ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Penerapan IndoBERT untuk Klasifikasi Curriculum Vitae Indonesia pada Tahap Early Screening

Egiofani Renedio[1], Rizky Parlika[2*], Budi Mukhamad Mulyo[3]

[1,2,3] Informatika, Universitas Pembangunan Nasional "Veteran" Jawa Timur
[1]21081010056@student.upnjatim.ac.id
[3]budi.m.mulyo.fasilkom@upnjatim.ac.id

*Corresponding author email: rizkyparlika.if@upnjatim.ac.id

***Abstrak*— Proses rekrutmen karyawan sering melibatkan penyaringan sejumlah besar dokumen Curriculum Vitae (CV) yang umumnya masih dilakukan secara manual. Kondisi ini memerlukan waktu yang relatif lama serta berpotensi menimbulkan subjektivitas dalam proses seleksi administrasi. Permasalahan tersebut semakin kompleks karena dokumen CV memiliki struktur yang beragam dan sering menggunakan campuran bahasa Indonesia dan bahasa Inggris. Penelitian ini menerapkan model bahasa berbasis Transformer, yaitu IndoBERT, untuk melakukan klasifikasi dokumen CV berbahasa Indonesia secara otomatis pada tahap *early screening*. Dataset yang digunakan terdiri dari 1140 dokumen CV dari proses rekrutmen nyata dengan distribusi kelas yang tidak seimbang. Tahapan pra-pemrosesan meliputi ekstraksi teks, anonimisasi data pribadi, normalisasi teks, serta segmentasi dokumen menggunakan teknik *sliding window*. Model kemudian dilakukan *fine-tuning* dan dievaluasi menggunakan metrik Accuracy, Precision, Recall, F1-score, serta Precision–Recall Area Under Curve (PR-AUC). Hasil evaluasi pada data uji menunjukkan bahwa model menghasilkan Accuracy sebesar 0,731, Precision sebesar 0,154, Recall sebesar 0,316, F1-score sebesar 0,207, dan PR-AUC sebesar 0,165. Analisis variasi *threshold* menunjukkan adanya *trade-off* antara precision dan recall yang dapat disesuaikan dengan kebutuhan proses rekrutmen. Hasil penelitian ini menunjukkan bahwa pendekatan berbasis IndoBERT memiliki potensi sebagai alat bantu prioritisasi kandidat pada tahap seleksi administrasi.**

Kata Kunci— **curriculum vitae, klasifikasi dokumen, IndoBERT, pemrosesan bahasa alami, rekrutmen.**

I. PENDAHULUAN

Proses rekrutmen karyawan merupakan tahapan strategis dalam manajemen sumber daya manusia karena berperan dalam menentukan kualitas tenaga kerja yang direkrut oleh suatu organisasi. Pada tahap awal rekrutmen, perusahaan umumnya melakukan seleksi administrasi dengan meninjau *Curriculum Vitae* (CV) yang dikirimkan oleh pelamar. Seiring meningkatnya jumlah pencari kerja, perusahaan dapat menerima ratusan hingga ribuan dokumen CV untuk satu posisi pekerjaan, sehingga proses penyaringan manual menjadi tidak efisien dan memerlukan waktu yang relatif lama serta berpotensi menimbulkan subjektivitas dalam pengambilan keputusan[1].

Di Indonesia, permasalahan ini semakin kompleks karena tidak adanya standar baku dalam penulisan CV. Setiap pelamar memiliki format, struktur, dan gaya penulisan yang berbeda-beda. Selain itu, dokumen CV sering menggunakan campuran bahasa Indonesia dan bahasa Inggris, terutama pada bagian pengalaman kerja dan keterampilan. Variasi struktur dan bahasa ini menyulitkan proses evaluasi manual dan menimbulkan tantangan dalam pengolahan dokumen secara otomatis [2].

Sejumlah penelitian terdahulu telah mengkaji otomasi penyaringan resume menggunakan pendekatan pembelajaran mesin konvensional. Representasi teks berbasis *Term Frequency–Inverse Document Frequency* (TF-IDF) yang dikombinasikan dengan algoritma klasifikasi seperti *Logistic Regression* atau *Support Vector Machine* (SVM) banyak digunakan dalam tugas klasifikasi dokumen dan resume screening [3], [4]. Pendekatan tersebut relatif sederhana dan efisien, namun memiliki keterbatasan dalam menangkap hubungan semantik dan konteks mendalam pada dokumen panjang dan tidak terstruktur seperti CV. Selain itu, sebagian besar penelitian sebelumnya menggunakan dataset resume berbahasa Inggris, sehingga evaluasi model bahasa pada dokumen CV berbahasa Indonesia masih terbatas.

Perkembangan signifikan dalam bidang *Natural Language Processing* (NLP) terjadi dengan diperkenalkannya arsitektur Transformer yang memanfaatkan mekanisme *self-attention* untuk memahami hubungan kontekstual antar kata dalam suatu teks [5]. Salah satu model yang paling berpengaruh adalah *Bidirectional Encoder Representations from Transformers* (BERT), yang menghasilkan representasi bahasa dua arah dan menunjukkan peningkatan performa pada berbagai tugas klasifikasi teks dibandingkan metode konvensional [6]. Dalam domain rekrutmen, pendekatan berbasis model bahasa pra-latih mulai diterapkan untuk tugas seperti resume classification dan *candidate-job matching*, dengan hasil yang menunjukkan kemampuan lebih baik dalam memahami konteks dokumen [7]. Untuk bahasa Indonesia, dikembangkan model IndoBERT yang dilatih menggunakan korpus besar berbahasa Indonesia sehingga lebih sesuai untuk tugas pemrosesan teks dalam konteks linguistik Indonesia dibandingkan model multibahasa [8]. Evaluasi pada berbagai benchmark NLP menunjukkan bahwa IndoBERT memberikan performa kompetitif dalam tugas klasifikasi teks berbahasa Indonesia [8],[9].

Meskipun berbagai penelitian telah membahas penggunaan model BERT untuk klasifikasi dokumen, studi yang secara spesifik mengevaluasi kinerja IndoBERT pada klasifikasi *Curriculum Vitae* berbahasa Indonesia menggunakan data

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

rekrutmen nyata dengan distribusi kelas tidak seimbang masih sangat terbatas. Selain itu, analisis pengaruh penentuan threshold terhadap performa model dalam skenario seleksi administrasi belum banyak dibahas secara eksplisit. Selain itu, sebagian besar penelitian sebelumnya belum secara eksplisit menganalisis pengaruh ketidakseimbangan data serta pemilihan threshold terhadap performa model dalam konteks seleksi administrasi.

Berdasarkan permasalahan tersebut, penelitian ini berfokus pada penerapan dan evaluasi model IndoBERT untuk klasifikasi Curriculum Vitae berbahasa Indonesia ke dalam kategori "Lolos" dan "Tidak Lolos". Penelitian ini menitikberatkan pada analisis performa model pada dataset yang tidak seimbang serta evaluasi berbagai nilai threshold untuk mendukung skenario early screening dalam proses rekrutmen. Diharapkan penelitian ini dapat memberikan kontribusi dalam pengembangan sistem penyaringan CV otomatis yang lebih efektif dan kontekstual di Indonesia.

II. LANDASAN TEORI

A. Klasifikasi Dokumen dalam Proses Rekrutmen

Klasifikasi dokumen merupakan salah satu tugas utama dalam *Natural Language Processing* (NLP) yang bertujuan untuk mengelompokkan teks ke dalam kategori tertentu berdasarkan isi dan karakteristiknya. Dalam konteks rekrutmen, klasifikasi dokumen digunakan untuk mengotomatisasi proses penyaringan *Curriculum Vitae* (CV) guna mendukung seleksi administrasi secara lebih efisien, konsisten, dan terukur.

Pemanfaatan kecerdasan buatan dalam bidang *Human Resource* Managemet telah menunjukkan potensi dalam meningkatkan efisiensi proses rekrutmen serta mengurangi bias subjektif dalam evaluasi kandidat [1], [2]. Selain itu, berbagai sistem informasi berbasis teknologi juga telah digunakan untuk mengotomatisasi proses evaluasi dan pengolahan data pada sistem digital guna meningkatkan efisiensi proses administrasi dan pengambilan keputusan [10]. Beberapa penelitian terbaru mengembangkan sistem penyaringan resume berbasis pembelajaran mesin dan deep learning yang mampu mengidentifikasi kesesuaian kandidat terhadap kebutuhan perusahaan secara otomatis [3], [4]. Pendekatan ini menjadi relevan ketika jumlah dokumen yang diproses sangat besar dan memiliki struktur yang tidak seragam seperti CV.

B. *Arsitektur Transformer dan BERT*

Perkembangan signifikan dalam NLP modern dipicu oleh diperkenalkannya arsitektur Transformer yang memanfaatkan mekanisme self-attention untuk memodelkan hubungan kontekstual antar kata dalam suatu teks [5]. Arsitektur ini memungkinkan pemrosesan paralel serta mampu menangkap dependensi jarak jauh secara lebih efektif dibandingkan model berbasis *Recurrent Neural Network* (RNN).

Model bahasa pra-latih seperti BERT memperluas kemampuan Transformer melalui pembelajaran representasi dua arah yang mempertimbangkan konteks dari kedua sisi suatu token [6]. Pendekatan ini terbukti meningkatkan performa pada berbagai tugas klasifikasi teks, terutama pada dokumen panjang dan tidak terstruktur.

Berbagai penelitian menunjukkan bahwa proses fine-tuning model bahasa pra-latih memberikan peningkatan performa signifikan dibandingkan metode berbasis fitur tradisional [7]–[9]. Dalam domain rekrutmen, model berbasis Transformer mulai diterapkan untuk tugas resume classification dan candidate screening, dengan hasil yang menunjukkan peningkatan kemampuan dalam memahami konteks dokumen yang kompleks dan beragam [3], [11].

C. *IndoBERT untuk Bahasa Indonesia*

Untuk bahasa Indonesia, IndoBERT dikembangkan sebagai model bahasa pra-latih yang dilatih pada korpus besar berbahasa Indonesia [12]. Model ini dirancang untuk menangkap karakteristik linguistik lokal yang mungkin tidak sepenuhnya terwakili dalam model multibahasa.

Evaluasi pada berbagai tugas NLP menunjukkan bahwa IndoBERT memberikan performa kompetitif dalam klasifikasi teks berbahasa Indonesia [11][13][14]. Penelitian lanjutan juga menunjukkan bahwa model Transformer berbasis bahasa Indonesia mampu meningkatkan akurasi klasifikasi pada berbagai domain aplikasi [15]. Oleh karena itu, IndoBERT menjadi pilihan yang relevan untuk menangani dokumen CV berbahasa Indonesia yang memiliki variasi struktur, gaya penulisan, serta penggunaan istilah campuran.

D. *Ketidakseimbangan Data (Class Imbalance)*

Ketidakseimbangan kelas (class imbalance) merupakan permasalahan umum dalam aplikasi klasifikasi dunia nyata, termasuk rekrutmen, di mana jumlah kandidat yang tidak lolos biasanya jauh lebih besar dibandingkan kandidat yang lolos. Kondisi ini dapat menyebabkan model cenderung bias terhadap kelas mayoritas dan mengabaikan kelas minoritas [16].

Penelitian menunjukkan bahwa penanganan ketidakseimbangan kelas dapat dilakukan melalui berbagai pendekatan, seperti *class weighting*, *oversampling*, maupun modifikasi fungsi loss [17], [18]. Dalam konteks model berbasis Transformer, penerapan bobot kelas pada fungsi loss merupakan strategi yang umum digunakan untuk meningkatkan sensitivitas terhadap kelas minoritas [19].

E. *Evaluasi Precision-Recall dan Threshold*

Pada dataset yang tidak seimbang, metrik akurasi sering kali kurang representatif karena model dapat memperoleh nilai tinggi hanya dengan memprediksi kelas mayoritas. Oleh karena itu, metrik seperti *Precision*, *Recall*, *F1-score*, serta *Precision-Recall Area Under Curve* (PR-AUC) lebih direkomendasikan untuk mengevaluasi performa model [20].

Kurva *Precision-Recall* dinilai lebih informatif dibandingkan ROC pada kondisi ketidakseimbangan kelas yang tinggi karena lebih sensitif terhadap performa kelas minoritas. Selain itu,

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

penentuan nilai threshold pada model klasifikasi berbasis probabilitas memengaruhi keseimbangan antara precision dan recall [21]. Dalam konteks rekrutmen, penyesuaian threshold menjadi penting untuk mendukung skenario early screening, di mana sistem berfungsi sebagai alat bantu prioritisasi sebelum dilakukan evaluasi manual oleh pihak Human Resource.

III. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimen untuk mengevaluasi kinerja model IndoBERT dalam melakukan klasifikasi dokumen *Curriculum* Vitae (CV) berbahasa Indonesia. Fokus penelitian adalah analisis performa model pada dataset rekrutmen nyata yang bersifat tidak terstruktur dan tidak seimbang.

A. Dataset

Dataset yang digunakan terdiri dari 1140 dokumen CV yang diperoleh dari proses rekrutmen nyata. Dataset berasal dari proses rekrutmen pada sektor teknologi informasi yang telah dianonimkan untuk menjaga privasi kandidat. Dokumen dikategorikan ke dalam dua kelas berdasarkan hasil seleksi administrasi, yaitu Lolos dan Tidak Lolos. Label Lolos dan Tidak Lolos ditentukan berdasarkan keputusan tim Human Resource pada tahap seleksi administrasi. Distribusi dataset secara keseluruhan ditunjukkan pada Tabel I.

TABEL I
DISTRIBUSI KESELURUHAN DATASET

Kelas	Jumlah Dokumen	Presentase (%)
Lolos	124	10.88
Tidak Lolos	1016	89,12
Total	1140	100

Distribusi tersebut menunjukkan ketidakseimbangan kelas yang signifikan, di mana kelas Lolos hanya mencakup sekitar 10,9% dari total data. Kondisi ini mencerminkan situasi nyata dalam proses seleksi administrasi.

Dataset dibagi secara *stratified* menjadi data latih, validasi, dan uji dengan proporsi 70:15:15. Pembagian dilakukan menggunakan *random seed* 42 untuk memastikan konsistensi eksperimen. Rincian pembagian data ditunjukkan pada Tabel II.

TABEL II
PEMBAGIAN DATA LATIH, VALIDASI, DAN UJI

Subset	Total	Lolos	Tidak lolos
Train	798	86	712
Validasi	171	19	152
Test	171	19	152

Pembagian dataset dilakukan secara *stratified* untuk menjaga proporsi distribusi kelas antara data latih, validasi, dan data uji tetap konsisten dengan distribusi pada dataset keseluruhan. Pendekatan ini penting untuk memastikan bahwa model dilatih dan dievaluasi pada distribusi data yang representatif, terutama mengingat adanya ketidakseimbangan kelas yang signifikan pada dataset. Data latih digunakan untuk proses pelatihan model, data validasi digunakan untuk pemilihan model terbaik serta penentuan nilai threshold, sedangkan data uji digunakan untuk evaluasi akhir performa model.

B. Pra-pemrosesan Data

Dokumen CV tersedia dalam format PDF, DOCX, dan TXT. Teks diekstraksi menggunakan pustaka yang sesuai dengan masing-masing format file. Selanjutnya dilakukan pembersihan karakter *non-printable* serta normalisasi spasi untuk menjaga konsistensi teks. Informasi sensitif seperti alamat email, nomor telepon, dan URL dianonimkan menggunakan pola ekspresi reguler guna menjaga privasi data.

Untuk menjaga kualitas korpus, dokumen dengan panjang teks kurang dari 200 karakter tidak disertakan dalam proses pelatihan. Tahapan pra-pemrosesan ini bertujuan untuk memastikan bahwa data yang digunakan memiliki kualitas dan konsistensi yang memadai sebelum diproses oleh model bahasa.

C. Segmentasi Dokumen

Model IndoBERT memiliki batas maksimum panjang token, sehingga dokumen dengan panjang teks yang melebihi batas tersebut perlu dibagi menjadi beberapa segmen menggunakan teknik sliding window. Parameter segmentasi yang digunakan dalam penelitian ini ditunjukkan pada Tabel III.

TABEL III
PARAMETER SEGMENTASI DOKUMEN

Parameter	Nilai
Panjang maksimum token	256
Stride	192
Metode agregasi dokumen	Mean Top-2

Dengan konfigurasi tersebut, jumlah segmen yang dihasilkan pada setiap subset ditunjukkan pada Tabel IV.

TABEL IV
JUMLAH SEGMEN SUBSET

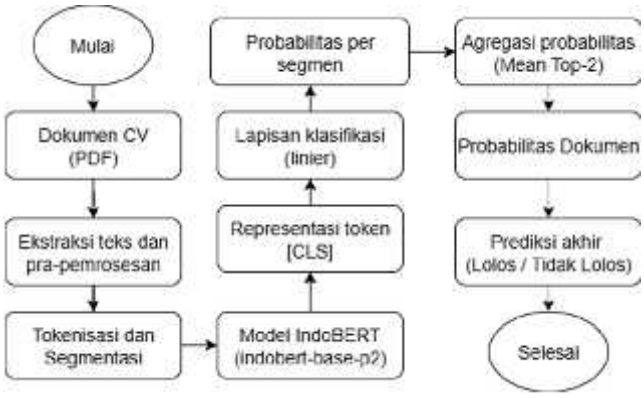
Subset	Jumlah Segmen
Train	3239
Validasi	676
Test	788

Proses pelatihan dilakukan pada tingkat segmen, sedangkan evaluasi dilakukan pada tingkat dokumen. Probabilitas prediksi dokumen diperoleh dengan metode agregasi Mean Top-2, yaitu rata-rata dari dua probabilitas tertinggi dalam satu dokumen. Pendekatan ini digunakan untuk mempertahankan informasi dari bagian dokumen yang paling relevan terhadap keputusan klasifikasi.

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

D. Konfigurasi Model dan Proses Pelatihan

Alur konfigurasi model dan proses pelatihan yang digunakan dalam penelitian ini ditunjukkan pada Gambar 1.



Gbr.1. Alur konfigurasi model dan proses pelatihan

Berdasarkan Gambar 1, proses dimulai dari dokumen Curriculum Vitae (CV) yang tersedia dalam format PDF yang kemudian melalui tahap ekstraksi dan pra-pemrosesan teks. Tahapan ini meliputi pembersihan teks, normalisasi, serta anonimisasi informasi sensitif.

Selanjutnya, teks diproses melalui tokenisasi dan segmentasi menggunakan teknik sliding window untuk mengatasi keterbatasan panjang token pada model IndoBERT. Setiap segmen teks kemudian dimasukkan ke dalam model IndoBERT (indobert-base-p2) untuk menghasilkan representasi kontekstual.

Representasi dari token [CLS] digunakan sebagai representasi keseluruhan segmen, yang selanjutnya diteruskan ke lapisan klasifikasi linier untuk menghasilkan probabilitas prediksi pada tingkat segmen. Probabilitas dari seluruh segmen dalam satu dokumen kemudian diagregasi menggunakan metode Mean Top-2 untuk memperoleh probabilitas pada tingkat dokumen.

Probabilitas hasil agregasi tersebut digunakan untuk menentukan prediksi akhir, yaitu kategori Lolos atau Tidak Lolos.

Model yang digunakan adalah indobenchmark/indobert-base-p2, yaitu model Transformer yang telah dilatih pada korpus bahasa Indonesia. *Fine-tuning* dilakukan dengan menambahkan lapisan klasifikasi linier pada keluaran representasi tersembunyi model.

Konfigurasi pelatihan yang digunakan dalam penelitian ini disajikan pada Tabel V.

TABEL V
KONFIGURASI PELATIHAN MODEL

Parameter	Nilai
Optimizer	AdamW
Learning rate	2×10^{-5}
Weight decay	0,01
Batch size	8
Jumlah epoch	4
Fungsi loss	BCEWithLogitsLoss
Penanganan imbalance	Pos-weight

Nilai pos-weight yang digunakan pada fungsi loss dihitung berdasarkan rasio distribusi kelas pada data latih dan bernilai 8,28. Seluruh eksperimen dilakukan menggunakan *framework* PyTorch dan pustaka *HuggingFace* Transformers dengan random seed 42 untuk menjaga reprodusibilitas.

Untuk menangani ketidakseimbangan kelas, digunakan parameter *pos-weight* pada fungsi loss yang dihitung berdasarkan rasio distribusi kelas pada data latih. Pendekatan ini bertujuan untuk mengurangi bias model terhadap kelas mayoritas.

E. Evaluasi dan Penentuan Threshold

Evaluasi dilakukan pada tingkat dokumen setelah agregasi probabilitas dari segmen, sehingga metrik yang dilaporkan merepresentasikan performa klasifikasi dokumen secara keseluruhan dokumen menggunakan metrik *Accuracy*, *Precision*, *Recall*, *F1-score*, serta *Precision-Recall Area Under Curve* (PR-AUC). Mengingat distribusi kelas yang tidak seimbang, PR-AUC dan *F1-score* digunakan sebagai indikator utama dalam menilai performa model.

Nilai threshold probabilitas tidak ditetapkan secara default pada 0,5, melainkan ditentukan berdasarkan performa terbaik pada data validasi. Penentuan threshold dilakukan dengan memaksimalkan nilai *F1-score* pada data validasi serta mempertimbangkan kebutuhan recall dalam skenario early screening. Threshold yang terpilih kemudian digunakan untuk evaluasi akhir pada data uji.

IV. HASIL DAN PEMBAHASAN

A. *Hasil Pelatihan Model*

Proses fine-tuning model dilakukan selama empat epoch menggunakan data latih yang terdiri dari 798 dokumen (3239 segmen). Evaluasi performa dilakukan pada data validasi pada setiap epoch menggunakan metrik *Precision-Recall Area Under Curve* (PR-AUC) dan F1-score.

Selama proses pelatihan, nilai PR-AUC pada data validasi menunjukkan fluktuasi antar epoch, dengan nilai tertinggi diperoleh pada *epoch* ke-3. Hal ini mengindikasikan bahwa model mulai mencapai konvergensi sebelum epoch terakhir, sehingga pemilihan model terbaik didasarkan pada performa validasi tertinggi untuk menghindari overfitting.

Model terbaik dipilih berdasarkan nilai PR-AUC tertinggi pada data validasi. Pendekatan ini digunakan karena karakteristik dataset yang tidak seimbang, sehingga PR-AUC dinilai lebih representatif dibandingkan akurasi dalam mengukur kemampuan model membedakan kelas minoritas.

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

B. *Hasil Evaluasi pada Data Uji*

Evaluasi akhir dilakukan pada 171 dokumen data uji menggunakan nilai threshold yang diperoleh dari proses optimasi pada data validasi. Penentuan threshold dilakukan dengan memaksimalkan nilai F1-score pada data validasi, sehingga nilai yang digunakan pada tahap ini merepresentasikan konfigurasi terbaik model untuk keseimbangan antara precision dan recall.

TABEL VI
HASIL EVALUASI MODEL PADA DATA UJI

Metrik	Nilai
Accuracy	0,731
Precision	0,154
Recall	0,316
F1-score	0,207
PR-AUC	0,165

Berdasarkan Tabel VI, model menghasilkan nilai Accuracy sebesar 0,731, Precision sebesar 0,154, Recall sebesar 0,316, F1-score sebesar 0,207, dan PR-AUC sebesar 0,165. Nilai akurasi yang relatif tinggi tidak sepenuhnya mencerminkan performa model karena adanya ketidakseimbangan kelas yang signifikan pada dataset.

Nilai recall sebesar 31,6% menunjukkan bahwa model mampu mengidentifikasi sebagian kandidat yang termasuk dalam kategori Lolos, meskipun masih terdapat kandidat yang tidak terdeteksi. Di sisi lain, nilai precision yang relatif rendah menunjukkan bahwa masih terdapat sejumlah dokumen Tidak Lolos yang diprediksi sebagai Lolos.

Penggunaan threshold hasil optimasi dari data validasi pada tahap ini bertujuan untuk memberikan evaluasi performa model yang paling representatif sebelum dilakukan analisis lebih lanjut terhadap variasi threshold.

C. *Analisis Confusion Matrix*

Hasil confusion matrix yang telah dilakukan pada data uji ditunjukkan pada Tabel VII.

TABEL VII
CONFUSION MATRIX DATA UJI

	Prediksi Tidak	Prediksi Lolos
Aktual Tidak	119	33
Aktual Lolos	13	6

Berdasarkan Tabel VII, model berhasil mengklasifikasikan 119 dokumen Tidak Lolos dengan benar (*True Negative*) dan 6 dokumen Lolos dengan benar (*True Positive*). Namun demikian, terdapat 33 dokumen Tidak Lolos yang salah diklasifikasikan sebagai Lolos (*False Positive*) serta 13 dokumen Lolos yang tidak terdeteksi (*False Negative*).

Dalam konteks *early screening*, kesalahan tipe *False Positive* masih dapat ditoleransi karena keputusan akhir tetap dilakukan oleh pihak *Human Resource*. Namun, keberadaan False Negative perlu menjadi perhatian karena berpotensi menyebabkan kandidat potensial terlewatkan.

D. *Analisis Pengaruh Threshold*

Untuk menganalisis pengaruh nilai threshold terhadap performa model, dilakukan pengujian menggunakan beberapa nilai threshold yang ditentukan secara manual. Nilai threshold yang digunakan dalam analisis ini meliputi 0,05; 0,10; 0,15; 0,20; 0,25; 0,30; 0,40; dan 0,50. Kinerja model pada berbagai nilai threshold ditunjukkan pada Tabel VIII.

TABEL VIII
KINERJA MODEL PADA BERBAGAI NILAI THRESHOLD

Threshold	Accuracy	Precision	Recall	F1-score
0,05	0,690	0,146	0,368	0,209
0,10	0,713	0,143	0,316	0,197
0,15	0,725	0,150	0,316	0,203
0,20	0,754	0,171	0,316	0,222
0,25	0,760	0,176	0,316	0,226
0,30	0,778	0,194	0,316	0,240
0,40	0,784	0,179	0,263	0,213
0,50	0,801	0,200	0,263	0,227

Tabel Berdasarkan Tabel VIII, terlihat adanya trade-off antara precision dan recall pada berbagai nilai threshold. Threshold yang lebih rendah cenderung meningkatkan nilai recall karena model menjadi lebih sensitif dalam mengklasifikasikan kandidat sebagai Lolos. Namun, hal ini diikuti dengan penurunan precision karena meningkatnya jumlah False Positive.

Sebaliknya, threshold yang lebih tinggi cenderung meningkatkan precision karena model menjadi lebih selektif, namun menyebabkan penurunan recall sehingga lebih banyak kandidat Lolos yang tidak terdeteksi.

Pada rentang threshold 0,20 hingga 0,30, model menunjukkan keseimbangan yang relatif lebih baik antara precision dan recall, yang tercermin dari peningkatan nilai F1-score dibandingkan threshold lainnya. Hal ini menunjukkan bahwa rentang tersebut dapat dipertimbangkan sebagai kompromi antara sensitivitas dan ketepatan prediksi.

Perlu dicatat bahwa analisis ini bersifat eksploratif dan bertujuan untuk memahami perilaku model terhadap variasi threshold. Nilai threshold yang digunakan pada evaluasi utama tetap mengacu pada hasil optimasi dari data validasi sebagaimana dijelaskan pada bagian sebelumnya.

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

E. *Pembahasan*

Hasil penelitian menunjukkan bahwa model IndoBERT mampu menangkap sebagian karakteristik dokumen CV berbahasa Indonesia, namun performa klasifikasi masih dipengaruhi oleh ketidakseimbangan kelas yang signifikan. Nilai PR-AUC sebesar 0,165 mengindikasikan bahwa kemampuan model dalam membedakan kelas minoritas masih terbatas, meskipun pendekatan class weighting telah diterapkan.

Kompleksitas dokumen CV yang tidak terstruktur, variasi format, serta penggunaan istilah campuran bahasa Indonesia dan Inggris berpotensi mengurangi konsistensi representasi semantik antar dokumen. Selain itu, proporsi kelas Lolos yang relatif kecil menyebabkan distribusi probabilitas prediksi condong ke kelas mayoritas, sehingga berdampak pada nilai precision dan recall yang belum optimal.

Analisis terhadap variasi *threshold* menunjukkan adanya trade-off yang konsisten antara precision dan recall. Threshold yang lebih rendah meningkatkan sensitivitas terhadap kelas minoritas, sedangkan threshold yang lebih tinggi meningkatkan presisi namun mengurangi cakupan deteksi kandidat Lolos. Dalam konteks early screening, pemilihan threshold dapat disesuaikan dengan kebijakan organisasi, apakah lebih memprioritaskan cakupan kandidat potensial (*recall*) atau mengurangi beban verifikasi manual (*precision*).

Secara keseluruhan, meskipun performa numerik belum menunjukkan nilai yang tinggi, pendekatan berbasis IndoBERT memberikan fondasi yang menjanjikan untuk pengembangan sistem penyaringan CV otomatis yang lebih adaptif terhadap karakteristik bahasa Indonesia.

V. KESIMPULAN

Penelitian ini mengevaluasi penerapan model IndoBERT untuk klasifikasi dokumen Curriculum Vitae (CV) berbahasa Indonesia pada tahap seleksi administrasi. Eksperimen dilakukan pada dataset rekrutmen nyata yang terdiri dari 1140 dokumen dengan distribusi kelas yang tidak seimbang. Hasil evaluasi menunjukkan bahwa model mampu melakukan klasifikasi dokumen dengan performa yang dipengaruhi oleh karakteristik data yang tidak terstruktur serta ketidakseimbangan kelas yang signifikan.

Hasil evaluasi pada data uji menunjukkan bahwa model IndoBERT yang digunakan menghasilkan nilai Accuracy sebesar 0,731, Precision sebesar 0,154, Recall sebesar 0,316, F1-score sebesar 0,207, serta PR-AUC sebesar 0,165. Nilai tersebut menunjukkan bahwa model memiliki kemampuan dalam melakukan klasifikasi dokumen CV, meskipun performa masih dipengaruhi oleh ketidakseimbangan kelas yang signifikan.

Analisis terhadap variasi nilai threshold menunjukkan adanya *trade-off* yang konsisten antara precision dan recall. Penentuan *threshold* berdasarkan data validasi memberikan fleksibilitas dalam menyesuaikan kebutuhan sistem, khususnya untuk skenario early screening yang menekankan prioritisasi kandidat. Temuan ini menunjukkan bahwa penyesuaian threshold merupakan komponen penting dalam penerapan model klasifikasi pada konteks rekrutmen nyata.

Secara keseluruhan, penelitian ini memberikan kontribusi berupa evaluasi empiris penggunaan IndoBERT pada klasifikasi CV berbahasa Indonesia dengan analisis ketidakseimbangan data dan strategi penentuan threshold. Penelitian selanjutnya dapat difokuskan pada eksplorasi teknik penanganan ketidakseimbangan data yang lebih lanjut, seperti oversampling atau focal loss, serta pengembangan pendekatan berbasis fitur domain rekrutmen untuk meningkatkan kemampuan diskriminatif model

REFERENSI

[1] A. Jain and P. Sharma, “Automated Resume Screening Using Machine Learning Techniques,” *Int. J. Eng. Res. Technol.*, vol. 11, no. 5, pp. 1–6, 2022.
[2] R. Nugroho and S. Dewi, “Digital Recruitment Challenges in Indonesia,” *J. Manaj. Teknol.*, vol. 21, no. 2, pp. 115–123, 2022.
[3] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge University Press, 2008.
[4] T. Joachims, “Text Categorization with Support Vector Machines: Learning with Many Relevant Features,” in *Proceedings of ECML*, 1998, pp. 137–142.
[5] A. Vaswani and others, “Attention Is All You Need,” in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.
[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.
[7] M. Malherbe and N. Rule, “Resume Classification Using Contextual Language Models,” in *IEEE International Conference on Big Data*, 2020, pp. 1–6.
[8] B. Wilie and others, “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” in *Findings of EMNLP*, 2020, pp. 843–857.
[9] F. Kurniawan and R. Sarno, “Deep Learning for Indonesian Text Classification: A Comparative Study,” *TELKOMNIKA*, vol. 18, no. 6, pp. 3109–3116, 2020.
[10] R. Parlika, A. Pratama, and A. Kurniawan, “The Online Test Application Uses Telegram Bots Version 1.0,” *J. Phys. Conf. Ser.*, vol. 1569, no. 2, 2020.
[11] J. Davis and M. Goadrich, “The Relationship Between Precision-Recall and ROC Curves,” in *Proceedings of ICML*, 2006, pp. 233–240.
[12] P. Cappelli, “AI in Human Resources: Opportunities and Risks,” *Calif. Manage. Rev.*, vol. 61, no. 4, pp. 15–42, 2019.
[13] H. Zhang, Y. Song, and Q. Li, “Automated Resume Screening with Pretrained Language Models,” *IEEE Access*, vol. 9, pp. 123456–123468, 2021.
[14] M. Kumar and S. Singh, “Deep Learning-Based Resume Classification System,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 5, pp. 245–252, 2022.
[15] X. Sun, Y. Cheng, and Z. Gan, “How to Fine-Tune BERT for Text Classification?,” in *Proceedings of CCL*, 2019, pp. 194–206.
[16] Y. Liu and others, “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” *arXiv Prepr. arXiv1907.11692*, 2019.
[17] S. Buda, A. Maki, and M. A. Mazurowski, “A Systematic Study of the Class Imbalance Problem in Deep Learning,” *Neural Networks*, vol. 106, pp. 249–259, 2018.
[18] C. Xie, F. Zheng, and L. Zhang, “Imbalanced Text Classification Using BERT with Cost-Sensitive Learning,” *IEEE Access*, vol. 8, pp. 12345–12356, 2020.
[19] T.-Y. Lin and others, “Focal Loss for Dense Object Detection,” in *Proceedings of ICCV*, 2017, pp. 2980–2988.
[20] I. Koto, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: Benchmarking Indonesian Language Models,” in *Proceedings of COLING*,

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

2020.
[21] Z. C. Lipton, C. Elkan, and B. Naryanaswamy, "Optimal Thresholding of Classifiers to Maximize F1 Measure," in *Proceedings of ECML PKDD*, 2014.