

Analisis Sentimen Berbasis Aspek dengan LDA dan IndoBERT pada Ulasan Aplikasi Stockbit

Dewa Made Sutha Raditya Mahattama^{a1}, Gst Ayu Vida Matrika Giri^{a2}

¹Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹mahattama.2308561137@student.unud.ac.id
²vida@unud.ac.id

Abstract

This study aims to analyze sentiment in user reviews of the Stockbit application using a topic modeling approach combined with IndoBERT-based sentiment classification. Aspect extraction was carried out using Latent Dirichlet Allocation (LDA), and the experimental results indicate that selecting five topics ($n_{\text{components}} = 5$) provides the most optimal representation, as evidenced by a topic coherence score of 0.6191. These five topics reflect semantic structures that are highly relevant to the content of the reviews. For the sentiment classification stage, the IndoBERT-base model achieved an accuracy of 90.86%. The best performance was observed for the positive class, with an F1-score of 93.73%, while the negative class yielded an F1-score of 83.12%. This performance gap is attributed to the imbalanced data distribution, where positive sentiments are more dominant. Nevertheless, the macro-average F1-score of 88.43% demonstrates that the model is still capable of classifying both classes in a relatively balanced manner.

Keywords: *Aspect-Based Sentiment Analysis (ABSA), Sentiment Analysis, Latent Dirichlet Allocation (LDA), IndoBERT, Apps Review*

1. Pendahuluan

Pada pertengahan 2021, Stockbit sudah memiliki lebih dari 980 ribu pengguna terdaftar, dengan sekitar 70% berasal dari kelompok usia 18–34 tahun. Menurut survei OJK indeks literasi keuangan di Indonesia meningkat sebesar 49.68%, Angka tersebut menunjukkan peningkatan dari tahun ke tahun, yaitu sebesar 21.84% pada 2013, naik menjadi 29.70% pada 2016, dan mencapai 38.03% pada 2019 [1]. Dengan banyaknya pengguna serta kontribusi yang diberikan dalam literasi indeks keuangan, ulasan dan tanggapan pengguna terhadap aplikasi Stockbit yang berfokus pada bidang keuangan yaitu saham menjadi sumber data dalam mengevaluasi kualitas layanan dan kepuasan pelanggan secara keseluruhan.

Aspect-Based Sentiment Analysis (ABSA) merupakan metode dalam *Natural Language Processing (NLP)* yang bertujuan untuk menemukan pola dari sesuatu yang sedang dibahas pada sebuah teks, seperti umpan balik, postingan media sosial, dan ulasan pelanggan [2]. ABSA berbeda dengan sentimen analisis konvensional karena fokus utama dari metode ini adalah mengidentifikasi dan mengevaluasi polaritas sentimen (netral, negatif, dan positif) [3]. Berbeda dengan metode konvensional yang hanya memberikan penilaian sentimen secara umum terhadap isi teks, ABSA memberikan evaluasi yang lebih detail dengan cara menganalisis sentimen terhadap unsur-unsur spesifik yang menyusun teks tersebut. ABSA digunakan di berbagai industri, dimana pemahaman terhadap opini dan sentimen pelanggan terhadap fitur atau produk sangat penting [4]. Analisis ini berperan penting dalam berbagai sektor, seperti kesehatan, otomotif, keuangan, dan lainnya. Dimana umpan balik pelanggan dan analisis sentimen merupakan peran penting sebagai strategi dalam perkembangan bisnis [5]. Salah satu tantangan penerapan ABSA adalah ketergantungan terhadap data yang berlabel aspek, yang seringkali tidak tersedia secara lengkap, terutama dalam bahasa Indonesia.

Topic modeling adalah metode yang digunakan untuk menemukan topik-topik unik yang terdapat di dalam sebuah *dataset* [6]. Pemodelan topik dengan *Latent Dirichlet Allocation* (LDA) adalah teknik analisis konten komputasional yang dapat digunakan untuk menyelidiki struktur tematik dari suatu kumpulan teks [7]. LDA mengidentifikasi kata-kata yang terkait dengan suatu topik, yang kemudian dianggap sebagai aspek dalam analisis sentimen.

Salah satu kemajuan terbaru dalam bidang Pemrosesan Bahasa Alami (NLP) adalah pengembangan *transfer learning* dan *model contextual word embedding*, seperti *Bidirectional Encoder Representations from Transformers* (BERT) [8]. Seperti yang telah disampaikan sebelumnya, tantangan utama penelitian ini adalah pemahaman konteks dan struktur bahasa Indonesia yang kompleks. Bahasa Indonesia memiliki kekayaan morfologi, penggunaan kata tidak baku, serta adanya campuran bahasa informal dan *slang* dalam ulasan pengguna, yang sering kali menyulitkan model dalam menangkap makna sebenarnya. Untuk mengatasi masalah ini, dapat digunakan model IndoBERT *pre-trained* korpus bahasa Indonesia yang lebih beragam (Indo4B), termasuk berita *online*, artikel *online*, media sosial, Wikipedia, *subtitle* video, dan *dataset* paralel [9].

Pada penelitian sebelumnya [10] LDA digunakan dalam mengekstraksi fitur dari *dataset* Amazon lalu model seperti *Naive Bayes*, SVM, *Random Forest*, LSTM, dan GRU digunakan untuk mengklasifikasikan sentimen berdasarkan aspek tersebut. *Random Forest* mencapai akurasi 94.5%, *F1-score* 95.45% untuk aspek *reliability*. Terdapat juga penelitian yang [11] mengajukan pendekatan dengan kombinasi LDA aspek dan TF-BERT untuk klasifikasi sentimen. Model kombinasi LDA-TF-BERT mendapat akurasi 82.19%, 79.52% *Macro-F1* untuk laptop dan 86.26%, 81.27% *Macro-F1* untuk restoran. Pada penelitian yang lain [12] juga LDA dievaluasi sebagai ekstraksi aspek yang kemudian dibandingkan dengan pelabelan manual. Dengan hasil *aspect accuracy* 96%, *aspect precision* 70%. Berdasarkan hasil dari penelitian yang terdahulu, pada penelitian ini diajukan pendekatan analisis sentimen berdasarkan aspek dengan LDA sebagai metode untuk ekstraksi aspek yang kemudian akan dilakukan sentimen analisis dengan model *pre-trained* IndoBERT. Adapun sistematika penulisan dari artikel ini sebagai berikut: Pendahuluan, Metode Penelitian, Hasil dan Pembahasan dan Kesimpulan.

2. Metode Penelitian

2.1. Tahapan Penelitian

Pada penelitian ini terdapat beberapa tahapan yang dilalui untuk mencapai hasil penelitian dengan pendekatan yang telah diajukan, tahapan tersebut dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Langkah awal penelitian ini adalah pengumpulan data yang akan digunakan dalam melatih dan menguji model analisis sentimen berdasarkan aspeknya. Setelah itu, dilakukan *preprocessing* data yang berfungsi dalam mempersiapkan data agar sesuai untuk digunakan saat pelatihan model. Setelah *preprocessing* data, selanjutnya dilakukan ekstraksi aspek dari data, agar setiap data memiliki aspek-aspek yang telah ditentukan. Setelah data sudah memiliki aspek akan dilanjutkan dengan melatih data tersebut agar dapat menganalisa sentimen dari data berdasarkan aspek tersebut. Tahap terakhir adalah evaluasi hasil dari metrik yang digunakan dalam analisis sentimen.

2.2. Sumber dan Pengumpulan Data

Data yang dikumpulkan pada penelitian ini adalah data kuantitatif dan sekunder. Dimana data pada penelitian ini merupakan data ulasan beserta *rating* aplikasi Stockbit pada platform GooglePlay yang diperoleh dari Kaggle. *Dataset* berjumlah 13.020 dalam format .csv dengan

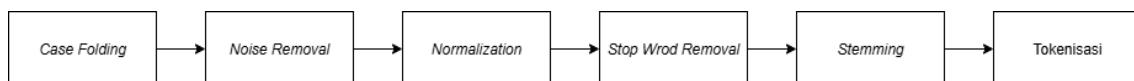
atribut, antara lain *reviewId*, *userName*, *userImage*, *content*, *score*, *thumbsUpCount*, *reviewCreatedVersion*, *at*, *replyContent*, *repliedAt*, *appVersion*. Namun atribut yang digunakan hanya *content* yang berisi ulasan pengguna dan *score* yang berisi *rating* (1-5) dari pengguna. *Rating* dari pengguna tersebut akan dikonversi ke dalam sentimen positif dan negatif. Dimana *rating* 1-3 (negatif) dan 4-5 (positif). Berikut merupakan distribusi dari data yang telah dikumpulkan:

Tabel 1. Distribusi Data

Sentimen Total	
Positif	9.328
Negatif	3.692
Total	13.020

2.3. Preprocessing Data

Pada penelitian ini dilakukan *preprocessing* data, tujuan utama dari proses ini adalah membersihkan, mengorganisir, dan mengubah data menjadi format yang sesuai dan dapat dimanfaatkan untuk melakukan analisis data. Terdapat berbagai proses yang dilalui dalam *preprocessing* data, sebagai berikut:



Gambar 2. Preprocessing Data

Proses pertama adalah *case folding*, pada proses ini akan mengubah semua teks pada data menjadi huruf kecil agar semua data direpresentasikan dengan seragam. Setelah itu proses *noise removal* untuk menghilangkan angka, simbol, URL, dan karakter non-alfabetik. Selanjutnya proses *normalization*, pada proses ini menghapus karakter berulang yang ada pada teks. Ini berfungsi untuk menyeragamkan format dan representasi data. Setelah itu dilakukan *stop word removal*, yaitu penghapusan kata-kata umum yang tidak memiliki makna signifikan dalam analisis teks, seperti “nya”, “aja”, “yg”, dan sebagainya. Selanjutnya adalah *stemming*, pada proses ini semua teks akan dihapus awalan, imbuhan, dan akhiran katanya. Ini bertujuan untuk mengembalikan kata ke bentuk dasarnya. Proses terakhir adalah tokenisasi, proses ini memecah teks menjadi unit-unit kata atau frasa. Proses ini penting karena banyak ulasan yang menggunakan kata tidak baku, *emoticon*, *slang*, dan tanda baca yang dapat menambah *noise* ke dalamnya. Berikut merupakan beberapa contoh ulasan yang harus melewati *preprocessing* data:

Tabel 2. Contoh Ulasan Setiap Kelas

Sentimen	Ulasan	Tantangan
Positif	"Tampilan aplikasi cukup ringkas namun lengkap dan sudah mencakup hal yg penting sehingga mempermudah dalam memantau dan bertransaksi.Helpdesk fastrespon jika ada masalah.Thebest lah aplikasi ini,jadi mempermudah dalam berinvestasi 😊👍"	Kata tidak baku ("yg", "Thebest lah"), campuran bahasa, <i>emoticon</i> "😊👍"
Negatif	"Deposit 2 jam ga masuk, komplek ke CS disuruh SS bukti trf pas sy coba kok sekarang sudah ga bisa kirim foto bukti trf padahal sebelumnya bisa, berasa dijebak bgt ni mah kalo smpe deposit ga masuk. Sebelum deposit sy masuk, bintang yg sy kasih tetap satu"	<i>Slang</i> ("ga", "bgt", "trf", "sy"), campuran formal-informal, struktur tidak baku

2.4. Ekstraksi Aspek *Latent Dirichlet Allocation* (LDA)

LDA merupakan metode statistik yang mendeteksi topik-topik tersembunyi di dalam dokumen dengan memodelkan dokumen sebagai campuran topik, dan topik sebagai campuran kata-kata. *Bag-of-Words* (BoW) merepresentasikan dokumen sebagai vektor frekuensi kata berdasarkan kosakata keseluruhan [13]. Ini memungkinkan teks dianalisis secara numerik oleh suatu algoritma *machine learning* seperti LDA, meskipun BoW tidak memperhatikan urutan dan konteks antar kata. Penelitian ini menerapkan *topic modeling* LDA untuk ekstraksi aspek, dengan representasi fitur teks berbasis BoW.

a. Representasi Fitur *Bag-of-Words* (BoW)

Pada penelitian ekstraksi fitur BoW dilakukan untuk mengubah data yang telah melewati proses pre-pemrosesan akan direpresentasikan ke dalam numerik berbentuk matriks frekuensi kata.

$$D = \{d_1, d_2, \dots, d_n\} \tag{1}$$

$$V = \{w_1, w_2, \dots, w_m\} \tag{2}$$

$$x_i = [f(w_1, d_i), f(w_2, d_i), \dots, f(w_m, d_i)] \tag{3}$$

Keterangan:

D = Document

V = Vocabulary

$f(w_j, d_i)$ = Frekuensi kemunculan w_j di dokumen d_i

m = Jumlah kata unik pada sebuah korpus

d_i = Dokumen ke i dalam korpus

w_j = Kata ke j dalam vocabulary

Dalam penelitian ini, proses ekstraksi fitur dilakukan dengan bantuan modul *CountVectorizer*, dengan menggunakan parameter sebagai berikut:

Tabel 3. Parameter BoW

Parameter	Nilai
max_df	0.85
min_df	5
$stop_word$	None
$ngram_range$	(1,2)

Parameter $max_df = 0.85$ digunakan untuk menghilangkan kata-kata yang terlalu umum, sementara $min_df = 5$ menghapus kata yang sangat jarang muncul. Selain itu, digunakan $ngram_range = (1,2)$ untuk mempertimbangkan unigram dan bigram guna menangkap konteks kata yang lebih luas.

b. Model *Latent Dirichlet Allocation* (LDA)

Pada penelitian ini LDA digunakan untuk mengekstraksi fitur aspek-aspek utama pada ulasan pengguna. LDA mendeteksi topik-topik yang tersembunyi yang tersembunyi di dalam dokumen, dimana setiap topik direpresentasikan oleh kata yang sering muncul pada dokumen. Pada penelitian ini, proses ekstraksi fitur dilakukan dengan bantuan modul *scikit-learn*, dengan menggunakan parameter sebagai berikut:

Tabel 4. Parameter LDA

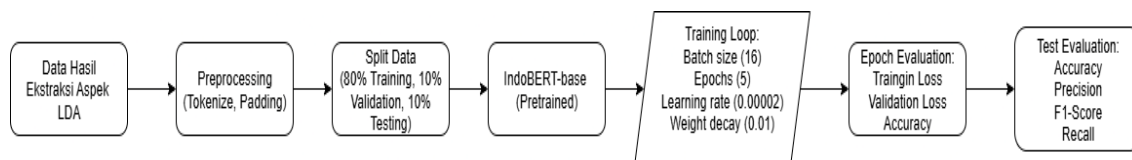
Parameter	Nilai
<i>n_components</i>	5
<i>learning_method</i>	'batch'
<i>random_state</i>	42
<i>max_iter</i>	20
<i>evaluate_every</i>	1

Model dikonfigurasi dengan jumlah topik $n_components = 5$, yang diasumsikan merepresentasikan lima aspek utama dari ulasan pengguna. Parameter *learning_method = 'batch'*, proses pelatihan menggunakan pendekatan *batch Variational Bayes* dengan jumlah maksimum iterasi sebanyak *max_iter = 20*. Nilai *random_state* berfungsi agar hasil model dapat direproduksi.

2.5. Analisis Sentimen IndoBERT

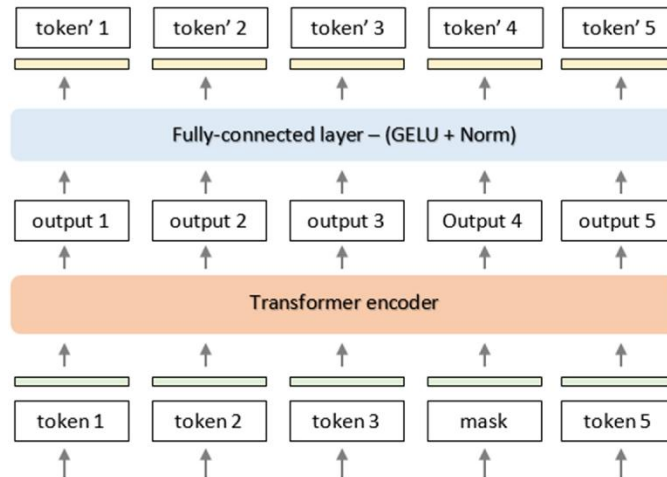
Penggunaan IndoBERT sebagai model klasifikasi pada penelitian ini bertujuan untuk menghasilkan representasi teks bahasa Indonesia yang akurat dan kaya secara linguistik. IndoBERT adalah model berbasis BERT yang telah melalui pretraining menggunakan korpus Bahasa Indonesia yang besar (korpus Indo4B) [14]. Arsitektur dari IndoBERT sama dengan BERT, yaitu arsitektur *transformer* dengan *bidirectional encoder*. Terdapat dua jenis arsitektur dari IndoBERT, yaitu: IndoBERT-*base* dan IndoBERT-*large* model [15]. Adapun alur dari klasifikasi dengan model IndoBERT pada penelitian ini, sebagai berikut:

Gambar 3. Alur Klasifikasi IndoBERT



Hasil dari ekstraksi aspek dengan LDA akan melalui *preprocessing* data, dimana akan dilakukan *tokenizing* dan *padding*. *Tokenizing* mengubah data ke dalam bagian-bagian kecil, kemudian *padding* menyamakan panjang dari input teks agar memiliki panjang token yang sama. Setelah *preprocessing* data, data kemudian dibagi ke dalam tiga bagian. Setelah itu data akan diolah dengan model IndoBERT-*base*. Setelah model selesai dilatih akan dievaluasi setiap *epoch* dengan menghitung *training loss*, *validation loss*, dan *accuracy*. Kemudian hasil evaluasi *epoch* dengan akurasi tertinggi akan digunakan dalam melatih kembali model dengan data pengujian, yang kemudian dievaluasi dengan *confusion matrix*.

Pada penelitian ini menggunakan IndoBERT-*base* yang memiliki arsitektur yang identik dengan BERT. Tanpa adanya modifikasi struktural, perbedaan utamanya hanya pada korpus bahasa yang digunakan. Arsitektur dari IndoBERT-*base* terdiri oleh 12 *layer* dari *transformer*, 12 *attention heads*, dan 110 juta parameter [16].



Gambar 4. Arsitektur BERT [17]

Setelah proses ekstraksi aspek sudah selesai, selanjutnya proses dalam melatih model IndoBERT untuk mengklasifikasi sentimen pada ulasan. Format data mencakup ulasan, aspek, sentimen. Selanjutnya data akan dibagi menjadi tiga bagian, yaitu: 80% data pelatihan, 10% data validasi, dan 10% data pengujian.

Tabel 5. Distribusi Data Pelatihan

Sentimen	Data Pelatihan	Data Validasi	Data Pengujian	Total
Positif	7.438	951	939	9.328
Negatif	2.978	351	363	3.692
Total	10.416	1.302	1.302	13.020

Proses pelatihan model dilakukan dengan menggunakan konfigurasi dari modul *Hugging Face Transformer*. Adapun parameter yang digunakan untuk melatih model, sebagai berikut:

Tabel 6. Parameter Pelatihan IndoBERT

Parameter	Nilai
Batch size	16
Epochs	5
Learning rate	0.00002
Weight decay	0.01

Meskipun arsitektur IndoBERT secara *default* menerapkan *dropout layer* dengan rasio 0.01, regulasi tambahan seperti *early stopping* dapat mencegah potensi dari *overfitting* pada model. Pada penelitian ini tidak menerapkan *early stopping* secara langsung. Sehingga pelatihan akan terus berlanjut hingga *epoch* yang kelima tanpa adanya interupsi evaluasi dari performa validasi. Selain *early stopping* penelitian ini juga hanya menerapkan regulasi *dropout* secara *default* yaitu 0.1, ini artinya 10% neuron akan dinonaktifkan secara acak selama pelatihan. Regulasi ini berfungsi dalam mencegah *overfitting* pada model.

2.6. Evaluasi Model IndoBERT

Evaluasi dari performa menggunakan *confusion matrix* dan evaluasi berdasarkan label. *Confusion matrix* adalah representasi evaluasi performa model klasifikasi yang berbentuk tabel dua dimensi. Dalam matriks ini, baris-baris menunjukkan label sebenarnya (*ground truth*) dari data masukan, sedangkan kolom-kolom menunjukkan label yang diprediksi oleh model [18]. Terdapat empat metrik evaluasi berdasarkan *confusion matrix*, yaitu: *accuracy*, *precision*, *recall*, dan *F1-score*.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F1 - score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (7)$$

Pada evaluasi berbasis label mempertimbangkan setiap label secara terpisah, ini berarti klasifikasi multi-label diubah ke dalam bentuk klasifikasi biner untuk setiap labelnya. Adapun empat kemungkinan dari *output* klasifikasi biner: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) [19].

2.7. Lingkungan Pengembangan

Penelitian ini diimplementasikan pada bahasa pemrograman Python dalam lingkungan komputasi berbasis *cloud* pada Google Collaboratory (Google Colab), dengan menggunakan *Graphics Processing Unit* (GPU) NVIDIA Tesla T4. Modul yang digunakan pada penelitian ini meliputi *re* untuk operasi pembersihan teks, *pandas* untuk manipulasi data, *Sastrawi* untuk *stop word removal* dan *stemming* bahasa Indonesia, *scikit-learn* untuk ekstraksi fitur *BoW* dan pemodelan topik LDA. Untuk tahap analisis sentimen, model IndoBERT dengan arsitektur *indobenchmark/indobert-base-p1* digunakan dan di-*fine-tune* sesuai dengan kebutuhan klasifikasi sentimen berbasis aspek.

3. Hasil dan Pembahasan

3.1. Hasil Preprocessing Data

Setelah *preprocessing* data, data teks mengalami perubahan struktur dan isi. Teks yang sebelumnya tidak seragam seperti “Hey admin, harap update berikutnya aplikasi ini bisa menjadi mode landscape”, setelah melalui proses pada Gambar 2 teks menjadi [“admin”, “harap”, “update”, “berikut”, “aplikasi”, “jadi”, “mode”, “landscape”, “potret”], sehingga sesuai format dan dapat digunakan untuk menganalisis data sentimen. Berikut adalah hasil dari setiap proses yang dilalui:

Tabel 7. Hasil Preprocessing

Proses	Teks	Sentimen
<i>Case Folding</i>	hey admin, harap update berikutnya aplikasi ini bisa menjadi mode landscape	Positif
<i>Noise Removal</i>	hey admin harap update berikutnya aplikasi ini bisa menjadi mode landscape	Positif
<i>Normalization</i>	hey admin harap update berikut aplikasi ini bisa menjadi mode landscape	Positif

Proses	Teks	Sentimen
Stop Word Removal	admin harap update berikutnya aplikasi menjadi mode landscape	Positif
Stemming	admin harap update berikut aplikasi jadi mode landscape	Positif
Tokenization	["admin", "harap", "update", "berikut", "aplikasi", "jadi", "mode", "landscape", "potret"]	Positif

3.2. Hasil Ekstraksi Aspek *Latent Dirichlet Allocation* (LDA)

Ekstraksi fitur dengan pendekatan *topic modeling* menggunakan LDA dilakukan dengan mengasumsikan aspek utama dari teks terdiri dari lima aspek utama, ini didapat setelah melakukan eksperimen beberapa jumlah aspek (4-6) dan mengukur nilai *topic coherence* yang paling tinggi. Hasil eksperimen dapat dilihat pada Tabel 8.

Tabel 8. *Topic Coherence*

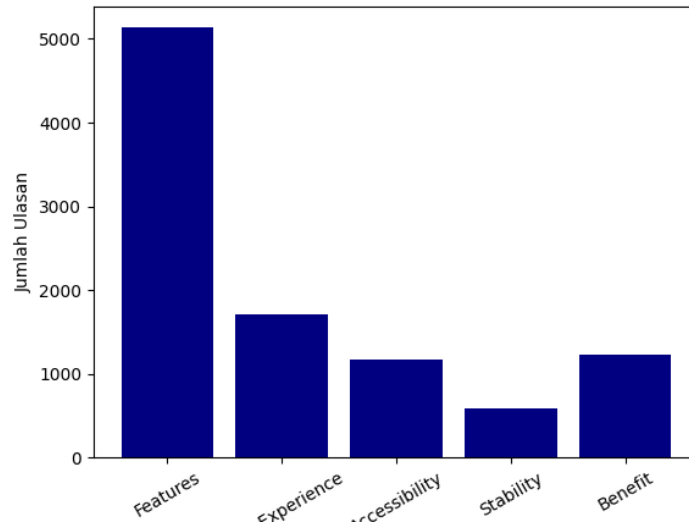
n_components	Nilai
4	0.5881
5	0.6191
6	0.6166

Berdasarkan hasil eksperimen, pemilihan parameter $n_components = 5$ menghasilkan nilai *topic coherence* tertinggi, yaitu sebesar 0.6191. Nilai ini lebih tinggi dibandingkan dengan $n_components = 4$ dan $n_components = 6$, yang masing-masing memperoleh skor sebesar 0.5881 dan 0.6166. Oleh karena itu, aspek yang diekstraksi ditetapkan sebanyak lima aspek utama, yaitu: *features*, *user experience*, *accessibility*, *stability*, dan *benefit*. Adapun kelima aspek beserta kata kunci utama dan distribusi datanya disajikan sebagai berikut:

Tabel 9. Hasil Ekstraksi Aspek

Aspek	Kata Kunci	Jumlah
<i>Features</i>	["mantap", "mudah", "fitur", "stockbit", "lengkap", "buat", "mula", "keren", "bagus", "paham"]	5.135
<i>User Experience</i>	["bagus", "aplikasi", "dulu", "stockbit", "baru", "nice", "buka", "coba", "suka", "pakai"]	1.705
<i>Accessibility</i>	["lama", "hari", "masuk", "baik", "daftar", "aplikasi", "saham", "dana", "stockbit", "akun"]	1.167
<i>Stability</i>	["sering", "jam", "error", "aplikasi", "maintenance", "bursa", "buka", "jadi", "eror", "stockbit"]	589
<i>Benefit</i>	["sangat", "good", "saham", "bantu", "aplikasi", "mudah", "sangat bantu", "bagus", "ajar", "app"]	1.226

Distribusi dari jumlah setiap aspek dapat dilihat pada Gambar 6, dimana aspek *features* memiliki ulasan yang paling banyak yaitu sebanyak 5.135 ulasan, sebaliknya aspek *stability* mendapatkan ulasan yang paling sedikit yaitu 589 ulasan. Dari distribusi tersebut didapatkan bahwa aspek *features* mendapatkan banyak perhatian dari pengguna, sebaliknya aspek *stability* bukan menjadi perhatian utama dari pengguna aplikasi Stockbit.



Gambar 6. Visualisasi Distribusi Aspek

Selain berdasarkan nilai dari *topic coherence*, visualisasi dengan WordCloud juga menunjukkan kata-kata yang paling sering muncul pada korpus teks, seperti “aplikasi”, “stockbit”, “saham”, “bagus”, “baik”, “fitur”, “mudah”, “paham”, “lama”. Kemunculan kata-kata ini juga mendukung dari lima aspek yang telah ditentukan berdasarkan *topic coherence* sebelumnya. Dimana kata-kata tersebut mewakili limas aspek utama, yaitu: *features*, *user experience*, *accessibility*, *stability*, dan *benefit*.



Gambar 7. WordCloud Ulasan

3.3. Hasil Analisis Sentimen IndoBERT

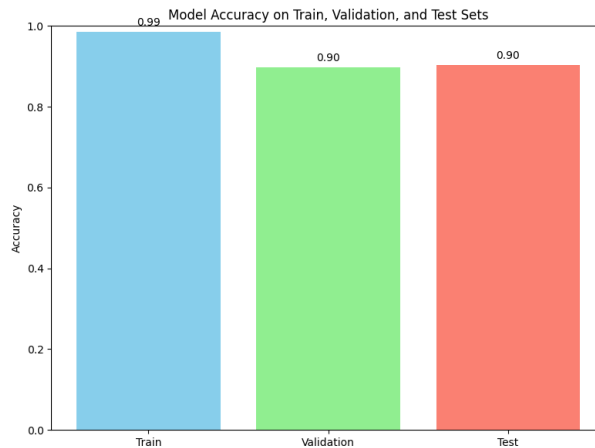
Hasil pelatihan pada Tabel 10. menunjukkan bahwa *training loss* menurun konsisten, namun *validation loss* justru meningkat dari 0.2459 pada *epoch* 1 menjadi 0.5030 pada *epoch* 5. Hal ini mengindikasikan terjadinya *overfitting*, dimana model terlalu menyesuaikan dengan data pelatihan sehingga model kurang dalam kemampuan generalisasi. Meskipun akurasi hanya sedikit menurun dari 91.01% menjadi 89.78%, tren ini menunjukkan bahwa performa terbaik kemungkinan dicapai pada *epoch* pertama, sebelum *overfitting* terjadi.

Tabel 10. Evaluasi Hasil Training

Epoch	Training Loss	Validation Loss	Accuracy
1	0.279700	0.245894	0.910138
2	0.221500	0.256626	0.907834
3	0.183100	0.329100	0.903994

Epoch	Training Loss	Validation Loss	Accuracy
4	0.123300	0.439745	0.903226
5	0.066500	0.502974	0.897849

Hasil evaluasi pada Gambar 8 menunjukkan bahwa akurasi pelatihan mencapai 98.54%, yang menandakan model terlalu menyesuaikan dengan data pelatihan. Berbeda dengan nilai akurasi validasi sebesar 89.78% dan akurasi pengujian sebesar 90.25% menunjukkan adanya *gap* yang cukup signifikan antara data pelatihan dan data baru. Perbedaan ini mengindikasikan bahwa model mengalami *overfitting*.



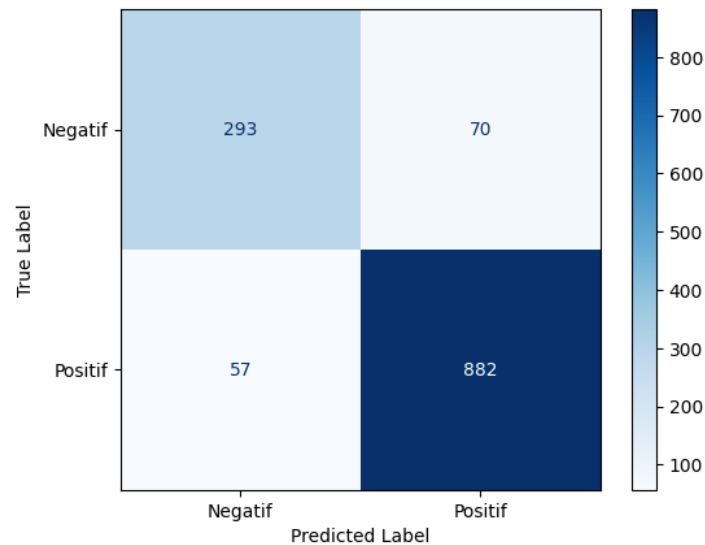
Gambar 8. Visualisasi Akurasi *Train*, *Validation*, & *Test*

Dilihat dari Tabel 11 performa model sudah dapat mengklasifikasikan sentimen, dengan akurasi 90.86%. Label positif memiliki skor tertinggi dengan *precision* 92.71%, *recall* 94.78%, dan *F1-score* 93.73%, menandakan model sangat andal dalam mengenali sentimen positif. Sementara itu, pada label negatif, performa sedikit lebih rendah dimana mendapatkan *F1-score* sebesar 83.12% dengan *recall* 80.72%, yang mengindikasikan bahwa model terkendala dalam mendeteksi sentimen negatif dibandingkan dengan sentimen positif. Nilai *macro average F1-score* sebesar 88.43% menunjukkan bahwa model mampu menangani kedua kelas dengan cukup seimbang.

Tabel 11. Hasil Evaluasi *Testing*

Label Sentimen	Precision	Recall	F1-score
Positif	92.71%	94.78%	93.73%
Negatif	85.67%	80.72%	83.12%
Macro Average	89.19%	87.75%	88.43%
Accuracy	90.86%		

Berdasarkan *confusion matrix* pada Gambar 9, model berhasil mengklasifikasikan 293 dari 363 data negatif secara benar, dan 882 dari 939 data positif secara akurat. Kesalahan klasifikasi terjadi pada 70 data negatif diprediksi sebagai positif, serta 57 data positif diprediksi sebagai negatif. Hal ini menunjukkan bahwa meskipun performa model secara umum tinggi, terdapat kecenderungan sedikit lebih banyak *false positive*, yang dapat mempengaruhi interpretasi hasil jika fokus utama adalah deteksi sentimen negatif.



Gambar 9. Confusion Matrix

3.4. Pembahasan

Hasil ekstraksi aspek menunjukkan bahwa pemilihan parameter $n_components = 5$ menghasilkan nilai *topic coherence* tertinggi sebesar 0.6191, yang mengindikasikan bahwa struktur lima topik merupakan representasi yang paling optimal terhadap data ulasan yang dianalisis. Nilai ini mencerminkan bahwa jumlah lima topik berhasil mencapai keseimbangan antara *coverage* dan *specificity*, serta sesuai dengan pola semantik alami yang terkandung dalam data. Dengan demikian, lima aspek yang dihasilkan tidak hanya koheren secara makna, tetapi juga relevan secara konseptual dalam konteks isi ulasan pengguna.

Hasil pelatihan model IndoBERT-base menunjukkan performa klasifikasi sentimen yang tinggi. Model mencapai akurasi pengujian sebesar 90.25%, dengan akurasi pelatihan sebesar 98.54% dan akurasi validasi sebesar 89.78%. Meskipun akurasi pelatihan sangat tinggi, terdapat selisih cukup besar dibandingkan akurasi validasi dan pengujian, yang mengindikasikan adanya potensi *overfitting*, yaitu kondisi ketika model terlalu menyesuaikan diri terhadap data latih sehingga kinerjanya kurang optimal dalam generalisasi ke data baru.

Evaluasi lebih lanjut melalui metrik klasifikasi menunjukkan bahwa model lebih baik dalam mengklasifikasikan sentimen positif dibandingkan negatif. Dapat dilihat dari hasil *F1-score* positif sebesar 93.73%, jauh lebih tinggi dibanding *F1-score* negatif yang hanya 83.12%. Ketimpangan ini dapat dijelaskan oleh distribusi data yang tidak seimbang, dimana jumlah data sentimen positif 9.328 jauh lebih banyak dibandingkan sentimen negatif 3.692. Ketidakseimbangan ini dapat menyebabkan model lebih terbiasa mendeteksi sentimen positif dan kesulitan mengenali pola sentimen negatif.

Confusion matrix pada Gambar 9 memperkuat temuan ini. Model mampu mengklasifikasikan 293 data negatif dan 882 data positif secara benar. Namun, terdapat 70 kesalahan klasifikasi negatif menjadi positif (*false positive*), serta 57 kesalahan klasifikasi positif menjadi negatif (*false negative*). Meskipun jumlah kesalahan relatif kecil, distribusi kesalahan ini menunjukkan bahwa model sedikit lebih cenderung memprediksi data ke dalam kelas positif.

4. Kesimpulan

Berdasarkan hasil eksperimen, pemilihan jumlah topik sebanyak lima ($n_component = 5$) merepresentasikan data ulasan secara optimal, ini didukung oleh nilai *topic coherence* yang diperoleh sebesar 0.6191. Ini menunjukkan bahwa lima topik tersebut mencerminkan struktur semantik yang paling sesuai terhadap ulasan pengguna.

Dari hasil klasifikasi sentimen, model menunjukkan performa baik, dengan akurasi sebesar 90.86%. Kinerja tertinggi tercapai pada label positif, dengan *precision* 92.71%, *recall* 94.78%, dan *F1-score* 93.73%, yang menunjukkan keandalan model dalam mendeteksi sentimen positif. Sebaliknya, pada label negatif, model mencatat *F1-score* sebesar 83.12% dan *recall* 80.72%, menunjukkan bahwa deteksi sentimen negatif masih menjadi tantangan. Meskipun demikian, macro average *F1-score* sebesar 88.43% menunjukkan bahwa model tetap mampu menangani kedua kelas dengan performa yang relatif seimbang. Dari hasil performa, dapat disimpulkan bahwa model cenderung lebih kuat dalam mengenali sentimen positif dibandingkan dengan negatif. Ini disebabkan oleh distribusi data yang tidak seimbang, dimana jumlah data sentimen positif adalah 9.328 sedangkan sentimen negatif hanya 3.692, sehingga model lebih terbiasa mengenali pola-pola sentimen positif. Untuk penelitian ke depan agar memperoleh hasil yang lebih maksimal, disarankan untuk memastikan bahwa data seimbang atau tidak terlalu timpang, menambahkan regulasi *early stopping* dan *dropout* untuk menghindari *overfitting* pada model.

Daftar Pustaka

- [1] Rolen, S. Lim, T. C. Djaja, and M. Linus, "Literasi Keuangan Digital: Bijak Mengelola Keuangan Pada Gen Z," *Jurnal Serina Abdimas*, vol. 2, no. 3, pp. 1728–1733, Nov. 2024, doi: <https://doi.org/10.24912/jsa.v2i4.33519>.
- [2] G. Kontonatsios *et al.*, "FABSA: An aspect-based sentiment analysis dataset of user reviews," *Neurocomputing*, vol. 562, Dec. 2023, doi: 10.1016/j.neucom.2023.126867.
- [3] S. Datta and S. Chakrabarti, "Aspect based sentiment analysis for demonetization tweets by optimized recurrent neural network using fire fly-oriented multi-verse optimizer", doi: 10.1007/s12046-021-01608-1S.
- [4] R. J. Dhanal and V. R. Ghorpade, "Aspect-based sentiment-analysis using topic modelling and machine-learning," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 6, pp. 6689–6698, Dec. 2024, doi: 10.11591/ijece.v14i6.pp6689-6698.
- [5] Y. C. Hua, P. Denny, J. Wicker, and K. Taskova, "A systematic review of aspect-based sentiment analysis: domains, methods, and trends," *Artif Intell Rev*, vol. 57, no. 11, Nov. 2024, doi: 10.1007/s10462-024-10906-z.
- [6] D. Maier *et al.*, "Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology," *Commun Methods Meas*, vol. 12, no. 2–3, pp. 93–118, 2018, doi: 10.1080/19312458.2018.1430754.
- [7] U. Chauhan and A. Shah, "Improving Semantic Coherence of Gujarati Text Topic Model Using Inflectional Forms Reduction and Single-letter Words Removal," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 20, no. 1, Mar. 2021, doi: 10.1145/3447760.
- [8] A. Turchin, S. Masharsky, and M. Zitnik, "Comparison of BERT implementations for natural language processing of narrative medical documents," *Inform Med Unlocked*, vol. 36, Jan. 2023, doi: 10.1016/j.imu.2022.101139.
- [9] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 348–354, Dec. 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.
- [10] D. R. I. M. Setiadi, D. Marutho, and N. A. Setiyanto, "Comprehensive Exploration of Machine and Deep Learning Classification Methods for Aspect-Based Sentiment Analysis with Latent Dirichlet Allocation Topic Modeling," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 1, pp. 12–22, May 2024, doi: 10.62411/faith.2024-3.
- [11] R. Dhanal and V. Ghorpade, "Aspect-based sentiment-analysis using topic modelling and machine-learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, p. 6689, Jul. 2024, doi: 10.11591/ijece.v14i6.pp6689-6698.
- [12] N. Ashikin, A. Samad, and R. Ibrahim, "Extraction of Aspect Categories for Sentiment Analysis towards Technical and Vocational Education and Training (TVET) in Malaysia using Topic Modelling Approach Politeknik & Extraction of Aspect Categories for Sentiment Analysis towards Technical and Vocational Education and Training (TVET) in Malaysia using Topic Modelling Approach," 2019.

- [13] M. Graff, D. Moctezuma, and E. S. Téllez, "Bag-of-Word approach is not dead: A performance analysis on a myriad of text classification challenges," *Natural Language Processing Journal*, vol. 11, p. 100154, 2025, doi: <https://doi.org/10.1016/j.nlp.2025.100154>.
- [14] B. Willie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," 2020. [Online]. Available: <https://arxiv.org/abs/2009.05387>
- [15] L. Geni, E. Yulianti, and D. Sensuse, "Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using BERT Language Models," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, pp. 746–757, Jul. 2023, doi: 10.26555/jiteki.v9i3.26490.
- [16] A. Patel, P. Oza, and S. Agrawal, "Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model," *Procedia Comput Sci*, vol. 218, pp. 2459–2467, 2023, doi: <https://doi.org/10.1016/j.procs.2023.01.221>.
- [17] C. H. Miranda, G. Sanchez-Torres, and D. Salcedo, "Exploring the Evolution of Sentiment in Spanish Pandemic Tweets: A Data Analysis Based on a Fine-Tuned BERT Architecture," *Data (Basel)*, vol. 8, no. 6, Jun. 2023, doi: 10.3390/data8060096.
- [18] L.-E. Pommé, R. Bourqui, R. Giot, and D. Auber, "Relative Confusion Matrix: Efficient Comparison of Decision Models," in *2022 26th International Conference Information Visualisation (IV)*, 2022, pp. 98–103. doi: 10.1109/IV56949.2022.00025.
- [19] D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić, "Multi-label Classifier Performance Evaluation with Confusion Matrix," Academy and Industry Research Collaboration Center (AIRCC), Jun. 2020, pp. 01–14. doi: 10.5121/csit.2020.100801.

Halaman ini sengaja dibiarkan kosong