

# Analisis Sentimen Ulasan Produk Kecantikan Menggunakan Bi-LSTM dan *Weighted FastText Embedding*

Shelomita Putrinda Culio<sup>a1</sup>, Luh Gede Astuti<sup>a2</sup>, Ida Ayu Gde Suwiprabayanti Putra<sup>a3</sup>, I Made Widiartha<sup>a4</sup>

<sup>a</sup>Program Studi Informatika, Universitas Udayana  
Bali, Indonesia

<sup>1</sup>culio.2208561002@student.unud.ac.id

<sup>2</sup>lg.astuti@unud.ac.id

<sup>3</sup>iagsuwiprabayantiputra@unud.ac.id

<sup>4</sup>madewidiartha@unud.ac.id

## Abstract

*The rapid expansion of the beauty product sector in Indonesia has contributed to a growing volume of consumer reviews, which hold significant potential for extracting insights into user perceptions toward products. This study seeks to conduct sentiment analysis on beauty product reviews by employing the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm, coupled with FastText word embeddings weighted through an Inverse Document Frequency (IDF) scheme. The data utilized in this research were sourced from the Female Daily platform. The research process encompasses text preprocessing, word representation using FastText, IDF-based word weighting, and Bi-LSTM model training with hyperparameter tuning involving the number of LSTM units, dropout rate, and learning rate. The evaluation results on the testing data reveal that the model achieved an accuracy of 85.96%. The optimal hyperparameter configuration consisted of 96 LSTM units, a dropout value of 0.3, and a learning rate of 0.0008. The best-performing model was subsequently deployed into a web-based system for automated sentiment analysis of beauty product reviews.*

**Keywords:** Sentiment Analysis, Bi-LSTM, FastText, IDF, Beauty Products

## 1. Pendahuluan

Pertumbuhan industri produk kecantikan di Indonesia menunjukkan peningkatan yang signifikan seiring meningkatnya minat masyarakat terhadap perawatan wajah dan tubuh. Data menunjukkan bahwa nilai pasar industri kosmetik nasional pada tahun 2025 mencapai Rp35,6 triliun dengan proyeksi pertumbuhan sebesar 4,73% setiap tahunnya [1]. Selain itu, segmen produk kecantikan juga secara konsisten termasuk dalam tiga besar kategori penjualan tertinggi di *marketplace* sejak tahun 2018 hingga 2022, yaitu tercatat adanya aktivitas transaksi dengan total nominal Rp13.287,4 triliun dan jumlah transaksi mencapai 145,44 juta [2]. Peningkatan tersebut juga diikuti oleh semakin banyaknya ulasan konsumen yang dibagikan secara daring sebagai bentuk pengalaman penggunaan produk. Salah satu platform yang menyediakan wadah bagi pengguna untuk berbagi ulasan produk kecantikan adalah Female Daily. Melalui fitur ulasan pada platform tersebut, pengguna dapat memberikan penilaian dan opini terhadap produk yang digunakan, sehingga ulasan tersebut dapat menjadi sumber informasi untuk memahami persepsi konsumen. Akan tetapi, jumlah ulasan yang terus bertambah membuat proses analisis manual menjadi kurang efisien dan memerlukan metode otomatis untuk mengekstraksi sentimen teks ulasan.

Pendekatan yang dapat digunakan adalah analisis sentimen, yaitu salah satu metode dalam *Natural Language Processing* (NLP) yang berfungsi untuk mengelompokkan opini dalam teks ke dalam kelas sentimen yang telah ditentukan, seperti positif atau negatif. Salah satu arsitektur yang banyak digunakan dalam analisis teks adalah *Bidirectional Long Short-Term Memory* (Bi-LSTM), yang mampu memproses urutan kata dari dua arah sehingga dapat mempertimbangkan konteks sebelumnya dan sesudahnya secara bersamaan [3]. Penelitian sebelumnya menunjukkan bahwa model Bi-LSTM dapat menghasilkan performa yang lebih baik dibandingkan beberapa model *recurrent* lainnya (RNN, LSTM, GRU) dalam tugas analisis sentimen layanan medis [4]. Di sisi lain, model transformer modern seperti

BERT dan IndoBERT telah menunjukkan hasil yang sangat baik pada berbagai tugas NLP berbahasa Indonesia [5][6]. Meskipun demikian, model transformer umumnya memerlukan sumber daya komputasi yang jauh lebih besar dan waktu pelatihan yang lebih lama [7], sehingga pendekatan Bi-LSTM digunakan sebagai model yang lebih ringan dengan performa yang kompetitif.

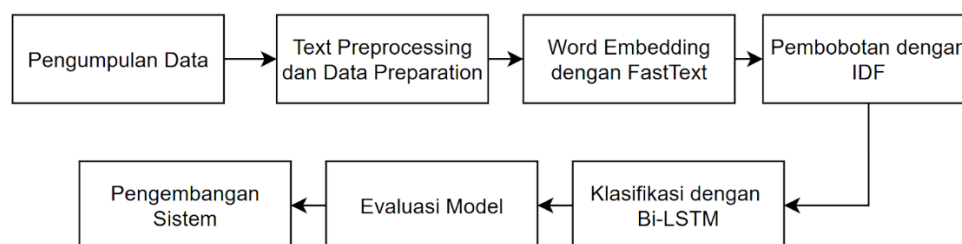
Dalam implementasinya, teks perlu direpresentasikan dalam bentuk numerik agar dapat diproses oleh model *deep learning*. Salah satu teknik yang umum digunakan adalah *word embedding*, yang merepresentasikan kata dalam bentuk vektor dengan mempertimbangkan hubungan semantik antar kata. *FastText* merupakan metode *word embedding* yang menawarkan keunggulan karena memanfaatkan representasi *subword* berbasis n-gram, sehingga lebih efektif menangani masalah *Out-of-Vocabulary* (OOV), kata tidak baku, singkatan, serta campuran bahasa Inggris yang sering muncul dalam ulasan produk kecantikan berbahasa Indonesia [8]. Beberapa penelitian sebelumnya telah membuktikan bahwa *FastText* memberikan performa yang lebih baik dibandingkan metode *embedding* lain seperti *Word2Vec* dan *GloVe* pada tugas klasifikasi teks [9].

Akan tetapi, *word embedding* seperti *FastText* umumnya hanya memberikan bobot yang sama terhadap setiap kata dalam dokumen (*uniform weighting*), padahal tidak semua kata memiliki tingkat kepentingan yang sama dalam menentukan sentimen [10]. Kata-kata opini, istilah khusus kecantikan, serta kata langka yang bermakna kuat sering kali lebih menentukan polaritas sentimen dibandingkan kata-kata umum. Kondisi ini menyebabkan representasi teks kurang optimal. Beberapa penelitian sebelumnya telah mengombinasikan *FastText* dan model *deep learning* seperti Bi-LSTM untuk analisis sentimen [11][12], tetapi penerapan mekanisme pembobotan kata pada *FastText* masih terbatas, khususnya pada domain ulasan produk kecantikan berbahasa Indonesia. Akibatnya, kata-kata yang lebih informatif belum memperoleh kontribusi yang optimal dalam representasi teks. Metode pembobotan seperti *Inverse Document Frequency* (IDF) dapat dikombinasikan dengan *FastText* untuk menekankan kata-kata yang lebih informatif dalam teks, dengan memberikan bobot yang lebih tinggi pada kata [13].

Oleh karena itu, penelitian ini mengusulkan penerapan model Bi-LSTM dengan representasi *weighted FastText embedding* berbasis IDF untuk melakukan analisis sentimen pada ulasan produk kecantikan dari platform Female Daily. Penelitian ini bertujuan mengatasi kelemahan representasi teks pada ulasan produk kecantikan berbahasa Indonesia yang umumnya mengandung bahasa tidak baku, singkatan, campuran bahasa asing, dan istilah khusus kecantikan. Kontribusi utama penelitian ini adalah penerapan *weighted FastText embedding* berbasis IDF untuk meningkatkan representasi kata-kata informatif sehingga diharapkan mampu menghasilkan representasi teks yang lebih representatif terhadap informasi sentimen

## 2. Metode Penelitian

Penelitian ini diawali dengan perolehan dataset ulasan produk kecantikan dari platform Female Daily. Data kemudian melalui tahap *text preprocessing* dan *data preparation* untuk mempersiapkan teks sebelum diproses lebih lanjut. Selanjutnya, teks direpresentasikan dalam bentuk vektor menggunakan *word embedding FastText* dan diberi pembobotan menggunakan skema IDF. Representasi tersebut digunakan sebagai input pada model klasifikasi Bi-LSTM dengan proses *hyperparameter tuning* untuk memperoleh model terbaik. Model dievaluasi menggunakan *confusion matrix* dengan empat metrik evaluasi. Model terbaik kemudian diimplementasikan ke dalam sistem berbasis web. Alur penelitian secara keseluruhan diperlihatkan pada Gambar 1.



Gambar 1. Alur Penelitian

## 2.1. Pengumpulan Data

Data yang digunakan berupa data sekunder yang bersumber dari penelitian sebelumnya [14]. Dataset berisi ulasan produk kecantikan dalam bahasa Indonesia yang dikumpulkan dari platform Female Daily. Setiap ulasan telah memiliki label sentimen berdasarkan nilai *rating*, yaitu *rating* 1 dan 2 dikategorikan sebagai sentimen negatif, *rating* 4 dan 5 sebagai sentimen positif, sedangkan *rating* 3 dikategorikan sebagai sentimen netral. Dataset awal terdiri dari 50.000 ulasan dengan distribusi 36.287 ulasan positif, 9.423 ulasan negatif, dan 4.290 ulasan netral. Data tersebut mencakup ulasan dari 10 kategori produk kecantikan dengan jumlah ulasan tertinggi pada platform Female Daily, seperti *Lipstick*, *Toner*, *Facial Wash*, *Wash-Off*, *Face*, dan *Serum & Essence*.

## 2.2. Text Preprocessing dan Data Preparation

Data ulasan terlebih dahulu diproses melalui serangkaian tahapan *text preprocessing* untuk meningkatkan kualitas teks. Proses ini mencakup penyeragaman huruf melalui *case folding*, pembersihan teks dari karakter yang tidak diperlukan melalui *cleaning*, serta pemecahan kalimat menjadi kata-kata tunggal menggunakan *tokenizing*. Selain itu, diterapkan normalisasi untuk menyesuaikan kata-kata nonstandar menjadi bentuk baku, dengan isi kamus yang disesuaikan dengan karakteristik dataset produk kecantikan. Tahap terakhir adalah penghapusan *stopword* [15].

Selanjutnya, dilakukan proses *data preparation* untuk memastikan kualitas dataset sebelum digunakan dalam pelatihan model. Pada tahap ini dilakukan penanganan *outlier* panjang teks dengan menentukan batas maksimum panjang token menggunakan pendekatan persentil ke-95 (P95). Selain itu, dilakukan analisis terhadap ulasan dengan *rating* 3 (netral), karena ditemukan bahwa kondisi data dengan *rating* ini tidak merepresentasikan sentimen yang benar-benar netral. Analisis dilakukan menggunakan pendekatan *n-gram* dengan menghitung tingkat kemiripan (*overlap*) pola kata dengan ulasan positif dan negatif untuk menentukan apakah data tersebut akan dihapus atau digabungkan ke salah satu kelas sentimen. Berdasarkan hasil analisis tersebut, klasifikasi dalam penelitian ini difokuskan pada dua label sentimen, yaitu positif dan negatif. Dataset selanjutnya diseimbangkan menggunakan teknik *random undersampling*, yang kemudian dibagi menjadi data latih, data validasi, dan data uji dengan rasio 80:10:10.

## 2.3. Word Embedding dengan FastText

Setiap data teks diubah ke dalam bentuk vektor numerik menggunakan pendekatan *word embedding* berbasis *FastText*. Metode ini merepresentasikan kata sebagai vektor dengan memanfaatkan *subword* berbasis *n-gram* [16]. Model *FastText* dilatih menggunakan data latih sehingga dapat menghasilkan representasi kata yang menyesuaikan karakteristik bahasa pada ulasan produk kecantikan. Dimensi vektor yang digunakan sebesar 300 dengan arsitektur *skip-gram*, yaitu pendekatan yang memprediksi kata konteks berdasarkan satu kata pusat, sehingga mampu menghasilkan representasi yang lebih baik untuk kata yang jarang muncul [17].

## 2.4. Pembobotan dengan IDF

Setelah diperoleh representasi kata oleh *FastText*, dilakukan pembobotan kata menggunakan metode *Inverse Document Frequency (IDF)* untuk mengukur tingkat kepentingan setiap kata dalam korpus. Metode ini memberikan penekanan bobot yang lebih besar pada istilah yang kemunculannya relatif sedikit di keseluruhan kumpulan dokumen [10]. Perhitungan IDF dilakukan pada seluruh data latih sehingga diperoleh bobot untuk setiap kata. Bobot tersebut kemudian digunakan untuk membentuk *weighted word embedding* dengan mengalikan vektor kata hasil *FastText* dengan nilai IDF masing-masing kata, sehingga kata yang lebih informatif memiliki peranan yang lebih besar dalam representasi teks, yang selanjutnya digunakan sebagai input pada model Bi-LSTM.

Pada ulasan produk kecantikan, terdapat banyak kata opini, istilah khusus kecantikan, serta campuran bahasa asing yang memiliki kontribusi besar dalam menentukan polaritas sentimen, seperti "*oily*", "*breakout*", "*glowing*", atau "*cakey*". Sebaliknya, kata umum seperti "dan", "ini", atau "yang" cenderung kurang memberikan informasi terhadap sentimen ulasan. Penerapan pembobotan IDF diharapkan mampu menekankan kata-kata yang lebih informatif tersebut.

## 2.5. Klasifikasi dengan Bi-LSTM

*Bidirectional Long Short-Term Memory (Bi-LSTM)* menjadi turunan lanjutan dari RNN yang dirancang untuk mengatasi permasalahan *vanishing gradient* ketika memproses data sekuensial. Berbeda dengan LSTM biasa, Bi-LSTM bekerja dengan membaca urutan kata dari dua sisi, yaitu arah maju

(*forward*) dan arah balik (*backward*), sehingga konteks kalimat dapat dipahami secara lebih lengkap [2]. Pada penelitian ini, model Bi-LSTM berperan untuk mengklasifikasikan ulasan produk kecantikan ke dalam dua kelas sentimen, yaitu positif dan negatif. Arsitektur model yang dibangun dalam penelitian ini tersusun atas beberapa lapisan, yaitu :

- Input *layer*, yang berfungsi menerima *weighted FastText word embedding sequence* sebagai representasi numerik dari teks ulasan.
- Bi-LSTM *layer*, yang bertugas mempelajari pola urutan kata dari dua arah untuk menangkap konteks kalimat secara lebih menyeluruh (*forward LSTM* dan *backward LSTM*)
- Dense *layer*, yang digunakan untuk menggabungkan informasi hasil ekstraksi fitur dari *layer* sebelumnya sebelum dilakukan proses klasifikasi.
- Output *layer*, yang menerapkan aktivasi sigmoid agar model dapat menghasilkan estimasi probabilitas sentimen positif ataupun negatif.

Untuk mendapatkan model terbaik, dilakukan proses *hyperparameter tuning* dengan menerapkan metode *grid search*. Parameter yang diuji meliputi jumlah unit LSTM, nilai *dropout*, dan *learning rate*. Jumlah unit LSTM diuji pada rentang 32 hingga 128 unit untuk menentukan kapasitas model dalam mempelajari pola data [18]. Parameter *dropout* diuji pada rentang 0,2 hingga 0,5 untuk mengurangi kemungkinan *overfitting* selama proses pelatihan [4], sedangkan *learning rate* diuji pada rentang 0,0001 hingga 0,002 untuk menjaga kestabilan proses pembelajaran model [19]. Kombinasi parameter terbaik yang diperoleh kemudian digunakan untuk melatih model akhir sebelum dilakukan evaluasi pada data uji. Rentang uji *tuning* disajikan pada Tabel 1.

Tabel 1. Rentang Uji *Hyperparameter Tuning*

No	Parameter	Nilai
1.	Jumlah Unit/Neuron	[32, 64, 96, 128]
2.	Nilai <i>Dropout</i>	[0.2, 0.3, 0.4, 0.5]
3.	<i>Learning Rate</i>	[0.0001, 0.0004, 0.0008, 0.001, 0.002]

## 2.6. Evaluasi Model

Tahap evaluasi dimaksudkan untuk menganalisis performa model dalam membedakan kelas sentimen pada data ulasan. Metode yang diterapkan adalah *Confusion Matrix*, sebuah matriks yang merangkum hasil prediksi model ke dalam empat nilai dasar. Pertama, *True Positive* (TP) mencerminkan banyaknya sampel berlabel positif yang berhasil dikenali model secara benar. Kedua, *True Negative* (TN) menggambarkan sampel berlabel negatif yang berhasil diidentifikasi dengan tepat. Ketiga, *False Positive* (FP) merujuk pada sampel yang bersifat negatif, tetapi keliru diklasifikasikan sebagai positif. Keempat, *False Negative* (FN), yaitu sampel positif yang keliru diprediksi sebagai negatif. Keempat nilai tersebut menjadi dasar perhitungan empat metrik evaluasi utama. *Accuracy* digunakan untuk melihat proporsi keseluruhan prediksi yang sesuai dengan label aslinya. *Precision* berfokus pada seberapa andal model menentukan suatu data sebagai kelas positif. *Recall* mengukur sejauh mana model mampu menjangkau seluruh sampel yang memang termasuk kelas positif. *F1-Score* menggabungkan nilai *precision* dan *recall* yang mencerminkan keseimbangan performa model.

## 2.7. Pengembangan Sistem

Model terbaik yang diperoleh dari proses pelatihan kemudian diimplementasikan ke dalam sistem analisis sentimen ulasan produk kecantikan otomatis berbasis web. Sistem dirancang melalui pendekatan *Waterfall* yang bersifat sekuensial, meliputi proses identifikasi kebutuhan, perancangan sistem, pengembangan aplikasi, pengujian, sampai tahap perawatan sistem [20]. Bahasa pemrograman Python beserta *framework* Flask dan basis data SQL digunakan sebagai teknologi implementasi. Pengujian sistem diterapkan melalui *Black Box Testing* dengan pendekatan *Equivalence Partitioning*, yaitu data input dikelompokkan untuk memverifikasi kesesuaian fungsi sistem dengan kebutuhan yang telah ditetapkan [21].

## 3. Hasil dan Pembahasan

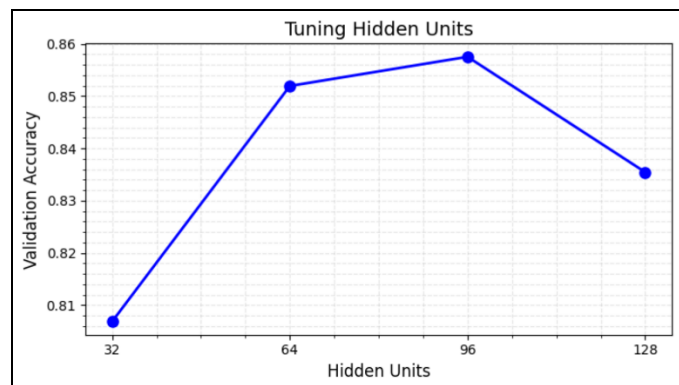
### 3.1. Hasil *Data Preparation*

Pada tahap penghapusan data *outlier*, ulasan dengan panjang teks lebih dari batas persentil ke-95 dihapus sehingga jumlah data berkurang dari 50.000 menjadi 47.618 ulasan. Selanjutnya, dilakukan

analisis terhadap ulasan *rating* 3 menggunakan pendekatan *n-gram* dengan menghitung tingkat *overlap* pola kata terhadap kelas positif dan negatif, yang menunjukkan kecenderungan sentimen positif sebesar 78,25%, sehingga *ulasan rating* 3 digabungkan ke dalam kelas positif. Setelah penggabungan label, jumlah data positif menjadi 38.465 dan negatif 9.153, kemudian dilakukan *random undersampling* ke kelas minoritas (kelas negatif) untuk menyeimbangkan data sehingga masing-masing kelas berjumlah 9.153 ulasan. Dataset yang telah seimbang kemudian dibagi menjadi data latih (14.644), data validasi (1.831), dan data uji (1.831) dengan rasio 80:10:10.

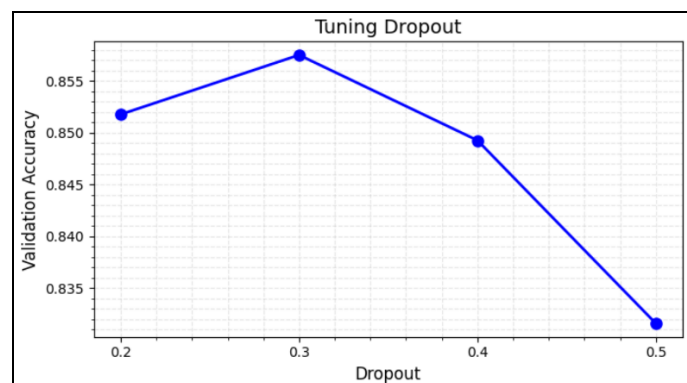
### 3.2. Hasil Hyperparameter Tuning

Tahap *hyperparameter tuning* bertujuan untuk menentukan kombinasi konfigurasi terbaik pada arsitektur Bi-LSTM yang diterapkan. Parameter yang diuji meliputi jumlah unit LSTM, nilai *dropout*, dan *learning rate* dengan rentang nilai yang ditampilkan pada Tabel 1. Setiap konfigurasi model dilatih menggunakan data pelatihan dan dievaluasi menggunakan data validasi untuk memperoleh nilai *validation accuracy* terbaik.



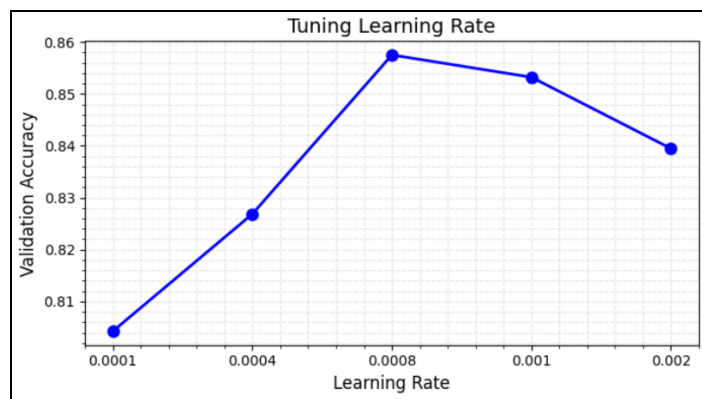
Gambar 2. Grafik Pengaruh Jumlah Unit terhadap *Validation Accuracy*

Berdasarkan grafik pada Gambar 2, variasi jumlah unit LSTM menunjukkan bahwa peningkatan jumlah unit dari 32 hingga 96 menghasilkan peningkatan performa model, yaitu dari 80,6% hingga mencapai puncak pada 96 unit dengan akurasi 85,7%. Akan tetapi, ketika jumlah unit ditingkatkan menjadi 128, akurasi model mengalami penurunan. Hal ini menunjukkan bahwa jumlah unit yang terlalu besar dapat membuat model menjadi terlalu kompleks sehingga menyebabkan performa model berkurang.



Gambar 3. Grafik Pengaruh *Dropout* terhadap *Validation Accuracy*

Pada Gambar 3, hasil *tuning* nilai *dropout* menunjukkan bahwa nilai *dropout* 0,3 menghasilkan performa terbaik, yaitu naik dari 85,1% menjadi 85,7%. Nilai tersebut memberikan tingkat regularisasi yang cukup sehingga model dapat mengurangi potensi *overfitting* selama proses pelatihan. Sementara itu, nilai *dropout* yang lebih besar menyebabkan penurunan akurasi pada data validasi. Hal ini dapat terjadi karena terlalu banyak neuron yang dinonaktifkan selama proses pelatihan.



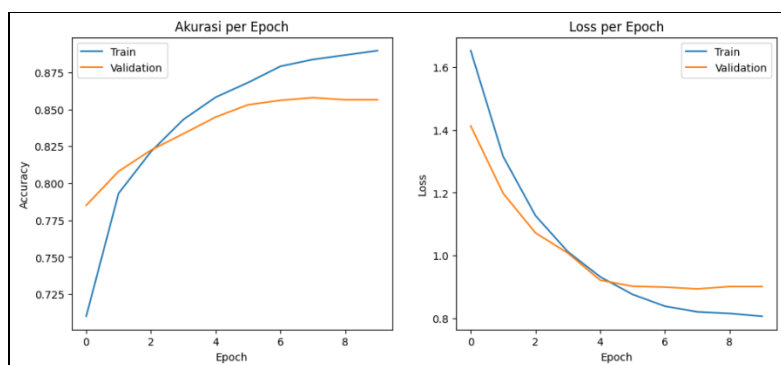
Gambar 4. Grafik Pengaruh *Learning Rate* terhadap *Validation Accuracy*

Sementara itu pada Gambar 4, pengujian *learning rate* menunjukkan bahwa nilai 0.0008 memberikan performa paling optimal dengan *validation accuracy* sebesar 85,84%, sedangkan nilai yang terlalu kecil atau terlalu besar menghasilkan akurasi yang lebih rendah. Berdasarkan keseluruhan proses tuning yang dilakukan, diperoleh kombinasi *hyperparameter* terbaik yaitu jumlah unit LSTM sebesar 96, nilai *dropout* sebesar 0.3, dan *learning rate* sebesar 0.0008. Kombinasi parameter ini kemudian digunakan pada tahap pelatihan model akhir untuk menghasilkan model klasifikasi yang optimal.

### 3.3. Hasil Pelatihan Model

Proses pelatihan menggunakan data pelatihan dan data validasi dengan *batch size* 32, maksimum 20 *epoch*, serta mekanisme *early stopping*. Selama pelatihan, sistem menyimpan bobot model terbaik berdasarkan nilai *validation loss* dan akan mengembalikannya setelah proses pelatihan selesai. Berdasarkan hasil pelatihan, proses pelatihan berhenti pada *epoch* ke-10 karena tidak terjadi penurunan *validation loss* pada *epoch* berikutnya. Performa terbaik model diperoleh pada *epoch* ke-7 dengan nilai *training accuracy* sebesar 88,41% dan *validation accuracy* sebesar 85,64%.

Grafik perkembangan *accuracy* dan *loss* model selama proses pelatihan ditunjukkan pada Gambar 5. Grafik tersebut memperlihatkan bahwa nilai *training accuracy* dan *validation accuracy* meningkat secara bertahap hingga mencapai performa terbaik pada sekitar *epoch* ke-6 hingga ke-7, sementara nilai *training loss* dan *validation loss* menunjukkan tren yang terus menurun sepanjang proses pelatihan. Pola yang ditandai dengan naiknya *accuracy* seiring berkurangnya *loss* pada data latih dan validasi, mengindikasikan bahwa model menangkap dan mempelajari pola yang terdapat dalam data.



Gambar 5. Kurva Perubahan *Accuracy* dan *Loss* Model pada Setiap *Epoch*

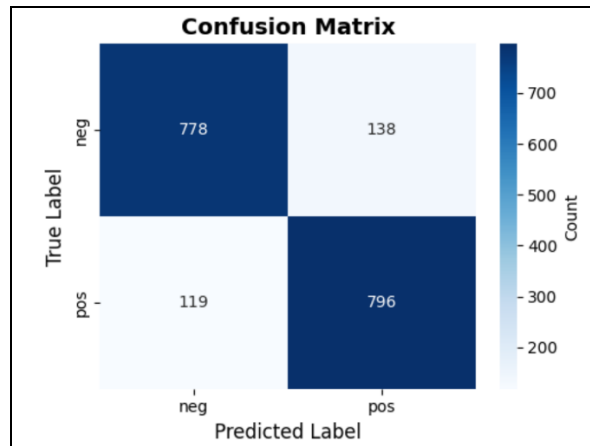
### 3.4. Hasil Evaluasi Model

Evaluasi terhadap performa model dilakukan pada tahap akhir dengan menggunakan data uji, yaitu data baru yang tidak terlibat selama tahap pelatihan model. Evaluasi dilakukan menggunakan *classification report* untuk memperoleh nilai empat metrik evaluasi. Hasil evaluasi menunjukkan model memperoleh *testing accuracy* sebesar 85,96%, yang tidak jauh berbeda dengan *validation accuracy* sebesar 85,64% pada proses pelatihan, kondisi ini memperlihatkan bahwa model mampu mempertahankan kinerja prediksi pada data di luar proses pelatihan.

**Tabel 2.** Hasil *Classification Report* Model pada Data Pengujian

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	0,8673	0,8493	0,8582	916
Positif	0,8522	0,8699	0,8610	915
<b><i>Accuracy</i></b>			<b>0,8596</b>	1.831

Hasil *classification report* pada Tabel 2 memperlihatkan bahwa model mengklasifikasikan sentimen negatif dengan *precision* sebesar 86,73%, *recall* 84,93%, dan *f1-score* 85,82%. Sementara itu, pada kelas positif model meraih *precision* sebesar 85,22%, *recall* 86,99%, dan *f1-score* 86,10%. Nilai *precision* dan *recall* pada kedua kelas relatif seimbang, yang menjadi indikasi bahwa model mampu mengidentifikasi karakteristik dari masing-masing kelas sentimen.



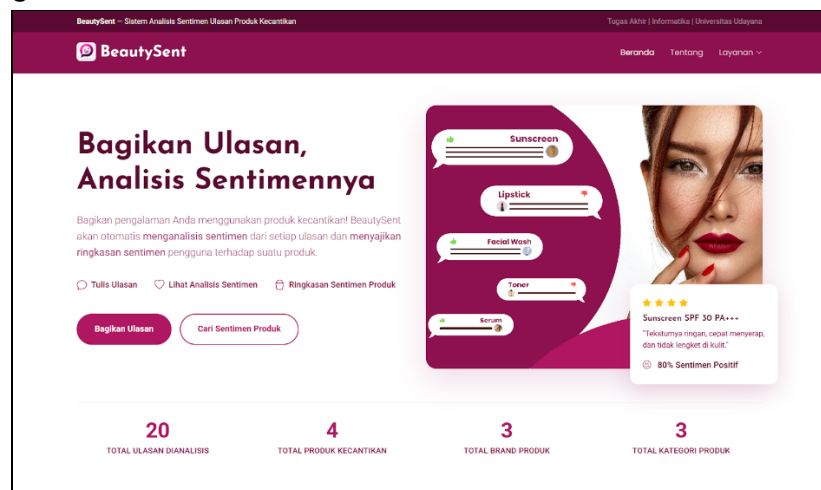
**Gambar 6.** *Confusion Matrix*

Selain itu, hasil evaluasi juga ditampilkan dalam bentuk *confusion matrix* pada Gambar 6 untuk melihat distribusi prediksi model secara lebih rinci. Berdasarkan *confusion matrix* tersebut, model berhasil mengklasifikasikan 778 data sentimen negatif dengan benar sebagai negatif dan 796 data sentimen positif dengan benar sebagai positif. Akan tetapi masih terdapat kesalahan prediksi, yaitu 138 data sentimen negatif yang diprediksi sebagai positif serta 119 data sentimen positif yang diprediksi sebagai negatif. Hasil tersebut memperlihatkan bahwa model sudah memiliki kemampuan klasifikasi yang baik terhadap mayoritas data.

### 3.5. Hasil Implementasi Sistem

Model analisis sentimen yang telah dilatih pada tahap sebelumnya kemudian diintegrasikan ke dalam sistem berbasis *website* untuk melakukan prediksi sentimen secara otomatis. Integrasi dilakukan dengan memanfaatkan model terbaik hasil pelatihan.

#### a. Landing Page



**Gambar 7.** Halaman *Landing Page* Sistem

Tampilan pada Gambar 7 berperan sebagai tampilan awal yang langsung terlihat oleh pengguna saat memasuki *website*. Halaman tersebut berisi informasi umum terkait sistem serta navigasi menuju fitur utama yang tersedia. Selain itu, ditampilkan pula ringkasan informasi data yang terdapat dalam sistem, seperti jumlah total ulasan, jumlah produk, jumlah *brand*, dan jumlah kategori produk. Informasi tersebut memberikan gambaran awal kepada pengguna mengenai data ulasan produk kecantikan yang tersedia dalam sistem.

## b. Fitur Analisis Sentimen Ulasan

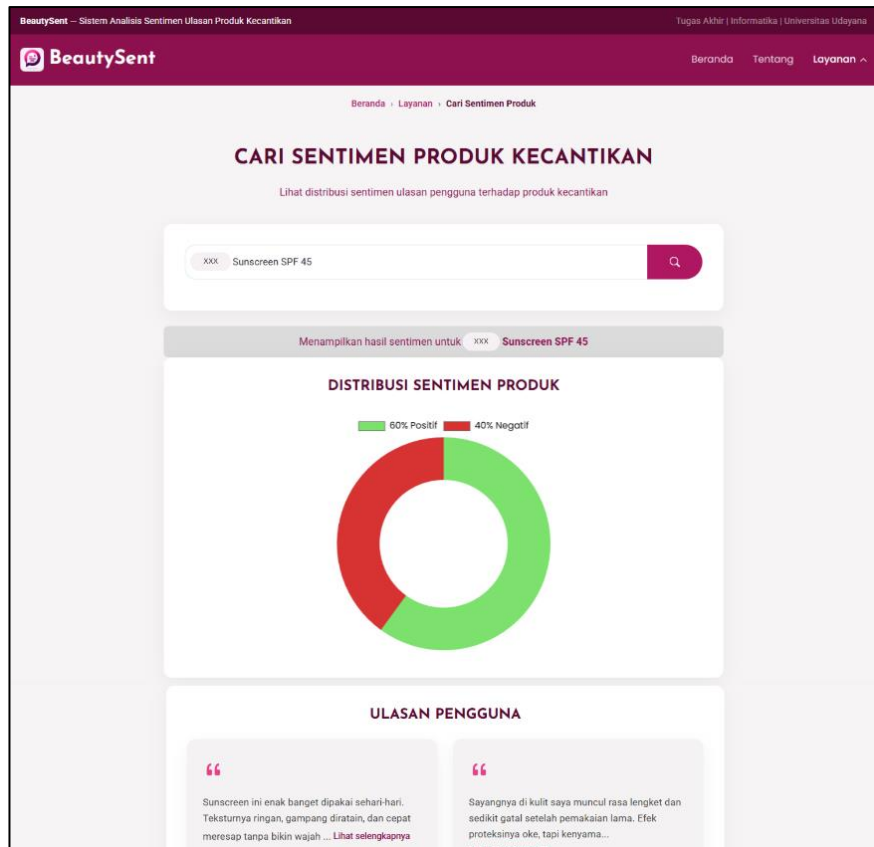
The screenshot shows the 'BeautySent' web application interface for sentiment analysis. The page has a purple header with the logo and navigation links. The main content area is titled 'BAGIKAN ULASAN DAN ANALISIS SENTIMEN'. Below the title, there is a form with the following fields: 'Nama Produk' (input field with 'xxx Sunscreen 30SPF'), 'Brand' (input field with 'xxx'), and 'Kategori Produk' (dropdown menu with 'Sun Protection'). A text area for 'Ulasan Produk' contains a sample review: 'Wah ini sunscreen bikin muka aku kayak minyak goreng habis dipakai goreng. Baru sekitar 10 menit dipakai aja wajah sudah kelihatan berminyak parah. Teksturnya juga terasa agak berat dan lama meresap di kulit. Wanginya pun aneh banget, mirip bau obat nyamuk jadi agak ganggu waktu dipakai. Setelah beberapa jam, muka makin kelihatan kusam dan nggak nyaman. Kayaknya habis ini bakal skip dan cari sunscreen lain yang lebih cocok.' Below the form is a purple 'Analisis' button. At the bottom of the form, there are two purple boxes: 'HASIL PREDIKSI SENTIMEN' with the result 'Negatif' and 'KEYAKINAN MODEL' with the result '90%'.

**Gambar 8.** Halaman Fitur Analisis Sentimen Ulasan

Gambar 8 menunjukkan fitur analisis sentimen ulasan yang dimanfaatkan untuk mengidentifikasi sentimen dari teks yang dimasukkan oleh pengguna. Pada halaman ini, pengguna dapat mengisi form yang tersedia berupa nama produk, *brand* produk, kategori produk, serta teks ulasan yang akan dianalisis. Sistem akan memproses teks ulasan menggunakan model analisis sentimen yang telah diintegrasikan dan menampilkan hasil prediksi dalam bentuk klasifikasi sentimen positif atau negatif beserta nilai persentase keyakinan model. Selain itu, ulasan yang telah dianalisis disimpan pada *database* untuk digunakan pada fitur ringkasan sentimen produk.

## c. Fitur Ringkasan Sentimen Produk

Fitur ringkasan sentimen produk (Gambar 9) digunakan untuk menampilkan rangkuman hasil analisis sentimen berdasarkan ulasan yang telah tersimpan dalam *database*. Pada halaman ini, pengguna dapat mencari produk tertentu melalui fitur pencarian yang tersedia. Sistem kemudian menampilkan ringkasan sentimen produk dalam bentuk persentase menggunakan visualisasi *doughnut chart*. Selain itu, sistem juga menampilkan contoh ulasan-ulasan pengguna lain terkait produk tersebut.



**Gambar 9.** Halaman Ringkasan Sentimen Produk

**d. Pengujian Sistem**

Pengujian difokuskan pada fitur utama sistem, yaitu fitur analisis sentimen ulasan dan fitur ringkasan sentimen produk. Hasil pengujian sistem disajikan pada Tabel 3. Berdasarkan hasil pengujian tersebut, seluruh fitur utama sistem berfungsi sesuai dengan rancangan yang telah ditetapkan pada tahap analisis kebutuhan, serta menghasilkan keluaran yang diharapkan. Sistem mampu memproses berbagai jenis input yang diberikan oleh pengguna serta menampilkan hasil analisis sentimen dan ringkasan sentimen produk secara tepat.

Selain itu, pengujian menggunakan data *ground truth* dilakukan untuk memastikan perilaku model setelah implementasi pada web tetap konsisten dengan tahap pelatihan. Hasil pengujian menunjukkan bahwa keluaran prediksi yang dihasilkan sistem web tetap sama dengan hasil prediksi pada tahap pelatihan, sehingga implementasi model pada *website* dapat berjalan dengan baik dan sesuai harapan.

**Tabel 3.** Hasil Pengujian *Black Box* Sistem

No	Fitur Sistem	Skenario Pengujian	Input	Harapan Output	Hasil
1.	Analisis Sentimen	Pengguna memasukkan ulasan dengan sentimen positif	Teks ulasan positif	Sistem menampilkan hasil sentimen positif	Berhasil
2.	Analisis Sentimen	Pengguna memasukkan ulasan dengan sentimen negatif	Teks ulasan negatif	Sistem menampilkan hasil sentimen negatif	Berhasil
3.	Analisis Sentimen	Pengguna memasukkan teks ulasan dengan karakter khusus	Ulasan dengan tanda baca/URL	Sistem tetap dapat memproses ulasan dan menampilkan hasil prediksi	Berhasil
4.	Analisis Sentimen	Sistem menampilkan probabilitas/persentase keyakinan prediksi	Teks ulasan	Sistem menampilkan probabilitas/	Berhasil

				persentase keyakinan prediksi	
5.	Analisis Sentimen	Sistem menyimpan hasil analisis ulasan	Nama produk dan ulasan	Data ulasan tersimpan pada <i>database</i>	Berhasil
6.	Analisis Sentimen	Pengguna tidak mengisi teks ulasan atau nama produk	Input kosong	Sistem menampilkan pesan bahwa ulasan atau nama produk tidak boleh kosong	Berhasil
7.	Ringkasan Sentimen Produk	Pengguna mencari produk yang tersedia di <i>database</i>	Nama produk yang ada	Sistem menampilkan persentase sentimen positif dan negatif serta daftar ulasan	Berhasil
8.	Ringkasan Sentimen Produk	Pengguna mencari produk yang tidak tersedia	Nama produk tidak ada	Sistem menampilkan pesan bahwa produk tidak ditemukan	Berhasil

#### 4. Kesimpulan

Berdasarkan temuan yang dihasilkan dari penelitian ini, model Bi-LSTM dengan representasi fitur IDF *weighted FastText word embedding* mampu meningkatkan representasi kata-kata informatif pada ulasan produk kecantikan berbahasa Indonesia yang mengandung bahasa tidak baku, singkatan, dan istilah khusus kecantikan, sehingga menghasilkan performa *accuracy* terbaik sebesar 85,96% pada data pengujian dengan nilai *precision*, *recall*, dan *f1-score* yang relatif seimbang pada kedua kelas sentimen. Proses *hyperparameter tuning* menunjukkan bahwa kombinasi 96 unit LSTM, *dropout* 0,3, dan *learning rate* 0,0008 menghasilkan performa pelatihan terbaik dengan *validation accuracy* sebesar 85,64%. Model terbaik tersebut kemudian diimplementasikan pada sistem berbasis web yang menyediakan fitur analisis sentimen ulasan dan ringkasan sentimen produk. Hasil pengujian sistem mengindikasikan bahwa seluruh fungsi dapat berjalan sesuai dengan kebutuhan fungsional yang telah ditentukan. Penelitian selanjutnya disarankan untuk membandingkan model dengan arsitektur yang lebih modern seperti CNN-LSTM, GRU, BERT, atau IndoBERT, khususnya dengan menerapkan *attention mechanism* atau model transformer untuk menangani data teks ulasan yang mengandung *mixed sentiment*.

#### Referensi

- [1] Kemenko Perekonomian RI, "Tumbuh Double Digit, Pemerintah Dukung Penguatan Ekosistem Industri Kecantikan", 4 Desember 2025. [Online]. Available: <https://www.ekon.go.id/publikasi/detail/6712/tumbuh-double-digit-pemerintah-dukung-penguatan-ekosistem-industri-kecantikan>. [1 Januari 2026]
- [2] Kemenko Perekonomian RI, "Hasilkan Produk Berdaya Saing Global, Industri Kosmetik Nasional Mampu Tembus Pasar Ekspor dan Turut Mendukung Penguatan Blue Economy", 3 Februari 2024. [Online]. Available: <https://www.ekon.go.id/publikasi/detail/5626/hasilkan-produk-berdaya-saing-global-industri-kosmetik-nasional-mampu-tembus-pasar-ekspor-dan-turut-mendukung->. [1 Januari 2026]
- [3] S. Siami-Namini, N. Tavakoli and A. S. Namin, "The Performance of LSTM and BiLSTM in Forecasting Time Series" in *IEEE International Conference on Big Data (Big Data)*, Los Angeles, 2019, pp. 3285-3292.
- [4] A. Rolangon, A. Weku and G. A. Sandag, "Perbandingan Algoritma LSTM Untuk Analisis Sentimen Pengguna Twitter Terhadap Layanan Rumah Sakit Saat Pandemi Covid-19" *Jurnal Teknologi Informasi dan Komunikasi (TelKa)*, vol. 13, no. 1, p. 31-40, 2023.
- [5] H. Imaduddin, F.Y. A'la and Y.S. Nugroho, "Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach" *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, p. 113-117, 2023.
- [6] V.G. Utomo, "Benchmarking IndoBERT and Transformer Models for Sentiment Classification on Indonesian E-Government Service Reviews" *Jurnal Transformatika*, vol. 23, no. 1, p. 86-95, 2025.
- [7] T. Indriyatmoko, M. Rahardi, H. Utama and A.C. Frobenius, "Comparative Analysis of BERT and LSTM Models for Sentiment Classification of Mobile Game User Reviews" *Journal of Applied Informatics and Computing*, vol. 10, no. 1, p.1093-1100, 2026.

- [8] M. A. Riza and N. Charibaldi, "Emotion Detection in Twitter Social Media Using Long Short-Term Memory (LSTM) and Fast Text." *International Journal of Artificial Intelligence & Robotics (IJAIR)*, vol.3, no. 1, p.15-26, 2021.
- [9] A. Nurdin, B. A. S. Aji, A. Bustamin and Z. Abidin, "Perbandingan Kinerja Word Embedding Word2Vec, Glove, Dan Fasttext Pada Klasifikasi Teks" *Jurnal Tekno Kompak*, vol. 14, no. 2, p. 74-79, 2020.
- [10] W. Zhao, L. Zhu, M. Wang, X. Zhang and J. Zhang, " WTL-CNN: A News Text Classification Method of Convolutional Neural Network Based on Weighted Word Embedding" *Connection Science*, vol. 34, no. 1, p.2291-2312, 2022.
- [11] T.A. Fitri, "Optimization of Deep Learning with FastText for Sentiment Analysis of the SIREKAP 2024 Application" *The Indonesian Journal of Computer Science*, vol. 14, no. 2, p. 2414-2429, 2025.
- [12] D. Ariyus, D. Manongga and I. Sembiring, "Enhancing Sentiment Analysis of Indonesian Tourism Video Content Commentary on TikTok: A FastText and Bi-LSTM Approach" *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, p.18020-18028, 2024.
- [13] A. Onan, "Sentiment Analysis on Product Reviews Based on Weighted Word Embeddings and Deep Neural Networks" *Concurrency and Computation Practice and Experience*, vol. 33, no. 23, p. 1-12, 2021.
- [14] A. Romadhony, S. Al Faraby, R. Rismala, U. N. Wisesti and A. Arifianto, " Sentiment Analysis on a Large Indonesian Product Review Dataset" *Journal of Information Systems Engineering & Business Intelligence*, vol. 10, no. 1, p.167-178, 2024.
- [15] L. Hickman, S. Thapa, L. Tay, M. Cao and P. Srinivasan, " Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations" *Organizational Research Methods*, vol. 25, no. 1, p. 114-146, 2022.
- [16] M. D. Rahman, A. Djunaidy and F. Mahananto, " Penerapan Weighted Word Embedding pada Pengklasifikasian Teks Berbasis Recurrent Neural Network untuk Layanan Pengaduan Perusahaan Transportasi" *Jurnal Sains dan Seni ITS*, vol. 10, no. 1, p. A1-A6, 2021.
- [17] V. Major, A. Surkis and Y. Aphinyanaphongs, " Utility of general and specific word embeddings for classifying translational stages of research" in *AMIA annual symposium proceedings*, City, 2018, pp. 1405-1414.
- [18] A. Sinaga, Y. Sibaroni and S. Prasetyowati, " Multi-Aspect Sentiment Analysis of Movie Reviews Using BiLSTM on Platform X Data" *Building of Informatics, Technology and Science (BITS)*, vol. 7, no. 1, p. 672-684, 2025.
- [19] I. N. Yulita, A. F. Al-Auza'i, A. S. Prabuwo, A. Sholahuddin, F. Ardiansyah, I. Sarathan and Y. Djuyandi, " Bidirectional Long Short-Term Memory for Analysis of Public Opinion Sentiment on Government Policy During the COVID-19 Pandemic" *International Journal of Advanced Computer Science & Applications*, vol. 14, no. 11, p. 877-885, 2023.
- [20] T. Thesing, C. Feldmann and M. Burchardt, " Agile Versus Waterfall Project Management: Decision Model for Selecting The Appropriate Approach to a Project" *Procedia Computer Science*, vol. 181, p. 746-756, 2021.
- [21] M. Sholeh, I. Gisfas and M. A. Fauzi, " Black Box Testing on ukmbantul. com Page with Boundary Value Analysis and Equivalence Partitioning Methods" *Journal of Physics: Conference Series*, vol. 1823, no. 1, p.012029, 2021.