Jurnal Elektronik Ilmu Komputer Udayana
Volume 14, No 4. May 2026

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Penggunaan LSTM dan Fast-Text dalam Analisis Sentimen Berbasis Aspek pada Ulasan Aplikasi dengan Seleksi Fitur Information Gain

Maedelien Tiffany Kariesta Simatupang[a1], I Made Widiartha[a2], Luh Gede Astuti[b3], I Gede Santi Astawa [b4]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Udayana, Indonesia
[1]simatupang.2208561065@unud.ac.id
[2]madewidiartha@unud.ac.id
[3]lg.astuti@unud.ac.id
[4]santi.astawa@unud.ac.id

Abstract

This study aims to perform aspect-based sentiment analysis on user reviews of the GoTube application from the Google Play Store. The large volume of user reviews makes manual analysis inefficient, requiring an automated approach to extract meaningful insights. Data were collected through web scraping, resulting in over 280,000 reviews, which were further processed through cleaning, labeling, and data balancing. To improve label consistency, a pseudo-labeling approach using IndoBERT was applied. The proposed method combines Information Gain for feature selection, FastText for word representation, and Long Short-Term Memory (LSTM) for sentiment classification. In addition, Latent Dirichlet Allocation (LDA) was used for aspect extraction. The experimental results show that the sentiment classification model achieved an accuracy of 95.37% and F1-score of 0.9537 using an optimal threshold of 0.59, with balanced precision (0.9539) and recall (0.9534). Meanwhile, the aspect classification model achieved an accuracy of 87.12% with a macro F1-score of 0.8697. These findings indicate that the combination of feature selection, subword-based representation, and sequential modeling is effective in producing accurate and informative aspect-based sentiment analysis.

Keywords: *sentiment analysis, aspect-based sentiment analysis, LSTM, FastText, information gain*

1. Pendahuluan

Perkembangan teknologi internet telah meningkatkan konsumsi konten digital, khususnya video. Media sosial tidak hanya berfungsi sebagai sarana komunikasi, tetapi juga sebagai ruang untuk berbagi informasi, hiburan, dan opini. Di Indonesia, penggunaan media sosial telah menjadi bagian dari aktivitas sehari-hari [1], sehingga mendorong munculnya berbagai aplikasi berbasis video dengan fitur yang semakin beragam. Fenomena ini turut mendorong pengguna untuk lebih aktif memberikan ulasan terhadap aplikasi yang mereka gunakan, terutama melalui platform distribusi aplikasi seperti Google Play Store.

Salah satu aplikasi tersebut adalah GoTube, sebuah aplikasi pemutar video yang menawarkan pengalaman menonton tanpa iklan serta fitur tambahan seperti mode malam, pemutaran latar belakang, dan pemutaran pop-up. Seiring dengan meningkatnya jumlah pengguna, jumlah ulasan pada Google Play Store juga semakin besar. Ulasan tersebut mengandung opini, kritik, dan saran yang penting bagi pengembang, namun sulit dianalisis secara manual karena volumenya yang tinggi. Selain itu, teks ulasan sering bersifat informal, mengandung variasi penulisan, singkatan, serta kesalahan penulisan, dan dalam satu ulasan pengguna dapat membahas beberapa aspek dengan sentimen yang berbeda-beda.

Untuk menganalisis sentimen secara otomatis, digunakan pendekatan *Natural Language Processing* (NLP), khususnya analisis sentimen. Namun, analisis sentimen konvensional umumnya hanya memberikan polaritas secara keseluruhan tanpa mempertimbangkan aspek spesifik yang dibahas, sehingga informasi yang dihasilkan kurang dapat ditindaklanjuti secara praktis oleh pengembang

aplikasi. Oleh karena itu, diperlukan pendekatan *Aspect-Based Sentiment Analysis* (ABSA) yang mampu menganalisis sentimen secara lebih rinci berdasarkan aspek tertentu [2].

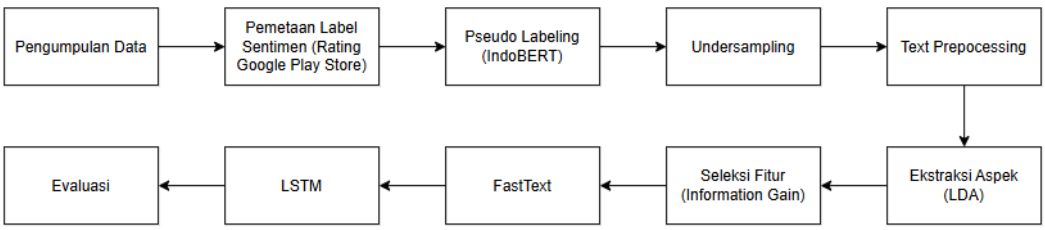
Meskipun penelitian ABSA telah cukup berkembang, sejumlah kelemahan mendasar masih ditemukan pada pendekatan yang ada. Pertama, banyak penelitian sebelumnya mengandalkan rating bintang sebagai sumber label utama [3][4], padahal inkonsistensi antara rating dan isi teks ulasan merupakan permasalahan yang umum terjadi — pengguna terkadang memberikan rating tinggi disertai kritik, atau sebaliknya. Ketergantungan pada rating tanpa validasi mengakibatkan data pelatihan yang bias dan menurunkan kualitas model. Kedua, sebagian besar penelitian ABSA pada ulasan berbahasa Indonesia belum menangani permasalahan variasi lekskal secara memadai, karena menggunakan metode *word embedding* berbasis token penuh seperti Word2Vec atau GloVe yang tidak mampu merepresentasikan kata-kata tidak baku, singkatan, dan kesalahan ejaan yang dominan pada teks informal [5][6]. Ketiga, penelitian yang menggabungkan seleksi fitur berbasis kontribusi informatif dengan representasi subword dalam satu kerangka ABSA masih sangat terbatas, padahal dimensi fitur yang tinggi tanpa seleksi dapat menurunkan efisiensi dan generalisasi model [7]. Keempat, ekstraksi aspek dalam penelitian sebelumnya sering dilakukan secara manual atau hanya berdasarkan kategori yang ditentukan a priori, tanpa memanfaatkan pendekatan berbasis topik yang dapat menemukan aspek secara data-driven dari ulasan itu sendiri [8].

Untuk mengatasi kelemahan-kelemahan tersebut, penelitian ini mengusulkan kerangka ABSA yang mengintegrasikan empat komponen secara terpadu. Validasi label berbasis *pseudo-labeling* diterapkan untuk menjamin konsistensi anotasi sentimen dan mengurangi bias dari rating bintang. FastText dengan representasi *subword* digunakan sebagai metode *word embedding* yang secara khusus lebih tangguh terhadap variasi penulisan informal dibandingkan Word2Vec dan GloVe [6]. Information Gain diterapkan sebagai tahap seleksi fitur untuk mereduksi dimensi sekaligus meningkatkan kontribusi fitur yang relevan terhadap klasifikasi [7]. LDA digunakan untuk mengekstraksi aspek secara *data-driven* dari korpus ulasan, sehingga aspek yang diperoleh mencerminkan topik nyata yang dibahas pengguna [8]. Keempat komponen tersebut diintegrasikan dengan model LSTM yang terbukti efektif dalam menangkap dependensi sekuensial pada teks [4]. Sistem kemudian diimplementasikan sebagai aplikasi web dan diuji menggunakan metode *black box testing*.

2. Metode Penelitian

2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan komputasional untuk melakukan analisis sentimen berbasis aspek pada ulasan pengguna aplikasi GoTube. Data dikumpulkan melalui scraping dari Google Play Store, kemudian dilakukan pembersihan awal untuk menghapus duplikat, nilai kosong, dan ulasan yang hanya berisi emoji. Alur keseluruhan penelitian ditunjukkan pada Gambar 1.



Gambar 1 Diagram Alur Penelitian

2.2. Pengumpulan dan Validasi Data

Data dikumpulkan dari ulasan pengguna aplikasi GoTube pada Google Play Store melalui proses web scraping menggunakan library google-play-scraper. Pengambilan data dibatasi pada rentang tahun 2021 hingga 2024 untuk menjaga relevansi terhadap kondisi aplikasi terbaru. Setiap data terdiri dari tiga atribut utama, yaitu tanggal ulasan, skor bintang (1–5), dan teks ulasan dalam bahasa Indonesia. Label sentimen awal ditentukan berdasarkan skor bintang, dengan ketentuan rating 1–2 dikategorikan sebagai sentimen negatif dan rating 4–5 sebagai sentimen positif, sedangkan rating 3 tidak digunakan karena bersifat ambigu dan tidak menunjukkan polaritas sentimen yang kuat [9]. Sebelum proses

validasi label, dilakukan preprocessing ringan yang meliputi case folding, penghapusan URL, mention, emoji, serta karakter non-alfabet. Validasi label dilakukan menggunakan pendekatan pseudo-labeling dengan model IndoBERT. Label dari skor bintang dibandingkan dengan hasil prediksi model, dan hanya data dengan label yang konsisten di antara keduanya yang dipertahankan untuk tahap selanjutnya [11]. Pendekatan ini bertujuan untuk mengurangi kesalahan pelabelan akibat ketidaksesuaian antara rating dan isi ulasan.

2.3. Undersampling

Setelah proses validasi label, dilakukan teknik random undersampling untuk menyeimbangkan distribusi kelas dalam dataset. Metode ini bekerja dengan cara memilih secara acak sejumlah sampel dari setiap kelas sentimen tanpa mempertimbangkan karakteristik atau atribut tertentu dari data [10]. Proses ini bertujuan agar model tidak bias terhadap kelas mayoritas yang dapat menyebabkan model cenderung memprediksi kelas dominan saja. Dari masing-masing kelas diambil sebanyak 10.000 ulasan secara acak, sehingga dataset akhir yang digunakan terdiri dari 20.000 ulasan dengan distribusi kelas yang seimbang, yaitu 10.000 ulasan untuk kelas positif dan 10.000 ulasan untuk kelas negatif.

2.4. Text Prepocessing

Setelah dataset tervalidasi dan memiliki distribusi kelas yang seimbang, dilakukan tahap text preprocessing utama untuk menghasilkan representasi teks yang lebih terstruktur dan konsisten. Tahapan ini meliputi beberapa proses sebagai berikut: (1) Case folding: menyeragamkan seluruh huruf menjadi huruf kecil untuk mengurangi variasi penulisan yang tidak perlu. (2) Normalisasi karakter berulang: mengganti karakter yang diulang berlebihan, misalnya 'mantaaap' menjadi 'mantap'. (3) Konversi emotikon: mengubah emotikon ke dalam bentuk teks yang bermakna agar informasi sentimen dari emotikon tidak hilang. (4) Cleaning: menghapus URL, angka, karakter non-alfabet, dan karakter khusus yang tidak membawa informasi semantik. (5) Normalisasi kata tidak baku: mengubah kata-kata informal dan singkatan ke bentuk baku menggunakan kamus normalisasi, misalnya 'gk' menjadi 'tidak', 'bgt' menjadi 'banget'. (6) Penanganan negasi: menggabungkan kata negasi dengan kata berikutnya menggunakan tanda underscore, misalnya 'tidak ada' menjadi 'tidak_ada', agar makna sentimen negatif tidak hilang saat stopword removal. (7) Stopword removal: menghapus kata-kata yang tidak membawa makna sentimen secara selektif, dengan tetap mempertahankan kata negasi dan kata bermakna opini. (8) Stemming: mengubah kata ke bentuk dasar menggunakan algoritma Sastrawi untuk bahasa Indonesia, sehingga mengurangi variasi morfologis dalam teks [12].

2.5. Latent Dirichlet Allocation (LDA)

Ekstraksi aspek dilakukan menggunakan Latent Dirichlet Allocation (LDA), yaitu metode probabilistik topic modeling yang mengelompokkan dokumen berdasarkan distribusi topik laten [7]. Secara teoritis, LDA mengasumsikan bahwa setiap dokumen merupakan campuran dari beberapa topik, dan setiap topik merupakan distribusi probabilistik atas kata-kata dalam kosakata. Model diterapkan pada data latih yang telah melalui tahap preprocessing dengan terlebih dahulu membentuk bigram untuk menangkap frasa dua kata yang bermakna, kemudian membangun dictionary dan corpus berbasis bag-of-words. Jumlah topik optimal ditentukan menggunakan coherence score dengan metrik c_v pada rentang K = 3 hingga K = 5, di mana nilai tertinggi menunjukkan kualitas topik yang lebih baik secara semantik [13]. Setiap ulasan kemudian diberi label aspek berdasarkan topik dengan probabilitas tertinggi (dominant topic), dan interpretasi aspek dilakukan secara manual berdasarkan kata-kata dominan pada setiap topik.

2.6. Information Gain

Seleksi fitur dilakukan menggunakan Information Gain (IG) untuk memilih kata yang paling relevan dalam membedakan kelas sentimen. Metode ini mengukur seberapa besar pengurangan entropi (ketidakpastian) dari label kelas setelah keberadaan suatu kata dalam dokumen diketahui [14]. Nilai IG yang tinggi menunjukkan bahwa kata tersebut memiliki kontribusi besar terhadap pembedaan kelas. Rumus Information Gain didefinisikan sebagai:

$$IG(Y,w) = H(Y) - H(Y|w) \quad (1)$$

dengan H(Y) menyatakan entropi kelas sebelum keberadaan kata w diketahui, dan H(Y|w) menyatakan entropi bersyarat terhadap kemunculan kata w dalam dokumen. Pada penelitian ini, fitur direpresentasikan secara biner berdasarkan kemunculan kata dalam dokumen. Proses seleksi fitur

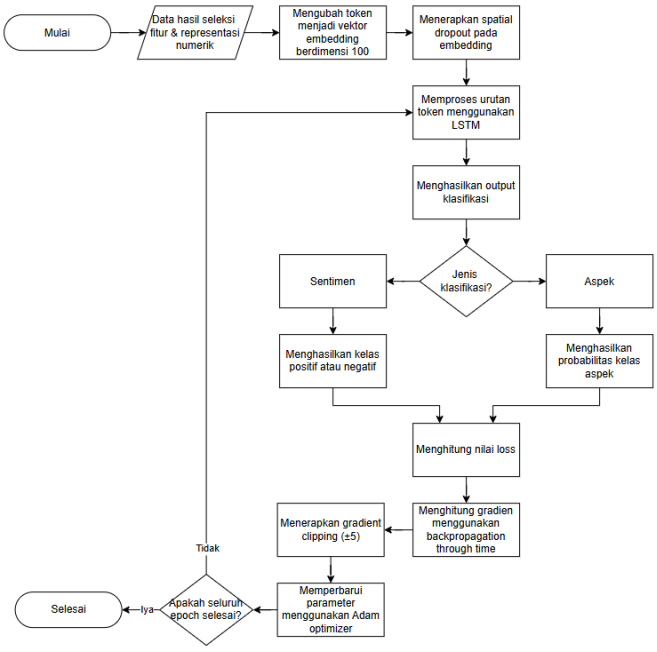
terdiri dari dua tahap: pertama, Document Frequency (DF) filtering untuk menghapus kata yang sangat jarang maupun terlalu umum; kedua, perhitungan IG pada kosakata yang tersisa, kemudian mempertahankan sebagian kata dengan nilai IG tertinggi sebagai vocabulary akhir untuk mengurangi dimensi data serta meningkatkan efisiensi model [15].

2.7. FastText

Representasi kata dilakukan menggunakan FastText, yaitu metode word embedding yang dikembangkan oleh Facebook AI Research dan memanfaatkan subword representation [6]. Berbeda dengan Word2Vec yang merepresentasikan setiap kata sebagai satu vektor, FastText merepresentasikan setiap kata sebagai gabungan beberapa character n-gram, sehingga kata-kata dengan struktur yang mirip akan memiliki representasi vektor yang berdekatan. Keunggulan utama FastText adalah kemampuannya menangani Out-of-Vocabulary (OOV) words, yaitu kata yang tidak muncul saat pelatihan, karena tetap dapat membangun representasinya dari n-gram karakter penyusunnya. Pendekatan ini sangat sesuai untuk data ulasan berbahasa Indonesia yang cenderung informal dan mengandung banyak variasi penulisan, singkatan, dan kesalahan ketik. Proses pelatihan FastText menggunakan algoritma Skip-Gram Negative Sampling, yaitu mempelajari representasi kata dengan cara memprediksi konteks dari kata target. Hasil representasi FastText kemudian digunakan sebagai input pada model LSTM untuk mempelajari pola sentimen dan aspek dalam teks.

2.8. LSTM

Klasifikasi dilakukan menggunakan Long Short-Term Memory (LSTM), yaitu varian dari model Recurrent Neural Network (RNN) yang dirancang untuk mengatasi masalah vanishing gradient pada RNN standar [16]. LSTM menggunakan mekanisme gate yang terdiri dari forget gate, input gate, dan output gate untuk mengontrol aliran informasi sepanjang urutan waktu, sehingga mampu menangkap dependensi jangka panjang dalam data teks. Arsitektur model yang digunakan pada penelitian ini terdiri dari embedding layer berdimensi 100 yang diinisialisasi dengan bobot dari FastText, diikuti oleh spatial dropout layer dengan rate 0,2 untuk mencegah overfitting, serta lapisan LSTM dengan 96 unit tersembunyi. Hasil representasi dari LSTM kemudian diteruskan ke dua output layer yang terpisah: (1) output layer untuk klasifikasi sentimen menggunakan aktivasi sigmoid yang menghasilkan probabilitas biner positif/negatif, dan (2) output layer untuk klasifikasi aspek menggunakan aktivasi softmax yang menghasilkan distribusi probabilitas pada kelas aspek. Model dilatih menggunakan optimizer Adam dengan learning rate 0,0005, ukuran batch 64, dan fungsi loss Binary Cross-Entropy untuk sentimen dan Categorical Cross-Entropy untuk aspek. Penentuan hyperparameter optimal dilakukan menggunakan pendekatan grid search dengan menguji 48 kombinasi dari tiga parameter utama yaitu hidden unit, dropout rate, dan learning rate.



Gambar 2. Arsitektur Model LSTM yang Digunakan

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2.9. Evaluasi

Evaluasi model dilakukan untuk mengukur performa klasifikasi sentimen dan aspek yang dihasilkan. Pengukuran dilakukan menggunakan confusion matrix yang menghasilkan empat nilai utama, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Berdasarkan nilai tersebut, digunakan empat metrik evaluasi utama sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Accuracy digunakan untuk mengukur proporsi prediksi yang benar secara keseluruhan. Precision mengukur ketepatan model dalam memprediksi kelas positif, sedangkan Recall mengukur kemampuan model dalam menemukan semua sampel positif yang sebenarnya. F1-Score memberikan keseimbangan antara precision dan recall, dan sangat berguna ketika distribusi kelas tidak seimbang. Selain itu, digunakan metode k-fold cross-validation untuk memperoleh hasil evaluasi yang lebih stabil dan mengurangi bias terhadap pembagian data tertentu [17].

2.10. BlackBox Testing

Black box testing merupakan teknik pengujian perangkat lunak yang memvalidasi fungsionalitas sistem berdasarkan input dan output tanpa memperhatikan proses internal [18]. Metode ini bersifat specification-based testing karena hanya mengacu pada kebutuhan dan spesifikasi sistem. Dalam penerapannya, penguji memberikan berbagai skenario input dan mengevaluasi apakah output yang dihasilkan sesuai dengan yang diharapkan. Pada penelitian ini, black box testing digunakan untuk menguji fitur utama sistem berbasis web, meliputi prediksi teks tunggal, prediksi batch, serta fitur pendukung seperti preprocessing, unggah file, unduh hasil, dan penanganan input tidak valid.

3. Result and Discussion

3.1. Pengumpulan dan Validasi Data

Data yang digunakan dalam penelitian ini berupa ulasan pengguna aplikasi GoTube yang diperoleh dari Google Play Store menggunakan google-play-scraper. Pengambilan data dibatasi pada rentang tahun 2021 hingga 2024 untuk menjaga relevansi terhadap kondisi aplikasi terbaru. Proses scraping menghasilkan sebanyak 284.853 ulasan pengguna.

Sebelum digunakan, data melalui tahap pembersihan awal dengan menghapus ulasan duplikat, nilai kosong, serta ulasan yang hanya berisi emoji. Setelah proses ini, dilakukan pelabelan sentimen berdasarkan rating, di mana rating 1–2 dikategorikan sebagai negatif dan rating 4–5 sebagai positif, sedangkan rating 3 tidak digunakan karena bersifat ambigu. Hasilnya diperoleh 155.080 ulasan yang siap digunakan sebagai dataset awal. Analisis distribusi sentimen menunjukkan bahwa dataset tidak seimbang, dengan dominasi ulasan positif sebesar 80,7% dan ulasan negatif sebesar 19,3%. Ketidakseimbangan ini berpotensi menyebabkan model bias terhadap kelas mayoritas.

Untuk meningkatkan kualitas label, dilakukan proses pseudo labeling menggunakan model IndoBERT. Validasi dilakukan dengan membandingkan label berbasis rating dan hasil prediksi model. Hanya data dengan label yang konsisten yang digunakan, sedangkan data yang konflik atau tidak memperoleh label dibuang. Hasil validasi menunjukkan bahwa dari 155.080 data awal, sebanyak 108.888 ulasan (70,2%) memiliki label yang konsisten dan digunakan pada tahap selanjutnya, sedangkan 23.743 data memiliki konflik label dan 22.449 data tidak memperoleh label. Ringkasan hasil pengumpulan dan pembersihan data ditunjukkan pada Tabel 1.

Tabel 1. Ringkasan Hasil Pengumpulan Data

Keterangan	Jumlah
Total ulasan hasil scraping	284.853
Ulasan hanya berisi emoji (dibuang)	1.388
Ulasan negatif (rating 1–2)	29.863
Ulasan positif (rating 4–5)	125.217
Total ulasan valid	155.080
Data konsisten setelah pseudo-labeling	108.888
Data konflik label (dibuang)	23.743
Data tanpa label (dibuang)	22.449

3.2. Undersampling

Dataset hasil validasi berjumlah 108.888 ulasan dengan distribusi yang masih tidak seimbang, terdiri dari 82.399 ulasan positif dan 26.489 ulasan negatif. Untuk mengatasi hal tersebut, dilakukan penyeimbangan menggunakan metode random undersampling dengan mengambil secara acak 10.000 ulasan dari masing-masing kelas. Dataset yang dihasilkan berjumlah 20.000 ulasan yang terdiri dari 10.000 ulasan positif dan 10.000 ulasan negatif. Dataset kemudian dibagi menjadi data latih (80%) dan data uji (20%) menggunakan stratified split. Ringkasan pembagian dataset ditunjukkan pada Tabel 2.

Tabel 2. Pembagian Dataset Setelah Undersampling

Set Data	Jumlah Sampel	Positif	Negatif
Data Latih (80%)	16.000	8.000	8.000
Data Uji (20%)	4.000	2.000	2.000
Total	20.000	10.000	10.000

3.3. Text Prepocessing

Pada tahap ini dilakukan proses text preprocessing untuk membersihkan dan menyeragamkan teks ulasan sebelum digunakan dalam pemodelan. Tahapan preprocessing yang digunakan meliputi case folding, normalisasi karakter berulang, konversi emotikon, cleaning, normalisasi kata tidak baku, penanganan negasi, penghapusan stopwords, serta stemming. Proses ini bertujuan untuk mengurangi noise pada data, mengatasi penggunaan bahasa tidak formal, serta menyeragamkan variasi penulisan agar model dapat menangkap pola semantik dengan lebih akurat. Contoh hasil preprocessing ditunjukkan pada Tabel 3.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Tabel 3. Contoh Hasil Text Preprocessing

Tahapan	Teks
Data Awal	Mantap!! gk ada iklannya, nonton pun puas 😍
Case Folding	mantap!! gk ada iklannya, nonton pun puas 😍
Normalisasi Karakter	mantap!! gk ada iklannya, nonton pun puas 😍
Cleaning	mantap gk ada iklannya nonton pun puas
Normalisasi Kata	mantap tidak ada iklannya nonton pun puas
Negasi	mantap tidak_ada iklan nonton puas
Stopwords Removal	mantap tidak_ada iklan nonton puas
Stemming	mantap tidak_ada iklan nonton puas

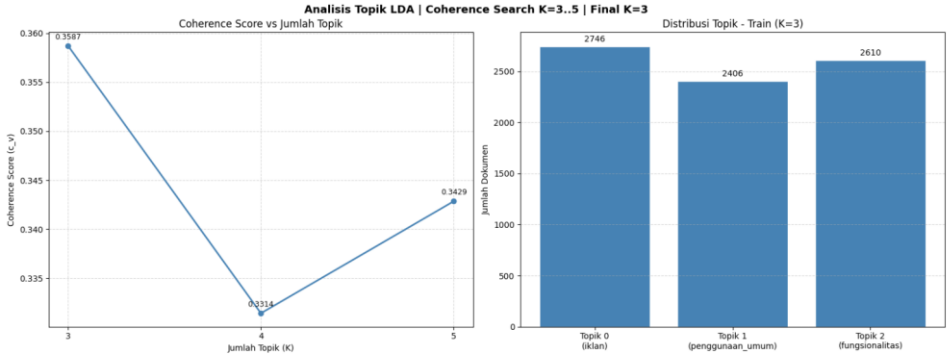
3.4. Ekstraksi Aspek dengan LDA

Pada tahap ini digunakan metode LDA untuk mengidentifikasi aspek yang dibahas dalam ulasan pengguna GoTube. Penentuan jumlah topik dilakukan dengan menguji beberapa nilai K pada rentang 3 hingga 5 menggunakan coherence score dengan metrik c_v. Hasil pengujian menunjukkan bahwa nilai coherence score tertinggi diperoleh pada K=3 dengan nilai 0,3587. Evaluasi manual terhadap kata-kata dominan setiap topik menunjukkan bahwa K=3 menghasilkan topik yang paling jelas, bermakna, dan mudah diinterpretasikan dibandingkan K=4 (0,3314) dan K=5 (0,3429). Berdasarkan hasil pemodelan, setiap ulasan dipetakan ke topik dengan probabilitas tertinggi sebagai dominant topic, yang kemudian diinterpretasikan secara manual menjadi tiga aspek utama. Pemetaan topik ke aspek ditunjukkan pada Tabel 4.

Tabel 4. Pemetaan Topik LDA ke Aspek Aplikasi

Topik	Kata Dominan	Aspek	Jumlah (Train)
Topik 0	iklan, aplikasi, tidak_iklan, tanpa_iklan, tidak, tonton, suka	Iklan	2.746
Topik 1	aplikasi, tidak, gotube, video, baik, putar, suka, lagu, buka	Penggunaan Umum	2.406
Topik 2	tidak, video, tidak_jelas, jadi, cari, muncul, baru, keluar	Fungsionalitas	2.610

Distribusi topik menunjukkan bahwa aspek iklan merupakan topik paling dominan (2.746 dokumen), diikuti fungsionalitas (2.610 dokumen) dan penggunaan umum (2.406 dokumen).



Gambar 3. Analisis LDA: *Coherence Score* dan Distribusi Topik

3.5. Seleksi Fitur dengan Information Gain

Seleksi fitur dilakukan untuk memilih kata yang paling relevan dalam membedakan sentimen. Tahap ini terdiri dari dua proses utama, yaitu Document Frequency (DF) filtering dan perhitungan Information Gain (IG). Pada tahap DF filtering, jumlah kosakata berkurang dari 7.113 menjadi 2.239 kata. Selanjutnya, perhitungan IG dilakukan pada 2.239 kata tersebut untuk mengukur kontribusi setiap kata terhadap label sentimen. Dipertahankan 70% kata dengan nilai IG tertinggi sehingga diperoleh 1.567 kata sebagai kosakata akhir dengan nilai cutoff sebesar 0,000068. Kata-kata dengan nilai IG tinggi seperti tidak (0,1428), mantap (0,1307), dan bagus (0,1053) memiliki kontribusi besar dalam membedakan sentimen positif dan negatif, sedangkan kata dengan nilai rendah seperti haha, sponsor, koneksi, blok, dan benar memiliki pengaruh yang sangat kecil dan dapat dihilangkan tanpa menurunkan performa secara signifikan.

3.6. Representasi Kata dengan FastText

Setelah proses seleksi fitur, teks ulasan diubah ke dalam bentuk representasi numerik menggunakan metode FastText. Sebelum pelatihan embedding, teks difilter menggunakan kosakata hasil seleksi fitur, kemudian diubah menjadi indeks numerik dan dipadatkan menggunakan padding dengan panjang sekuens maksimum 40 token. Pada FastText, setiap kata direpresentasikan sebagai kumpulan n-gram karakter dengan rentang 3 hingga 6 karakter.

Proses pelatihan FastText dilakukan menggunakan algoritma Skip-Gram Negative Sampling untuk mempelajari hubungan distribusional antara kata target dan konteksnya. Pelatihan dilakukan untuk tiga varian kosakata yang berbeda. Varian IG menghasilkan matriks embedding berukuran 1.568 × 100, varian NO_IG berukuran 2.240 × 100, dan varian LDA berukuran 5.265 × 100, seluruhnya dengan dimensi embedding 100. Nilai loss mengalami penurunan yang konsisten pada ketiga varian, di mana varian IG menghasilkan loss akhir paling rendah (1,8022), diikuti LDA (1,8424), dan NO_IG (1,8523), yang menunjukkan bahwa model berhasil mempelajari representasi kata dengan baik. Embedding yang dihasilkan pada tahap ini selanjutnya digunakan sebagai representasi awal kata pada masing-masing model LSTM untuk klasifikasi sentimen berbasis aspek.

3.7. Hasil Model LSTM

Setelah representasi kata diperoleh dari FastText, dilakukan pelatihan model LSTM untuk klasifikasi sentimen dan aspek. Model sentimen digunakan untuk membedakan ulasan ke dalam dua kelas, yaitu positif dan negatif, sedangkan model aspek mengklasifikasikan ulasan ke dalam tiga kategori, yaitu iklan, penggunaan umum, dan fungsionalitas.

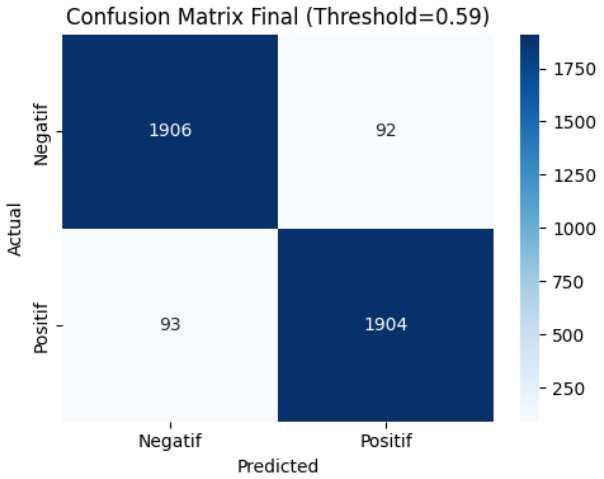
Sebelum pelatihan model final, dilakukan grid search terhadap 48 kombinasi hyperparameter yang diuji secara paralel menggunakan dua sesi komputasi. Kombinasi yang diuji meliputi hidden_dim ∈ {32, 64, 96, 128}, dropout ∈ {0,2; 0,3; 0,5}, dan learning_rate ∈ {0,0001; 0,0005; 0,001; 0,005}. Hasil grid search menunjukkan bahwa kombinasi terbaik adalah hidden_dim=96, dropout=0,2, dan learning_rate=0,0005 dengan nilai F1 validasi sebesar 0,9724. Selanjutnya dilakukan threshold tuning untuk menyeimbangkan precision dan recall, dan diperoleh threshold optimal sebesar 0,59 dengan selisih precision-recall terkecil sebesar 0,0005. Hasil evaluasi model sentimen pada data uji ditunjukkan pada Tabel 5..

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Tabel 5. Hasil Evaluasi Model LSTM Sentimen

Metrik	Nilai
Accuracy	0,9537
Precision	0,9539
Recall	0,9534
F1-Score	0,9537

Model sentimen mencapai accuracy sebesar 95,37% dan F1-score sebesar 0,9537 pada threshold optimal 0,59. Model menunjukkan nilai precision (0,9539) dan recall (0,9534) yang sangat seimbang dengan selisih hanya 0,0005, mengindikasikan bahwa model tidak bias terhadap salah satu kelas sentimen. Berdasarkan confusion matrix, sentimen negatif berhasil diprediksi dengan benar sebanyak 1.906 dari 1.998 data (92 salah), sedangkan sentimen positif berhasil diklasifikasikan dengan benar sebanyak 1.904 dari 1.997 data (93 salah).



Gambar 4. Confusion Matrix Model Sentimen

Untuk memvalidasi stabilitas model, dilakukan 5-fold cross validation yang menghasilkan rata-rata F1 sebesar 0,9646 ± 0,0046 dan rata-rata accuracy sebesar 0,9644 ± 0,0049. Nilai standar deviasi yang kecil menunjukkan model stabil dan konsisten di berbagai pembagian data.

Selain itu, dilakukan ablation study untuk mengukur kontribusi seleksi fitur Information Gain. Hasil perbandingan ditunjukkan pada Tabel berikut.

Tabel 6. Hasil *Ablation Study* IG vs NO_IG

Kondisi	Vocab Size	F1	Accuracy	Precision	Recall
IG	1.567	0,9537	0,9537	0,9539	0,9534
NO_IG	2.239	0,9557	0,9691	0,9691	0,9414

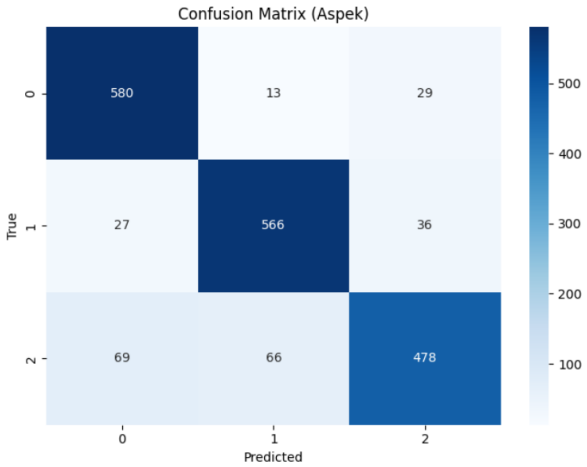
Meskipun F1 tanpa IG sedikit lebih tinggi (0,9550 vs 0,9537), model tanpa IG menunjukkan ketidakseimbangan yang signifikan antara precision (0,9691) dan recall (0,9414) dengan selisih 0,0277. Sebaliknya, model dengan IG memiliki precision (0,9539) dan recall (0,9534) yang jauh lebih seimbang dengan selisih hanya 0,0005. Selain itu, penggunaan IG berhasil mereduksi vocab size sebesar 30% (672 kata) tanpa mengorbankan keseimbangan prediksi, sehingga IG terbukti meningkatkan efisiensi dan keseimbangan model secara signifikan.

Sementara itu, model aspek memperoleh accuracy sebesar 87,12% dengan macro F1-score sebesar 0,8697. Rincian performa per kelas aspek ditunjukkan pada Tabel 7.

Tabel 7. Hasil Evaluasi Model LSTM Aspek per Kelas pada Data Uji

Kelas Aspek	Precision	Recall	F1-Score	Support
Iklan	0,8600	0,9325	0,8948	622
Pengguna Umum	0,8800	0,8998	0,8898	629
Fungsionalitas	0,8800	0,7796	0,8268	613
Macro Average	**0,8733**	**0,8706**	**0,8697**	**1.864**

Berdasarkan Tabel 7, aspek iklan memperoleh F1-score tertinggi (0,8948) diikuti penggunaan umum (0,8898), sedangkan aspek fungsionalitas memperoleh F1-score terendah (0,8268). Rendahnya performa aspek fungsionalitas disebabkan kosakata yang lebih tumpang tindih dengan aspek penggunaan umum keduanya sering membahas perilaku aplikasi secara umum sehingga model mengalami kesulitan membedakannya. Hal ini terlihat dari nilai recall fungsionalitas yang lebih rendah (0,7796) dibandingkan dua aspek lainnya.



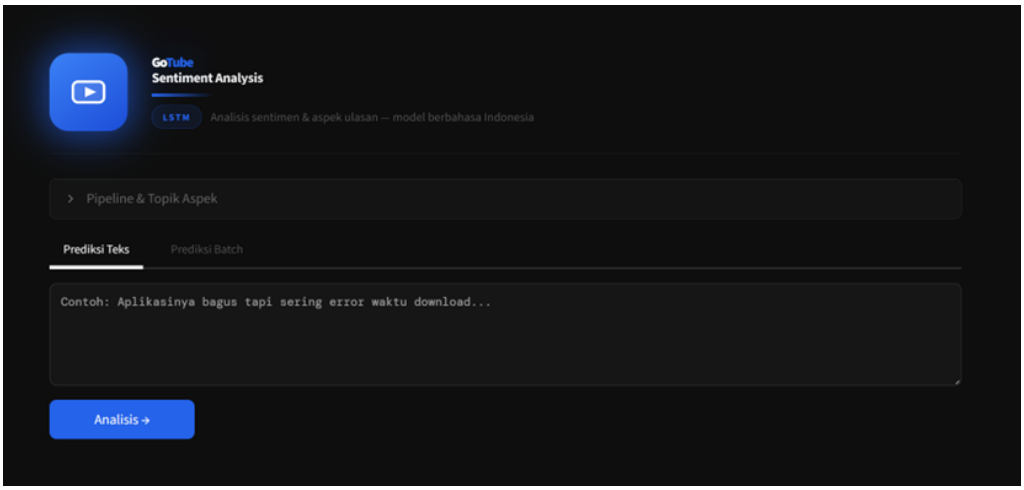
Gambar 5. Confusion Matrix Model Aspek

Untuk validasi stabilitas model aspek, dilakukan 5-fold cross validation yang menghasilkan rata-rata F1 sebesar 0,8668 ± 0,0078 dan rata-rata accuracy sebesar 0,8685 ± 0,0073. Nilai F1 test set (0,8697) berada di atas rata-rata K-Fold (0,8668), mengindikasikan model aspek mampu melakukan generalisasi dengan baik pada data yang belum pernah dilihat sebelumnya.

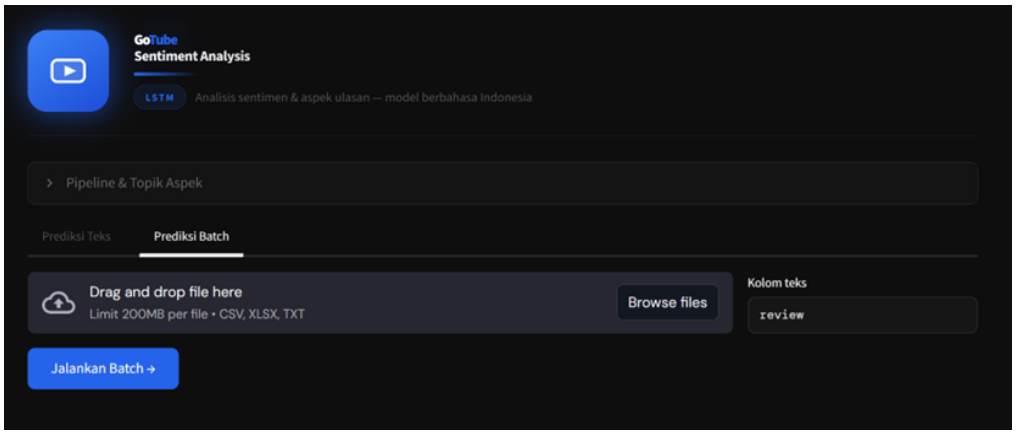
3.8. Implementasi Hasil

Model yang telah dilatih kemudian diimplementasikan dalam bentuk aplikasi berbasis web menggunakan framework Streamlit. Model disimpan terlebih dahulu menggunakan pickle dan dimuat kembali saat aplikasi dijalankan sehingga proses prediksi dapat dilakukan secara langsung tanpa perlu pelatihan ulang. Sistem dirancang dengan antarmuka yang sederhana dan intuitif sehingga dapat digunakan oleh pengguna non-teknis sekalipun. Sistem memiliki dua mode utama. Pertama, mode prediksi teks tunggal (single prediction) memungkinkan pengguna memasukkan satu ulasan secara langsung pada kolom teks yang tersedia. Teks yang dimasukkan akan melalui tahap preprocessing yang sama seperti saat pelatihan, kemudian diproses oleh kedua model untuk menghasilkan label sentimen, probabilitas sentimen, aspek yang terdeteksi, serta probabilitas aspek. Kedua, mode prediksi batch (batch prediction) memungkinkan pengguna mengunggah file berisi banyak ulasan dalam format CSV, XLSX, atau TXT. Sistem akan memproses seluruh data secara otomatis dan menampilkan ringkasan hasil analisis, termasuk distribusi sentimen dan aspek dalam bentuk tabel. Setelah proses prediksi selesai, pengguna dapat mengunduh hasil analisis dalam format CSV untuk keperluan analisis lebih lanjut.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 6. Tampilan Prediksi Teks Tunggal



Gambar 7 Tampilan Upload dan Hasil Prediksi Batch

3.9. Pengujian Sistem (BlackBox Testing)

Pengujian sistem dilakukan menggunakan metode black box testing untuk memastikan seluruh fitur berjalan sesuai dengan yang diharapkan. Pengujian difokuskan pada fitur utama sistem tanpa memperhatikan proses internal. Skenario pengujian mencakup tujuh fitur utama sistem. Hasil pengujian selengkapnya ditunjukkan pada Tabel 8.

Tabel 8. Hasil Pengujian Black Box Testing

No	Fitur Diuji	Input	Output yang Diharapkan	Status
1	Prediksi teks tunggal	Teks ulasan valid	Menampilkan sentimen, probabilitas, dan aspek	Berhasil
2	Preprocessing teks	Teks mentah berbahasa Indonesia	Menampilkan hasil preprocessing yang telah dibersihkan	Berhasil
3	Prediksi aspek	Teks ulasan valid	Menampilkan label aspek dan nilai probabilitas	Berhasil
4	Upload file batch	File CSV/XLSX/TXT	Menampilkan pratinjau data yang diunggah	Berhasil
5	Prediksi batch	File berisi banyak ulasan	Menampilkan tabel hasil analisis sentimen dan aspek	Berhasil

6	Download hasil	Klik tombol unduh	File CSV hasil analisis berhasil diunduh	Berhasil
7	Input kosong	Tidak ada input teks	Menampilkan pesan peringatan	Berhasil

Hasil pengujian menunjukkan bahwa seluruh fitur sistem berjalan dengan baik dan menghasilkan output sesuai dengan yang diharapkan pada setiap skenario uji. Sistem berhasil menangani berbagai kondisi input, termasuk kondisi input kosong yang menghasilkan pesan peringatan yang informatif bagi pengguna. Hal ini menunjukkan bahwa sistem telah siap digunakan untuk analisis sentimen dan aspek secara optimal.

4. Kesimpulan

Penelitian ini mengembangkan sistem analisis sentimen berbasis aspek pada ulasan pengguna aplikasi GoTube dengan memanfaatkan kombinasi metode LDA, Information Gain, FastText, dan LSTM. Penggunaan pseudo-labeling dengan IndoBERT menunjukkan kemampuan dalam meningkatkan kualitas label dataset, dengan mempertahankan 70,2% data yang memiliki label konsisten. Information Gain mampu mereduksi dimensi fitur dari 2.239 menjadi 1.567 kata sehingga fitur yang digunakan menjadi lebih relevan dan efisien. FastText dengan pendekatan subword representation membantu menangani variasi kata dan kesalahan penulisan pada ulasan berbahasa Indonesia informal. Model LSTM yang dilatih menggunakan embedding FastText menghasilkan performa yang baik, dengan akurasi 95,37% dan F1-score 0,9537 pada klasifikasi sentimen menggunakan threshold optimal 0,59, serta akurasi 87,12% dan macro F1-score 0,8697 pada klasifikasi aspek. Metode LDA mengidentifikasi tiga aspek utama, yaitu iklan, penggunaan umum, dan fungsionalitas, dengan aspek iklan sebagai topik yang paling dominan. Berdasarkan pengujian black box testing, seluruh fitur sistem berjalan sesuai fungsinya, sehingga pendekatan yang digunakan menunjukkan potensi yang baik dalam analisis sentimen berbasis aspek pada ulasan aplikasi berbahasa Indonesia.

Referensi

[1] M. Christianto, J. Andjarwirawan, and A. Tjondrowiguno, "Aplikasi analisa sentimen pada komentar berbahasa Indonesia dalam objek video di website YouTube menggunakan metode Naïve Bayes Classifier," Jurnal INFRA, 2020.
[2] T. Shaik et al., "Sentiment analysis and opinion mining on educational data: A survey," Natural Language Processing Journal, vol. 2, 2023.
[3] B. Ramadhani and R. R. Suryono, "Komparasi algoritma Naïve Bayes dan logistic regression untuk analisis sentimen metaverse," Jurnal Media Informatika Budidarma, 2024. https://doi.org/10.30865/mib.v8i2.7458
[4] D. Melati and Mulyati, "Penerapan metode Long Short-Term Memory (LSTM) dalam analisis sentimen terhadap pelaksanaan pilkada di masa pandemi COVID-19," Jurnal Informatika dan Komputer, vol. 22, no. 1, pp. 22–28, 2024. https://doi.org/10.35508/jicon.v12i1.9899
[5] H. S. Utama et al., "Sentimen analisis kebijakan ganjil genap di tol Bekasi menggunakan algoritma Naive Bayes dengan optimalisasi information gain," Jurnal Pilar Nusa Mandiri, vol. 15, no. 2, pp. 247–254, 2019. https://doi.org/10.33480/pilar.v15i2.705
[6] E. M. Dharma et al., "The accuracy comparison among Word2Vec, GloVe, and FastText towards Convolution Neural Network (CNN) text classification," J. Theor. Appl. Inf. Technol., vol. 31, no. 2, 2022.
[7] Y. Kustyaningsih and Y. Permana, "Penggunaan Latent Dirichlet Allocation (LDA) dan Support-Vector Machine (SVM) untuk menganalisis sentimen berdasarkan aspek dalam ulasan aplikasi EdLink," TEKNIKA, vol. 13, no. 1, pp. 127–136, 2024. https://doi.org/10.34148/teknika.v13i1.746
[8] J. Gan and Y. Qi, "Selection of the optimal number of topics for LDA topic model—taking patent policy analysis as an example," Entropy, vol. 23, no. 10, p. 1301, 2021. https://doi.org/10.3390/e23101301
[9] K. S. Nugroho et al., "BERT fine-tuning for sentiment analysis on Indonesian mobile apps reviews," in Proc. 6th Int. Conf. Sustain. Inf. Eng. Technol. (SIET), pp. 258–264, 2021. https://doi.org/10.1145/3479645.3479679
[10] H. Syafutra and Kusrini, "Random undersampling untuk menangani ketidakseimbangan kelas pada klasifikasi teks," Jurnal Ilmiah Ilmu Komputer, 2025.

[11] R. Hidayat and W. Gata, "Pemanfaatan IndoBERT dan Long Short-Term Memory (LSTM) pada analisis sentimen ulasan aplikasi Depok Single Window," Inf. Syst. Educ. Prof., vol. 10, no. 1, pp. 13–24, 2025.
[12] A. Santoso, A. Nugroho, and A. S. Sunge, "Analisis sentimen tentang mobil listrik dengan metode Support Vector Machine dan feature selection Particle Swarm Optimization," J. Pract. Comput. Sci., vol. 2, no. 1, 2022.
[13] C. Meaney et al., "Quality indices for topic model selection and evaluation: A literature review and case study," BMC Med. Inform. Decis. Mak., vol. 23, p. 132, 2023. https://doi.org/10.1186/s12911-023-02216-1
[14] C. Destitus, "Support Vector Machine vs information gain: Analisis sentimen cyberbullying di Twitter Indonesia," ULTIMA InfoSys, vol. XI, no. 2, p. 107, 2020.
[15] F. Pramono, D. Rosiyadi, and W. Gata, "Integrasi N-gram, information gain, particle swarm optimization di Naïve Bayes untuk optimasi sentimen Google Classroom," Jurnal RESTI, vol. 3, no. 3, 2019.
[16] R. Faza Inaku and J. C. Chandra, "Implementasi data mining dalam prediksi harga saham menggunakan metode Long Short Term Memory (LSTM)," Jurnal TICOM, vol. 12, no. 1, 2023.
[17] Z. Mahendra and A. Ridok, "Analisis sentimen opini masyarakat terhadap fenomena TikTokShop di Indonesia menggunakan metode K-Nearest Neighbor berbasis N-gram dengan seleksi fitur information gain," J. Pengemb. Teknol. Inf. Ilmu Komput. http://j-ptiik.ub.ac.id
[18] A. Verma, A. Khatana, and S. Chaudhary, "A comparative study of black box testing and white box testing," Int. J. Comput. Sci. Eng., vol. 5, no. 12, pp. 301–304, 2017.