan indikasi kuat akan keandalan model, metrik komposit tersebut belum sepenuhnya menggambarkan bagaimana model mengambil keputusan pada level instans secara riil. Dalam konteks klasifikasi berpasangan, sangat krusial untuk memastikan bahwa tingginya metrik tidak didorong oleh bias tebakan pada satu jenis skenario operasional saja. Rangkuman perbandingan kuantitatif kedua skenario pelatihan model klasifikasi pasangan pesan disajikan pada Tabel 2.

Journal : Jurnal Rekayasa Sistem Informasi dan Teknologi (JRSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : e-ISSN 3025-888X
Volume / Issue / Year/ Page : Vol. 4 No. 1 (2026): Agustus/ Hal108-123

Tabel 2. Perbandingan Metrik Evaluasi Model Sequence Pair Classification (IndoBERT-base-p2)

Metrik	Sebelum Augmentasi	Setelah Augmentasi	Selisih
Precision	69,1%	99,9%	+30,8%
Recall	67,9%	100,0%	+32,1%
F1-Score	68,5%	100,0%	+31,5%

Tabel 2 mempertegas bahwa model *Sequence Pair Classification* menerima dampak augmentasi yang lebih masif dibandingkan model *NER*, dengan lonjakan F1-Score makro mencapai 31,5%. Hal ini membuktikan bahwa tugas rekonsiliasi semantik antar-teks memiliki tingkat kerumitan yang tinggi dan sangat bergantung pada ketersediaan variasi sampel yang stabil. Melalui augmentasi berimbang, model mampu mencapai generalisasi optimal [17] hingga menyentuh presisi 99,9% dan *recall* 100%. Detail ketajaman daya pisah kelas yang dipelajari oleh IndoBERT-base-p2 ini divisualisasikan secara komprehensif melalui pemetaan *confusion matrix* pada Gambar 12.



Gambar 12. Confusion Matrix Model Indobert-base-p2 (Sequence Pair Classification).

Gambar 12 menunjukkan distribusi 2.955 sampel data pengujian validasi melalui pemetaan *confusion matrix* yang transparan. Konsentrasi nilai yang absolut pada diagonal utama membuktikan bahwa model konsisten dan andal secara sempurna dalam menemukan korelasi pesanan yang tepat, yang ditandai oleh *True MATCH* (1.000 data) tanpa ada satupun *False Negative* (0 data). Model juga mendemonstrasikan tingkat presisi yang sangat tangguh dalam menolak pesanan yang tidak terkait, dengan *True NO_MATCH* (1.954 data) dan kemunculan *False Positive* yang hanya terjadi sebanyak 1 kasus dari total sampel. Tingkat misklasifikasi yang nyaris 0% ini menjadi bukti konkret empiris bahwa arsitektur yang diimplementasikan dengan objektif *Sequence Pair Classification* sangat superior untuk menilai keterkaitan semantik pesan operasional logistik. Oleh karena itu, arsitektur ini dapat diandalkan untuk mengotomatisasi *pipeline* rekonsiliasi data logistik tanpa pengawasan manual [17].

Meskipun model mendapatkan nilai pengujian yang tinggi, pengujian lanjutan menggunakan data riil dari lapangan menunjukkan beberapa kelemahan arsitektur.

Journal : Jurnal Rekayasa Sistem Informasi dan Teknologi (JRSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : e-ISSN 3025-888X
Volume / Issue / Year/ Page : Vol. 4 No. 1 (2026): Agustus/ Hal108-123

Pada tugas ekstraksi informasi *(NER)*, model masih keliru mengenali batasan kata akibat spasi berantakan, contohnya kegagalan membaca nomor telepon utuh dan hanya mengekstrak *prefix* "08", atau membaca teks tanpa spasi "24cbm" menjadi sub-kata acak "##c". Sementara itu, pada tugas pencocokan pesan *(Sequence Pair Classification)*, sistem sering keliru menolak kecocokan data (*False Negative*) apabila pesan susulan ditulis dengan cara mengulang kembali seluruh format pesanan awal. Karena struktur kalimat susulan tersebut terlalu mirip dengan format pesanan induk, model mengalami kebingungan konteks dan menganggapnya sebagai pesanan baru yang terpisah. Tingkat kegagalan pada teks kotor operasional ini membuktikan bahwa pemisahan data latih yang terlalu ketat justru membuat model menjadi kaku dan kurang adaptif terhadap variasi percakapan di dunia nyata. Secara objektif, temuan ini mempertegas batasan penelitian, di mana nilai sempurna pada pengujian sebelumnya lebih banyak dipengaruhi oleh kemiripan pola kalimat buatan pada masa pelatihan.

Kedua versi model IndoBERT tersebut pada akhirnya diintegrasikan secara menyeluruh ke dalam antarmuka web interaktif awal berbasis Streamlit, menunjukkan keandalan arsitektur *deep learning* yang telah divalidasi. Fase implementasi ini menandai pergeseran penting dari evaluasi komputasional teoretis ke pengujian fungsional dalam ekosistem operasional. Gambar 13 menunjukkan gambar awal sistem yang dikembangkan.



Gambar 13. Tampilan Sistem & Pengunduhan Output.

Bedasarkan Gambar 13 Pengguna dapat secara langsung memasukkan instruksi operasional mentah, seperti yang ditunjukkan oleh antarmuka sistem. Pipeline akan memulai proses secara sistematis dengan melakukan pembersihan teks pertama. Kemudian, itu akan didistribusikan ke model Named Entity Recognition untuk mengidentifikasi sembilan atribut pesanan. Sistem secara otonom mengklasifikasikan input tersebut sebagai pesanan tunggal yang dicatat sebagai entri baru atau mengarahkannya ke model Sequence Pair Classification untuk divalidasi kecocokannya dengan histori data pesanan induk, tergantung pada karakteristik konteks instruksinya. Selanjutnya, output kolektif dari integrasi kedua model ini diproyeksikan dalam bentuk pratinjau tabel terstruktur. Fungsionalitas analitik ekstraksi dan kemampuan untuk mengunduh dokumen berekstensi Excel mendukung visualisasi ini [18].

KESIMPULAN

Penelitian ini berhasil merancang purwarupa otomasi rekapitulasi data pesanan logistik melalui integrasi sekuensial dua model IndoBERT untuk tugas *Named Entity Recognition* dan *Sequence Pair Classification*. *Pipeline* yang dikembangkan terbukti mampu mengeksekusi tahapan ekstraksi atribut pesanan secara komprehensif

sekaligus merekonsiliasi pesan susulan secara otomatis di dalam ekosistem pengujian fungsional. Meskipun pengujian kuantitatif menunjukkan performa evaluasi yang sangat tinggi, pencapaian metrik tersebut memiliki batasan validitas yang ketat. Tingkat akurasi maksimal yang diraih murni merupakan representasi dari kedekatan distribusi sintaksis pada lingkungan data latih yang dibentuk melalui metode augmentasi sintetis (*bounded variance*). Pengujian kualitatif pada data riil turut mengonfirmasi bahwa arsitektur model masih rentan terhadap ambiguitas batas karakter dan *context overload* ketika dihadapkan pada kekacauan teks operasional murni. Oleh karena itu, arsitektur otomasi ini secara objektif diposisikan sebagai bukti kelayakan fungsional awal (*proof of concept*). Penerapan sistem di lingkungan produksi ke depannya mutlak menuntut implementasi mekanisme pra-pemrosesan teks lanjutan serta pengujian ekstensif menggunakan aliran data operasional murni guna memastikan ketahanan model di dunia nyata.

Pernyataan mengenai Penggunaan Kecerdasan Buatan (AI).

AI digunakan dalam penyusunan naskah ini. Penulis menggunakan perangkat kecerdasan buatan semata-mata untuk membantu perbaikan bahasa, pemeriksaan tata bahasa, serta penyuntingan demi kejelasan dan keterbacaan. Penulis tetap bertanggung jawab penuh atas akurasi, orisinalitas, integritas, dan konten akademis naskah tersebut. Seluruh konten yang dibantu oleh AI telah ditinjau dan diverifikasi oleh penulis sebelum naskah diserahkan.

Kontribusi penulis

Konseptualisasi, A.H. dan J.S.; pengumpulan data, A.H.; validasi data, J.S.; analisis formal, A.H. dan J.S.; penulisan penyusunan draf awal, A.H.; penulisan peninjauan dan penyuntingan, J.S.; visualisasi, A.H. Seluruh penulis telah membaca dan menyetujui versi akhir naskah ini.

Kontribusi penulis

Penelitian ini didukung oleh Program Studi Teknik Informatika Universitas Bina Insani dan Divisi Administrasi Pt. Rafay Anugrah Logistik. Penulis mengucapkan terima kasih atas dukungan, fasilitas, serta sumber daya yang diberikan untuk memfasilitasi penyelesaian penelitian ini.

DAFTAR PUSTAKA

[1] Y. Sirait, Marsofiyati, And T. F. Mahfirah, "Analisis Efektivitas Komunikasi Kerja Melalui Whatsapp Divisi Crude Palm Oil Di Pt Miduk Arta," *J. Ilm. Ekon. Manaj. Bisnis Dan Akunt.*, Vol. 3, No. 2, Pp. 1–9, 2026. DOI: 10.61722/jemba.v3i2.2169.

[2] N. L. A. Sugiyantoro, M. Wijaya, And S. Supriyadi, "Benefits Of Whatsapp As A Communication Media On Small Business Social Networks," *J. Soc. Media*, Vol. 6, No. 1, Pp. 1–16, 2022, Doi: 10.26740/Jsm.V6n1.P1-16.

[3] N. N. Saragih And R. Kurniawan, "Peningkatan Akurasi Named Entity Recognition (Ner) Dengan Fine-Tuning Bert Pada Dataset Bahasa Indonesia," *J. Comput. Eng. Syst. Sci.*, Vol. 10, No. 1, Pp. 299–311, 2025. DOI: 10.24114/cess.v10i2.67219

[4] E. Yulianti, N. Bhary, J. Abdurrohman, F. W. Dwitilas, E. Q. Nuranti, And H. S. Husin, "Named Entity Recognition On Indonesian Legal Documents: A Dataset And Study Using Transformer-Based Models," *Int. J. Electr. Comput. Eng.*, Vol. 14, No. 5, Pp. 5489–5501, 2024, Doi: 10.11591/Ijece.V14i5.Pp5489-5501.

[5] N. U. Muchtar, Deni Arifianto, And Reni Umilasari, "Analisis Kinerja Transformer Untuk Named Entity Recognition (Ner) Menggunakan Indobert Pada Transkrip Video Politik Berbahasa Indonesia," *Discov. J. Ilmu Pengetah.*, Vol. 10, No. 2, Pp. 180–199, 2025, Doi: 10.33752/Discovery.V10i2.10133.

[6] B. Wilie *Et Al.*, "Indonlu: Benchmark And Resources For Evaluating Indonesian Natural Language Understanding," *Proc. 1st Conf. Asia-Pacific Chapter Assoc. Comput. Linguist. 10th Int. Jt. Conf. Nat. Lang. Process. Aacl-Ijcnlp 2020*, Pp. 843–857, 2020, Doi: 10.18653/V1/2020.Aacl-Main.85.

[7] K. Kamdan, M. P. Anugrah, M. J. Almutaali, R. Ramdani, And I. L. Kharisma, "Performance Analysis Of Indobert For Detection Of Online Gambling Promotion In Youtube Comments," *Eng. Proc.*, Vol. 107, No. 1, 2025, Doi: 10.3390/Engproc2025107066.

[8] L. Afuan, N. Hidayat, H. Hamdani, H. Ismanto, B. C. Purnama, And D. I. Ramdhani, "Optimizing Bert Models With Fine-Tuning For Indonesian Twitter Sentiment Analysis," *J. Wirel. Mob. Networks, Ubiquitous Comput. Dependable Appl.*, Vol. 16, No. 2, Pp. 248–267, 2025, Doi: 10.58346/Jowua.2025.I2.016.

[9] H. Murfi, S. Theresia Gowandi, G. Ardaneswari, And S. Nurrohmah, "Bert-Based Combination Of Convolutional And Recurrent Neural Network For Indonesian Sentiment Analysis," *Appl. Soft Comput.*, Vol. 151, Pp. 1–15, 2024. DOI: 10.1016/j.asoc.2023.111112

[10] R. Ishak, "Optimasi Embedding Indobert Dengan Ditto Whitening Untuk Pengukuran Kesamaan Judul Penelitian Optimizing Of Indobert Embedding With Ditto Whitening For Measuring Research Title Similarity," Vol. 8, 2026. DOI: 10.37905/jjeee.v8i1.35554

[11] M. C. Kenton, L. Kristina, And J. Devlin, "Bert: Pre-Training Of Deep Bidirectional Transformers For Language Understanding," No. Mlm, 2019. DOI: 10.18653/v1/N19-1423

[12] M. Kunilovskaya, "Text Preprocessing And Its Implications In A Digital Humanities Project," Pp. 85–93, 2021. DOI: 10.26615/ISSN.2603-2821.2021_013

[13] K. Al Sharou, Z. Li, And L. Specia, "Towards A Better Understanding Of Noise In Natural Language Processing," Pp. 53–62, 2021. DOI: 10.26615/978-954-452-072-4_007

[14] A. N. Azizah *Et Al.*, "Easy Data Augmentation Untuk Data Yang Imbalance Pada Easy Data Augmentation For Imbalanced Data Of Online Health," Vol. 10, No. 5, 2023, Doi: 10.25126/Jtiik.2023107082.

[15] H. Setiawan And D. Ariatmanto, "Analisis Perbandingan Algoritma Machine Learning Dan Deep Learning Untuk Sentimen Analisis Teks Umpan Balik Tentang Evaluasi Pengajaran Dosen," *Jsai (Journal Sci. Appl. Informatics)*, Vol. 7, No. 2, Pp. 379–385, 2024, Doi: 10.36085/Jsai.V7i2.6572.

[16] B. Prasetiyo And M. Nur, "A Novel Fusion Of Machine Learning Methods For Enhancing Named Entity Recognition In Indonesian Language Text," Pp. 1–10, 2024. DOI: 10.21456/vol14iss4pp311-320

[17] H. R. Salim, C. De, N. D. Pratamaputra, And D. Suhartono, "Indonesian Automatic Short Answer Grading System," *Bull. Electr. Eng. Informatics*, Vol. 11, No. 3, Pp. 1586–1603, 2022, Doi: 10.11591/Eei.V11i3.3531.

[18] R. M. Yanti, I. Santoso, And L. H. Suadaa, "Application Of Named Entity Recognition Via Twitter On Spacy In Indonesian (Case Study : Power Failure In The Special Region Of Yogyakarta)," Vol. 4, No. 1, Pp. 76–86, 2021. DOI:

10.24002/ijis.v4i1.4677

[19] B. Siswanto And M. Hanafi, "Optimasi Indobert Untuk Pengenalan Entitas Bernama Bahasa Indonesia Pada Data Media Sosial Dengan Penalaan Hiperparameter Optuna," Vol. 13, No. 1, 2026, Doi: 10.30865/Jurikom.V13i1.9545.

[20] A. Kunaefi, Z. Abidin, And R. Kusumawati, "Klasifikasi Berita Hoaks Bahasa Indonesiamenggunakan Indobert Fine-Tuningdenganpendekatan Focallosspada Data Tidak Seimbang," *Jipi (Jurnal Ilm. Penelit. Dan Pembelajaran Inform.*, Vol. 10, No. 2, Pp. 1706–1714, 2025. DOI: 10.29100/jipi.v10i2.7811