

Klasifikasi *Multi-Label* Dan Ekstraksi Entitas Pada Ulasan Aplikasi Blu by BCA Digital Menggunakan *IndoBERT*

Multi-Label Classification and Entity Extraction on Blu by BCA Digital Application Reviews Using IndoBERT

Bhagyas Satrya Dewa*¹, Eka Dyar Wahyuni², Nur Cahyo Wibowo³
^{1,2,3}Sistem Informasi, Universitas Pembangunan Nasional "Veteran" Jawa Timur
E-mail: 122082010164@student.upnjatim.ac.id

Abstrak

Pertumbuhan layanan perbankan digital di Indonesia mendorong kebutuhan untuk memahami faktor-faktor yang memengaruhi keberlanjutan penggunaan aplikasi oleh pengguna. Namun, penelitian sebelumnya masih terbatas pada pendekatan klasifikasi *single-label* dan analisis sentimen umum, sehingga belum mampu menangkap kompleksitas informasi dalam ulasan pengguna berbahasa Indonesia secara terstruktur. Penelitian ini bertujuan melakukan klasifikasi *multi-label* berdasarkan empat faktor Expectation-Confirmation Model (ECM), yaitu Confirmation, Perceived Usefulness, E-satisfaction, dan Perceived Security, serta ekstraksi enam entitas Named Entity Recognition (NER) pada ulasan aplikasi Blu by BCA Digital menggunakan IndoBERT. Dataset diperoleh dari Google Play Store dan Apple App Store periode Januari–Desember 2025, menghasilkan 3.389 ulasan berbahasa Indonesia setelah proses filtering. Penelitian menerapkan pendekatan *single-task* dengan oversampling dan Focal Loss untuk klasifikasi *multi-label*, serta token augmentation dan Conditional Random Field (CRF) untuk NER. Validasi anotasi dilakukan menggunakan Krippendorff's Alpha yang menghasilkan nilai rata-rata 0,856 untuk label intent dan 0,919 untuk entitas NER. Hasil pengujian menunjukkan model klasifikasi terbaik mencapai F1-Score sebesar 0,798 dan Hamming Loss sebesar 0,131, sedangkan model NER terbaik mencapai F1-Score sebesar 0,812. Penelitian ini menunjukkan bahwa IndoBERT efektif digunakan untuk menganalisis ulasan pengguna aplikasi perbankan digital dalam mengidentifikasi faktor ECM dan mengekstraksi entitas domain secara bersamaan, sehingga berpotensi mendukung pengembang dalam memahami kebutuhan pengguna secara otomatis.

Kata kunci: IndoBERT, *multi-label*, NER, ECM, bank digital

Abstract

The growth of digital banking services in Indonesia has heightened the need to understand factors influencing users' application continuance intention. However, prior studies remain limited to *single-label* classification and general sentiment analysis, lacking the ability to capture the complexity of information in Indonesian-language user reviews in a structured manner. This study aims to perform *multi-label* classification based on four Expectation-Confirmation Model (ECM) factors—Confirmation, Perceived Usefulness, E-satisfaction, and Perceived Security—and to extract six Named Entity Recognition (NER) entities from Blu by BCA Digital application reviews using IndoBERT. The dataset was collected from Google Play Store and Apple App Store covering January to December 2025, yielding 3,389 Indonesian-language reviews after filtering. The study employs a *single-task* approach, applying oversampling and Focal Loss for *multi-label* classification, and token augmentation with Conditional Random Field (CRF) for NER. Annotation validation using Krippendorff's Alpha yielded average values of 0.856 for intent labels and 0.919 for NER entities. Results show that the best classification model achieved an F1-Score of 0.798 with a Hamming Loss of 0.131,

while the best NER model achieved an F1-Score of 0.812. This study demonstrates that IndoBERT is effective for analyzing digital banking application reviews in identifying ECM factors and extracting domain-specific entities, thereby offering potential support for developers in automatically understanding user needs.

Keywords: IndoBERT, multi-label, NER, ECM, digital banking

1. PENDAHULUAN

Perkembangan teknologi informasi mendorong transformasi digital pada sektor perbankan melalui hadirnya layanan bank digital yang menawarkan kemudahan transaksi secara cepat dan fleksibel. Pertumbuhan penggunaan layanan perbankan digital di Indonesia juga meningkat secara signifikan. Bank Indonesia memprediksi transaksi perbankan digital pada tahun 2025 akan terus tumbuh seiring meningkatnya penggunaan internet dan smartphone [1]. Salah satu layanan bank digital di Indonesia adalah Blu by BCA Digital yang didukung oleh reputasi kuat Bank Central Asia (BCA). Berdasarkan survei Populix tahun 2024, Blu by BCA Digital memiliki tingkat brand awareness sebesar 85%, namun hanya sekitar 26% pengguna yang masih aktif menggunakan aplikasi tersebut dalam satu bulan terakhir [2]. Kondisi ini menunjukkan adanya kesenjangan antara tingkat kesadaran merek dengan penggunaan berkelanjutan aplikasi. Selain itu, ulasan pengguna pada Google Play Store dan App Store masih menunjukkan berbagai keluhan terkait kendala login, performa aplikasi yang lambat, maupun hambatan transaksi yang dapat memengaruhi pengalaman pengguna.

Fenomena keberlanjutan penggunaan aplikasi dapat dijelaskan menggunakan *Expectation-Confirmation Model* (ECM) yang dikembangkan oleh Bhattacherjee [3]. Model ini menjelaskan bahwa *IS Continuance Intention* dipengaruhi oleh beberapa faktor, seperti *confirmation*, *perceived usefulness*, dan *e-satisfaction*. Dalam konteks layanan perbankan digital, penelitian ini juga mempertimbangkan faktor *perceived security* karena keamanan menjadi aspek penting dalam penggunaan layanan keuangan digital [4]. Faktor-faktor tersebut dapat tercermin dalam ulasan pengguna aplikasi yang berisi pengalaman, kepuasan, maupun keluhan terhadap layanan yang digunakan. Namun, sebagian besar penelitian sebelumnya masih menggunakan pendekatan survei atau kuesioner sehingga belum mampu menggambarkan opini pengguna secara lebih mendalam melalui data teks ulasan pengguna.

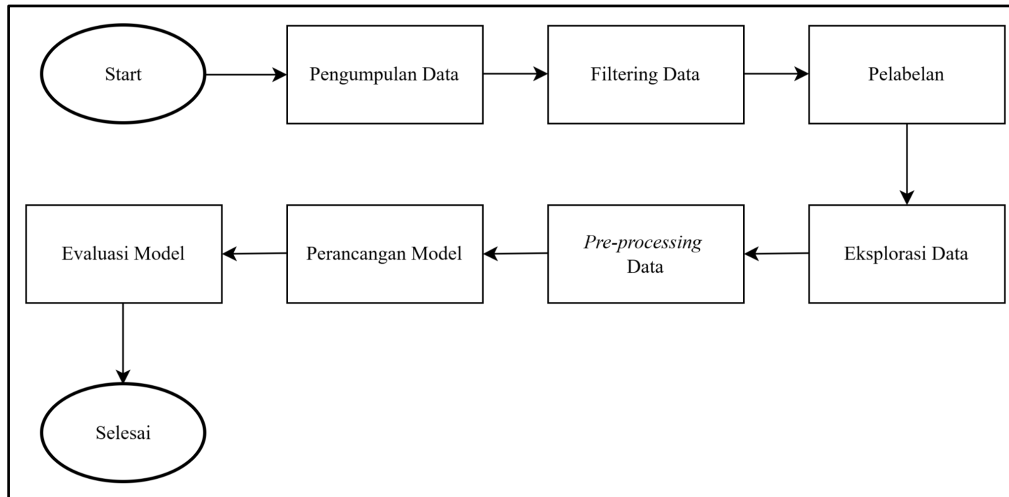
Ulasan pengguna umumnya mengandung lebih dari satu informasi dalam satu kalimat, misalnya pengguna merasa puas terhadap fitur tertentu tetapi mengeluhkan performa aplikasi secara bersamaan. Kondisi tersebut menyebabkan pendekatan klasifikasi *single-label* kurang mampu merepresentasikan kompleksitas informasi pada ulasan pengguna [5]. Oleh karena itu, penelitian ini menerapkan pendekatan klasifikasi *multi-label* untuk mengklasifikasikan faktor ECM pada ulasan aplikasi Blu by BCA Digital. Pendekatan multi-label relevan digunakan karena satu ulasan dapat memuat lebih dari satu aspek atau kategori secara bersamaan, sebagaimana ditunjukkan dalam penelitian klasifikasi *multi-label* pada ulasan pelanggan berbahasa Indonesia menggunakan IndoBERT [6].

Selain itu, penelitian ini juga menerapkan *Named Entity Recognition* (NER) untuk mengekstraksi entitas penting seperti fitur layanan, performa aplikasi, maupun kendala teknis yang dibahas pengguna [7]. NER diperlukan karena ulasan

pengguna berbentuk teks tidak terstruktur, sehingga informasi spesifik yang menjadi objek keluhan atau apresiasi perlu dikenali secara otomatis. Pada teks berbahasa Indonesia, tugas NER memiliki tantangan tersendiri karena variasi bentuk kata, gaya bahasa informal, serta struktur kalimat yang tidak selalu baku. Penelitian Kurniawan dan Louvan [8] menunjukkan bahwa NER pada teks percakapan berbahasa Indonesia menghadapi tantangan berupa variasi kata dan tingginya *out-of-vocabulary*, sehingga pendekatan berbasis *neural sequence labeling* diperlukan untuk meningkatkan kemampuan pengenalan entitas.

Untuk mendukung proses tersebut, penelitian ini menggunakan *IndoBERT* sebagai model berbasis *Transformer* yang memiliki kemampuan memahami konteks bahasa Indonesia secara *bidirectional* dan telah menunjukkan performa yang baik pada tugas klasifikasi teks maupun NER [9][10]. Namun, penelitian sebelumnya pada ulasan aplikasi perbankan digital umumnya masih berfokus pada analisis sentimen, klasifikasi tunggal, atau ekstraksi informasi umum dari ulasan aplikasi. Penelitian yang secara khusus memetakan ulasan bank digital berbahasa Indonesia ke dalam beberapa faktor ECM melalui klasifikasi *multi-label* serta mengekstraksi entitas spesifik melalui NER masih terbatas. Oleh karena itu, nilai lebih penelitian ini terletak pada penerapan klasifikasi *multi-label* dan ekstraksi entitas pada ulasan aplikasi perbankan digital berbahasa Indonesia menggunakan *IndoBERT*, sehingga mampu menghasilkan analisis ulasan yang lebih rinci dan terstruktur.

2. METODOLOGI PENELITIAN



Gambar 1. Alur Penelitian

Untuk memudahkan pemahaman, rangkaian penelitian ini divisualisasikan melalui diagram alir pada Gambar 1. Bagan ini berfungsi sebagai peta jalan yang merangkum seluruh proses dari awal hingga akhir. Detail mengenai apa yang terjadi di setiap fase, termasuk langkah-langkah yang diambil, akan dikupas secara mendalam pada bagian di bawah ini.

2.1. Pengumpulan Data

Data penelitian diperoleh dari ulasan pengguna aplikasi Blu by BCA Digital pada platform Google Play Store dan Apple App Store periode Januari hingga

Desember 2025. Proses pengambilan data dilakukan menggunakan teknik *web scraping* dengan bantuan pustaka *google-play-scraper* untuk Google Play Store dan *app_store_scraper* pada Apple App Store. Pengambilan data difokuskan pada ulasan berbahasa Indonesia dengan parameter *lang='id'* dan *country='id'*. Dari proses tersebut diperoleh sebanyak 3.905 ulasan dari Google Play Store dan 463 ulasan dari Apple App Store sehingga menghasilkan total 4.368 ulasan pengguna. Dataset yang dikumpulkan mencakup informasi berupa isi ulasan, rating, tanggal ulasan, serta sumber platform. Selanjutnya, seluruh data digabungkan dan digunakan sebagai dataset utama untuk proses berikutnya.

2.2. Filtering Data

Tahap *filtering* dilakukan untuk memastikan kualitas dataset yang digunakan pada penelitian. Proses ini diawali dengan penghapusan data kosong, ulasan yang hanya berisi spasi, serta ulasan duplikat untuk menghindari redundansi data. Selanjutnya, dilakukan penyaringan panjang ulasan minimal tiga kata agar setiap data memiliki konteks informasi yang memadai untuk proses klasifikasi. Penelitian ini juga menerapkan deteksi bahasa menggunakan pustaka *langdetect* untuk memastikan ulasan yang digunakan hanya berbahasa Indonesia sehingga sesuai dengan karakteristik model IndoBERT. Selain itu, dilakukan standarisasi atribut tanggal dan penyesuaian format data dari kedua platform agar dataset memiliki struktur yang seragam. Berdasarkan proses *filtering* tersebut, jumlah data yang semula sebanyak 4.368 ulasan berkurang menjadi 3.389 ulasan yang digunakan pada tahap penelitian selanjutnya.

2.3. Pelabelan

Tahap pelabelan dilakukan secara manual terhadap 3.389 ulasan pengguna menggunakan platform *Label Studio*. Proses anotasi melibatkan tujuh *annotator* untuk menjaga kualitas dan konsistensi pelabelan data. Setiap ulasan dapat memiliki lebih dari satu label (*multi-label*) yang terdiri dari *Confirmation*, *Perceived Usefulness*, *E-satisfaction*, dan *Perceived Security*. Selain pelabelan klasifikasi intent, penelitian ini juga menerapkan pelabelan *Named Entity Recognition* (NER) untuk mengidentifikasi entitas penting pada ulasan pengguna, yaitu *ORGANIZATION*, *FEATURE*, *SERVICE*, *PERFORMANCE*, *COST*, dan *SECURITY* menggunakan format BIO (*Begin*, *Inside*, *Outside*). Selanjutnya, validasi hasil anotasi dilakukan menggunakan metode *Krippendorff's Alpha* untuk mengukur tingkat kesepakatan antar *annotator*. *Krippendorff's Alpha* merupakan metode pengukuran reliabilitas anotasi yang fleksibel karena mampu digunakan pada berbagai jenis data kategorikal dan mendukung jumlah *annotator* lebih dari dua orang [11]. Hasil perhitungan *Krippendorff's Alpha* pada masing-masing label ditunjukkan pada Tabel 1 dan Tabel 2, yang menunjukkan tingkat reliabilitas anotasi berada pada kategori baik sehingga dataset dinilai layak digunakan pada tahap pemodelan.

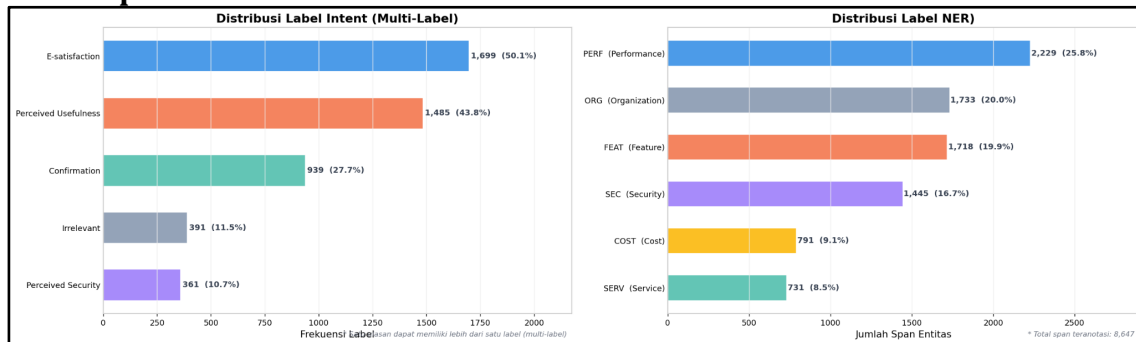
Tabel 1. Perhitungan *Krippendorff*
Alpha Intent

Label	Nilai
<i>Confirmation</i>	0.8701
<i>Perceived Usefulness</i>	0.8213
<i>E-satisfaction</i>	0.8558
<i>Perceived Security</i>	0.8770

Tabel 2. Perhitungan *Krippendorff*
Alpha NER

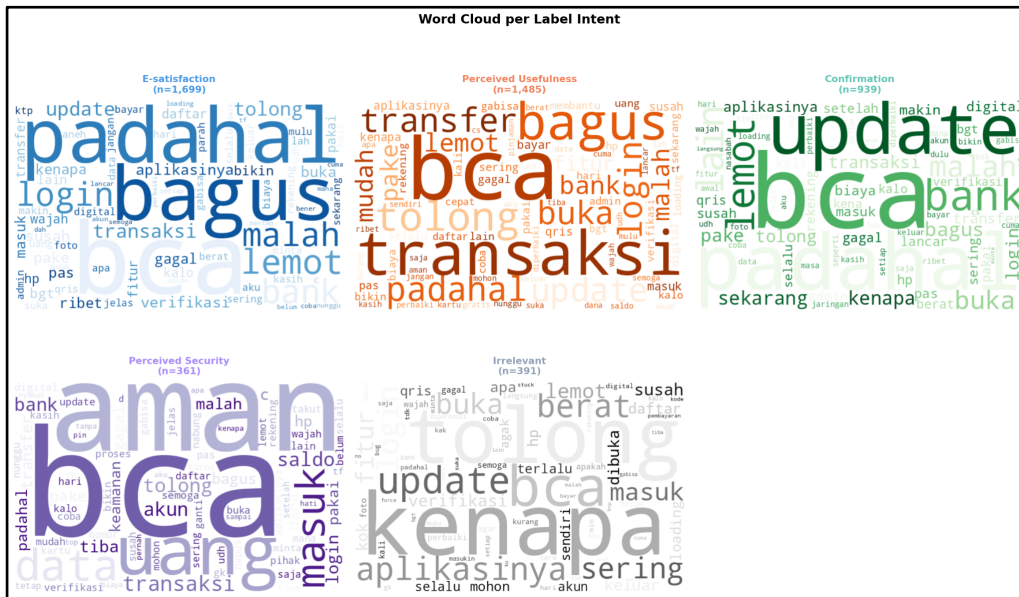
Label	Nilai
ORG	0.9614
FEAT	0.8971
SERV	0.8538
PERF	0.9202
COST	0.9562
SEC	0.9245

2.4. Eksplorasi Data

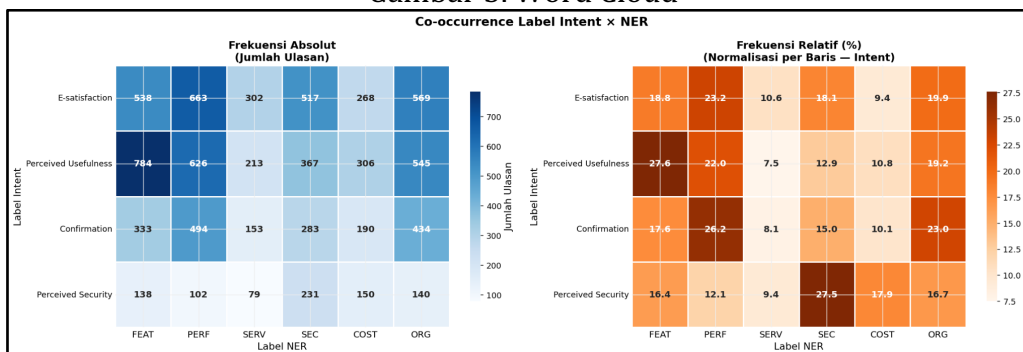


Gambar 2. Distribusi Label

Setelah data dilabeli, tahap *Exploratory Data Analysis* (EDA) dilakukan untuk memperoleh pemahaman mengenai karakteristik, pola, dan distribusi data sebelum proses pemodelan [12]. Berdasarkan Gambar 2, label *E-satisfaction* memiliki proporsi tertinggi sebesar 50,1%, diikuti *Perceived Usefulness* sebesar 43,8% dan *Confirmation* sebesar 27,7%, sedangkan label *Perceived Security* memiliki proporsi paling rendah sebesar 10,7%. Distribusi label intent bersifat *multi-label*, sehingga satu ulasan dapat memiliki lebih dari satu label secara bersamaan. Sementara itu, distribusi label NER menunjukkan bahwa entitas *PERF* (*Performance*) menjadi entitas paling dominan dengan proporsi sebesar 25,8%, diikuti *ORG* dan *FEAT*. Dominasi entitas tersebut menunjukkan bahwa sebagian besar ulasan pengguna membahas performa aplikasi, fitur layanan, dan penyebutan institusi atau aplikasi secara langsung. Selain itu, visualisasi *word cloud* pada Gambar 3 digunakan untuk melihat kata-kata yang paling sering muncul pada setiap label intent. Kata seperti “bagus”, “transfer”, “update”, “bug”, dan “lemot” terlihat dominan pada beberapa label, terutama *E-satisfaction* dan *Confirmation*, yang menggambarkan pengalaman pengguna terhadap performa aplikasi.



Gambar 3. Word Cloud



Gambar 4. Co-occurrence Intent dan NER

Analisis eksplorasi juga dilakukan menggunakan *co-occurrence matrix* antara label intent dan entitas NER seperti ditunjukkan pada Gambar 4. Visualisasi tersebut digunakan untuk melihat keterkaitan antara faktor *IS Continuance Intention* dan entitas yang dibahas pengguna dalam ulasan. Berdasarkan hasil analisis, label *E-satisfaction* dan *Perceived Usefulness* memiliki frekuensi kemunculan yang tinggi pada entitas *PERF*, *FEAT*, dan *ORG*. Hal ini menunjukkan bahwa kepuasan dan persepsi manfaat pengguna banyak dipengaruhi oleh performa aplikasi serta fitur layanan yang digunakan. Sementara itu, pada label *Perceived Security*, entitas *SEC* memiliki proporsi relatif paling tinggi dibandingkan entitas lainnya, yang menunjukkan bahwa pengguna yang membahas keamanan cenderung menyoroti aspek verifikasi, perlindungan akun, dan keamanan transaksi. Hasil eksplorasi ini digunakan untuk memahami pola hubungan antar label sebelum tahap pemodelan dilakukan.

2.5. Pre-processing

Tahap *pre-processing* dilakukan untuk menyiapkan teks ulasan agar lebih bersih dan terstruktur sebelum digunakan dalam proses pemodelan. Proses ini bertujuan untuk mengurangi *noise* serta menyeragamkan bentuk teks sehingga dapat diproses secara lebih optimal oleh model IndoBERT. Tahapan *pre-processing* yang diterapkan meliputi *data cleaning*, *case folding*, *normalization*, dan

tokenization. Pada tahap *data cleaning*, dilakukan penghapusan ulasan yang hanya memiliki label *Irrelevant* tanpa entitas NER, penghapusan elemen tidak relevan seperti URL, simbol, dan karakter berulang, serta konversi emoji menjadi representasi teks untuk mempertahankan makna emosional dalam ulasan. Selain itu, dilakukan penyesuaian posisi span NER secara otomatis agar anotasi entitas tetap sesuai dengan perubahan teks yang terjadi.

Tahap berikutnya adalah *case folding* dengan mengubah seluruh teks menjadi huruf kecil untuk menyeragamkan representasi kata. Selanjutnya, normalisasi dilakukan menggunakan kamus normalisasi bahasa informal dan istilah perbankan digital untuk menyederhanakan variasi penulisan kata agar lebih konsisten. Setelah proses normalisasi selesai, dilakukan *tokenization* menggunakan *tokenizer IndoBERT* berbasis *WordPiece* untuk mengubah teks menjadi token yang sesuai dengan format masukan model. Proses tokenisasi juga diikuti penyesuaian label NER agar tetap konsisten dengan hasil token yang dihasilkan. Contoh hasil sebelum dan sesudah tahap *pre-processing* ditunjukkan pada Tabel 3.

Ulasan	Sesudah <i>Pre-processing</i>
Sudah 3 hari lebih tidak bisa masuk cape banget tolong yah kaka 🙏	["sudah", "3", "hari", "lebih", "tidak", "bisa", "masuk", "capek", "banget", "tolong", "ya", "kakak", "emoji_doa"]
sangat membantu semoga sukses selalu 🙏🙏	["sangat", "membantu", "semoga", "sukses", "selalu", "emoji_jempol", "emoji_doa"]
👍👍👍👍👍👍👍 perfect pokoknya top deh dr bank lainnya	["emoji_jempol", "emoji_oke", "perfect", "pokoknya", "top", "deh", "dari", "bank", "lainnya"]

Tabel 3. Hasil *Pre-processing*

2.6. Perancangan Model

Penelitian ini menggunakan pendekatan *single-task learning* untuk tugas klasifikasi intent *multi-label* dan *Named Entity Recognition* (NER) secara terpisah menggunakan model IndoBERT. Pada tugas klasifikasi intent, keluaran representasi teks diteruskan ke lapisan klasifikasi dengan fungsi aktivasi *sigmoid* untuk memprediksi label *Confirmation*, *Perceived Usefulness*, *E-satisfaction*, dan *Perceived Security*. Untuk mengatasi ketidakseimbangan distribusi label, penelitian ini menerapkan teknik *oversampling* pada data latih sehingga distribusi label menjadi lebih seimbang. Pendekatan *oversampling* diketahui efektif dalam meningkatkan kemampuan model dalam mengenali label minoritas pada tugas klasifikasi *multi-label* [13]. Sementara itu, pada tugas NER, setiap token hasil tokenisasi diprediksi menggunakan skema BIO (*Begin*, *Inside*, *Outside*) untuk mengidentifikasi entitas *ORG*, *FEAT*, *SERV*, *PERF*, *COST*, dan *SEC*. Pada model NER diterapkan teknik *token augmentation* untuk meningkatkan variasi data pelatihan serta lapisan *Conditional Random Field* (CRF) untuk meningkatkan konsistensi prediksi antar token. Penggunaan CRF pada tugas NER juga telah terbukti mampu meningkatkan performa prediksi label token dengan mempertimbangkan dependensi antar label dalam satu urutan teks [14].

2.7. Evaluasi Model

Tahap evaluasi dilakukan untuk mengukur performa model klasifikasi intent *multi-label* dan *Named Entity Recognition* (NER). Pada tugas klasifikasi intent, evaluasi menggunakan metrik *precision*, *recall*, *F1-score*, dan *Hamming Loss* untuk mengukur kemampuan model dalam memprediksi lebih dari satu label pada setiap ulasan. Sementara itu, pada tugas NER digunakan metrik *precision*, *recall*, dan *F1-score* untuk mengukur ketepatan identifikasi entitas pada teks ulasan. Pengujian dilakukan pada data uji menggunakan beberapa skenario pembagian data untuk melihat stabilitas performa model.

3. HASIL DAN PEMBAHASAN

Bagian ini membahas hasil pengujian model klasifikasi intent *multi-label* dan *Named Entity Recognition* (NER) pada ulasan aplikasi Blu by BCA Digital. Evaluasi dilakukan untuk mengukur performa model dalam mengidentifikasi faktor *Expectation-Confirmation Model* (ECM) serta mengekstraksi entitas penting dari ulasan pengguna. Hasil evaluasi kemudian dianalisis untuk melihat kemampuan model dalam memahami konteks ulasan pengguna pada layanan perbankan digital.

3.1. Hasil Klasifikasi Intent *Multi-Label*

Tabel 3. Hasil Klasifikasi Intent *Multi-Label*

Model	Skenario	Data	Precision	Recall	F1-Score	Hamming Loss
M1	<i>BCE Loss</i>	80:20	0.791	0.793	0.791	0.134
		70:30	0.760	0.770	0.762	0.158
M2	<i>BCE Loss + Oversample</i>	80:20	0.748	0.781	0.764	0.135
		70:30	0.749	0.786	0.767	0.158
M3	<i>Focal Loss</i>	80:20	0.736	0.809	0.795	0.158
		70:30	0.732	0.796	0.761	0.165
M4	<i>Focal Loss + Oversample</i>	80:20	0.794	0.801	0.798	0.131
		70:30	0.754	0.780	0.763	0.162

Hasil pengujian model klasifikasi intent *multi-label* ditunjukkan pada Tabel 3. Pengujian dilakukan menggunakan empat skenario model dengan variasi fungsi *loss* dan penerapan *oversampling* pada dua skenario pembagian data, yaitu 80:20 dan 70:30. Berdasarkan hasil pengujian, model M4 dengan kombinasi *Focal Loss* dan *Oversampling* pada pembagian data 80:20 memperoleh performa terbaik dengan nilai *Precision* sebesar 0,794, *Recall* sebesar 0,801, *F1-Score* sebesar 0,798, dan *Hamming Loss* sebesar 0,131. Hasil tersebut menunjukkan bahwa kombinasi *Focal Loss* dan *Oversampling* mampu meningkatkan kemampuan model dalam menangani distribusi label yang tidak seimbang pada tugas klasifikasi *multi-label*.

Penerapan *Focal Loss* membantu model memberikan perhatian lebih besar terhadap label minoritas, sedangkan teknik *oversampling* membantu

menyeimbangkan distribusi data latih sehingga model lebih mampu mengenali label dengan jumlah data yang lebih sedikit, seperti *Perceived Security*. Selain itu, hasil pengujian juga menunjukkan bahwa skenario pembagian data 80:20 secara umum menghasilkan performa yang lebih baik dibandingkan skenario 70:30. Hal ini menunjukkan bahwa jumlah data latih yang lebih besar memberikan representasi pola yang lebih baik bagi model dalam memahami konteks ulasan pengguna. Secara keseluruhan, hasil pengujian menunjukkan bahwa model IndoBERT mampu melakukan klasifikasi faktor *Expectation-Confirmation Model* (ECM) pada ulasan aplikasi Blu by BCA Digital dengan performa yang baik.

3.2. Hasil Named Entity Recognition

Tabel 4. Hasil NER

Model	Skenario	Data	Precision	Recall	F1-Score
M5	NER <i>Baseline</i>	80:20	0.768	0.811	0.788
		70:30	0.763	0.816	0.787
M6	NER + <i>CRF</i>	80:20	0.806	0.811	0.807
		70:30	0.769	0.813	0.790
M7	NER + <i>Augmentation</i>	80:20	0.789	0.820	0.804
		70:30	0.759	0.819	0.787
M8	NER + <i>Augmentation</i> + <i>CRF</i>	80:20	0.802	0.823	0.812
		70:30	0.757	0.809	0.781

Hasil pengujian model *Named Entity Recognition* (NER) ditunjukkan pada Tabel 4. Pengujian dilakukan menggunakan empat skenario model dengan variasi penggunaan *Conditional Random Field* (CRF) dan teknik *token augmentation* pada dua skenario pembagian data, yaitu 80:20 dan 70:30. Berdasarkan hasil pengujian, model M8 dengan kombinasi *token augmentation* dan *CRF* pada pembagian data 80:20 memperoleh performa terbaik dengan nilai *Precision* sebesar 0,802, *Recall* sebesar 0,823, dan *F1-Score* sebesar 0,812. Hasil tersebut menunjukkan bahwa kombinasi *token augmentation* dan CRF mampu meningkatkan kemampuan model dalam mengenali serta mempertahankan konsistensi prediksi entitas pada teks ulasan pengguna.

Penerapan *token augmentation* membantu meningkatkan variasi data pelatihan sehingga model menjadi lebih adaptif terhadap variasi penulisan kata pada ulasan pengguna, seperti singkatan, bahasa informal, maupun kesalahan penulisan. Sementara itu, penggunaan lapisan CRF membantu model memahami dependensi antar label token sehingga prediksi entitas menjadi lebih konsisten pada skema BIO (*Begin, Inside, Outside*). Selain itu, hasil pengujian menunjukkan bahwa skenario pembagian data 80:20 secara umum menghasilkan performa yang lebih baik dibandingkan pembagian data 70:30. Secara keseluruhan, hasil pengujian menunjukkan bahwa model IndoBERT mampu melakukan ekstraksi entitas pada

ulasan aplikasi Blu by BCA Digital dengan performa yang baik, terutama pada entitas yang sering muncul seperti *PERF* dan *FEAT*.

KESIMPULAN

Penelitian ini berhasil menerapkan klasifikasi intent *multi-label* dan *Named Entity Recognition* (NER) pada ulasan aplikasi Blu by BCA Digital menggunakan IndoBERT untuk mengidentifikasi faktor-faktor pembentuk *IS Continuance Intention* berdasarkan *Expectation-Confirmation Model* (ECM). Hasil pengujian menunjukkan bahwa model terbaik pada tugas klasifikasi intent diperoleh oleh skenario *Focal Loss* dan *Oversampling* dengan nilai *F1-Score* sebesar 0,798 dan *Hamming Loss* sebesar 0,131. Sementara itu, pada tugas NER, model terbaik diperoleh dari kombinasi *token augmentation* dan *Conditional Random Field* (CRF) dengan nilai *F1-Score* sebesar 0,812. Hasil tersebut menunjukkan bahwa pendekatan yang digunakan mampu mengidentifikasi faktor ECM dan mengekstraksi entitas penting pada ulasan pengguna dengan performa yang baik.

Selain menghasilkan performa klasifikasi yang baik, penelitian ini juga menunjukkan bahwa ulasan pengguna dapat dimanfaatkan untuk memahami hubungan antara kepuasan pengguna, persepsi manfaat, keamanan, serta aspek layanan yang dibahas dalam aplikasi perbankan digital. Pendekatan ini berpotensi membantu pengembang aplikasi dalam memahami kebutuhan dan keluhan pengguna secara lebih terstruktur melalui analisis otomatis berbasis NLP. Penelitian selanjutnya dapat mengembangkan pendekatan yang lebih kompleks, seperti *multi-task learning*, penggunaan model *large language model* (LLM), maupun integrasi analisis sentimen untuk memperoleh pemahaman yang lebih mendalam terhadap perilaku pengguna layanan perbankan digital.

DAFTAR PUSTAKA

- [1]. M. Puspadini, "Transaksi Bank Digital Tumbuh 40,1%, Didominasi Gen Z & Milenial," *CNBC Indonesia*, Mar. 14, 2025. [Online]. Available: <https://www.cnbcindonesia.com/market/20250314201903-17-618812/transaksi-bank-digital-tumbuh-401-didominasi-gen-z-milenial>. [Accessed: May 19, 2026].
- [2]. Populix, "Studi Analisis Ekosistem dan Persepsi terhadap Bank Digital di Indonesia," 2024. [Online]. Available: <https://info.populix.co/data-hub/reports/digitalbanking2024>. [Accessed: May 19, 2026].
- [3]. A. Bhattacharjee, "Understanding Information Systems Continuance: An Expectation-Confirmation Model," *MIS Quarterly*, vol. 25, no. 3, pp. 351–370, 2001.
- [4]. M. Khoirul Umam and D. Puspawati, "Continuance Use Intention in the use of E-wallets by using the Expectation Confirmation Model through E-Satisfaction," *Dinasti Int. J. Econ. Financ. Account.*, vol. 5, no. 5, pp. 4815–4827, 2024, doi: 10.38035/dijefa.v5i5.3523.
- [5]. F. K. Ihtada, R. Alfianita, and O. Q. Aziz, "Aspect-based Multilabel Classification of E-commerce Reviews Using Fine-tuned IndoBERT," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 10, no. 1, 2025, doi: 10.22219/kinetik.v10i1.2088.

- [6]. N. K. Nissa and E. Yulianti, "Multi-label text classification of Indonesian customer reviews using bidirectional encoder representations from transformers language model," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 5, pp. 5641–5652, 2023, doi: 10.11591/ijece.v13i5.pp5641-5652.
- [7]. R. Massenon et al., "Mobile app review analysis for crowdsourcing of software requirements : a mapping study of automated and semi-automated tools," pp. 1–60, 2024, doi: 10.7717/peerj-cs.2401.
- [8]. K. Kurniawan and S. Louvan, "Empirical Evaluation of Character-Based Model on Neural Named-Entity Recognition in Indonesian Conversational Texts," in *Proceedings of the 2018 International Conference on Asian Language Processing (IALP)*, 2018, pp. 1–6, doi: 10.1109/IALP.2018.8629151.
- [9]. A. G. Rais Kumar, Y. Sukmono, and A. E. Burhandenny, "Comparison Of Support Vector Machine And Indobert In Non-Functional Requirement Classification Of Application User Reviews", *J. Tek. Inform. (JUTIF)*, vol. 5, no. 4, pp. 1035–1042, Jul. 2024.
- [10]. M. T. Manurung, I Gusti Ngurah Lanang Wijayakusuma, and I Putu Winada Gautama, "Named Entity Recognition for Medical Records of Heart Failure Using a Pre-trained BERT Model," *J. Appl. Informatics Comput.*, vol. 9, no. 2, pp. 341–348, 2025, doi: 10.30871/jaic.v9i2.9170.
- [11]. K. H. Krippendorff, *Content Analysis: An Introduction to Its Methodology*, 3rd Editio. Thousand Oaks, California: SAGE Publications, 2018.
- [12]. A. Wibowo, *Perbankan Digital*. Semarang, Indonesia: Yayasan Prima Agus Teknik, 2022.
- [13]. F. Sulianta, *Basic Data Mining from A to Z*. Feri Sulianta, 2023. [Online]. Available: <https://books.google.co.id/books?id=JcLhEAAAQBAJ>.
- [14]. Z. A. Annisa, R. S. Perdana, and P. P. Adikara, "Kombinasi Intent Classification dan Named Entity Recognition pada Data Berbahasa Indonesia dengan Metode Dual Intent and Entity Transformer," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 5, pp. 1017–1024, 2024, doi: 10.25126/jtiik.2024117985.