



Department of Digital Business

**Journal of Artificial Intelligence and Digital Business (RIGGS)**

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 4 (2025) pp: 7722-7732

P-ISSN: 2963-9298, e-ISSN: 2963-914X

---

## Analisis Sentimen Ulasan Aplikasi TikTok Menggunakan Convolutional Neural Network

Kelvin Dwi Pranata Sembiring<sup>1</sup>, Agam Setiawan<sup>2</sup>, Harun Al Rosyid<sup>3</sup>

<sup>1,2,3</sup>Pendidikan Teknologi Informasi, Fakultas Teknik, Universitas Negeri Surabaya

[kelvin.22085@mhs.unesa.ac.id](mailto:kelvin.22085@mhs.unesa.ac.id), [agam.22091@mhs.unesa.ac.id](mailto:agam.22091@mhs.unesa.ac.id), [harunrosyid@unesa.ac.id](mailto:harunrosyid@unesa.ac.id)

### Abstrak

Analisis sentimen terhadap ulasan pengguna merupakan instrumen krusial dalam memahami persepsi serta pengalaman pengguna terhadap aplikasi mobile seperti TikTok. Di tengah pesatnya pertumbuhan platform digital, pemrosesan umpan balik secara otomatis menjadi kebutuhan mendesak bagi pengembang. Penelitian ini bertujuan untuk menganalisis sentimen ulasan aplikasi TikTok dengan membandingkan efektivitas metode *machine learning* dan *deep learning*, yaitu Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Support Vector Machine (SVM), dan Logistic Regression (LR). Data penelitian diperoleh melalui teknik *scraping* pada Google Play Store periode Oktober 2025, menghasilkan 19.569 ulasan mentah yang kemudian diproses menjadi 11.368 data bersih melalui tahapan *preprocessing*. Tahapan tersebut meliputi pembersihan teks, *stopword removal*, tokenisasi, serta pelabelan sentimen dan kategori dengan rasio pembagian data 80:20. Evaluasi performa model dilakukan secara komprehensif menggunakan metrik Accuracy, Macro-F1, dan Weighted-F1. Hasil eksperimen menunjukkan bahwa model CNN mencapai performa terbaik dengan nilai Accuracy 0,9895, Macro-F1 0,9407, dan Weighted-F1 0,9897, secara konsisten mengungguli LSTM, SVM, dan Logistic Regression. Skor tinggi ini membuktikan bahwa arsitektur CNN sangat efektif dalam mengekstraksi fitur lokal dan menangani karakteristik teks pendek yang informal pada ulasan media sosial. Temuan ini memberikan kontribusi signifikan bagi peneliti NLP dalam mengoptimalkan sistem klasifikasi sentimen, serta memberikan wawasan strategis bagi pengembang aplikasi untuk meningkatkan kualitas layanan berdasarkan sentimen pengguna secara akurat.

**Kata Kunci:** Analisis Sentimen, TikTok, Deep Learning, CNN, Machine Learning, NLP

### 1. Latar Belakang

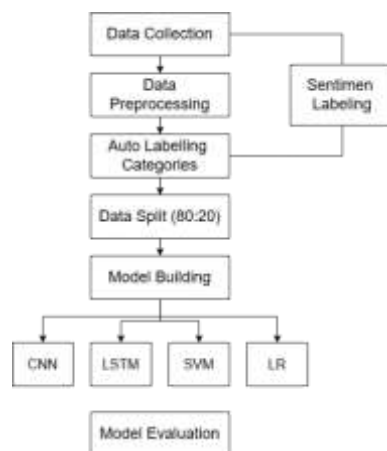
Pertumbuhan pesat teknologi digital telah mendorong peningkatan signifikan penggunaan aplikasi mobile, termasuk *TikTok*, yang kini menjadi salah satu jaringan media sosial yang paling banyak digunakan di seluruh dunia. Setiap hari, jutaan orang meninggalkan ulasan dan komentar yang mencerminkan persepsi serta pengalaman mereka terhadap fitur aplikasi. Informasi ini bernilai tinggi bagi *developer* untuk memahami kebutuhan *user* dan meningkatkan kualitas layanan. Namun, volume ulasan yang sangat besar menjadikan analisis manual tidak efisien dan rawan bias, sehingga diperlukan sistem otomatis berbasis *machine learning* untuk membaca dan mengklasifikasikan sentimen pengguna secara akurat [1,2].

Pendekatan *deep learning* dan *machine learning* telah terbukti efektif mengenai analisis sentimen, terutama untuk teks media sosial yang padat dan informal. Model klasik seperti *Support Vector Machine* (SVM) dan *Logistic Regression* (LR) unggul dalam efisiensi, sedangkan model *deep learning* seperti *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM) lebih mampu menangkap konteks semantik dan emosi pengguna [3,4]. Beberapa studi juga menegaskan bahwa metode berbasis CNN dan LSTM menunjukkan performa lebih baik dibandingkan algoritma tradisional dalam menganalisis opini di media sosial, termasuk *TikTok* [5,6].

Penelitian ini bertujuan membangun dan membandingkan model CNN, LSTM, SVM, dan LR dalam analisis sentimen ulasan *TikTok* periode Oktober 2025. Hasil evaluasi dilakukan menggunakan metrik *accuracy*, *macro-F1*, dan *weighted-F1 score* untuk menentukan model terbaik. Hasil penelitian diharapkan dapat membantu pengembang memahami persepsi pengguna secara objektif serta menjadi referensi bagi peneliti *Natural Language Processing* dalam mengembangkan model analisis sentimen untuk teks pendek di media sosial.

## 2. Metode Penelitian

### 2.1 Alur Penelitian



Gambar 1 Alur Penelitian

Penelitian ini dilaksanakan melalui alur kerja yang sistematis dan terstruktur, meliputi: (1) *data collection* ulasan pengguna aplikasi TikTok d Google Play Store, (2) tahap prapemrosesan teks, (3) pelabelan sentimen dan kategori ulasan, (4) *data split* menjadi data latih dan data uji, (5) pembangunan serta pelatihan model klasifikasi berbasis *Convolutional Neural Network* (CNN), *Long Short-Term Memory* (LSTM), *Support Vector Machine* (SVM), dan *Logistic Regression*, serta (6) model evaluasi kinerja menggunakan metrik akurasi, *macro-F1*, dan *weighted-F1*. Pendekatan penelitian ini mengacu pada kerangka kerja analisis sentimen modern berbasis *machine learning* dan *deep learning* yang menekankan pentingnya alur pemrosesan yang jelas dan terintegrasi, mulai dari ekstraksi data hingga evaluasi performa model, sebagaimana dijelaskan dalam penelitian sebelumnya [7,8].

### 2.2 Pengumpulan Data

Data penelitian diperoleh dari ulasan pengguna aplikasi TikTok pada Google Play Store dengan periode pengambilan data bulan Oktober 2025. Proses akuisisi data dilakukan menggunakan *Google Play Scraper* berbasis Python untuk mengekstraksi teks ulasan, nilai rating, serta tanggal unggahan. Dataset awal berjumlah 19.569 entri ulasan yang selanjutnya diseleksi untuk memastikan relevansi dan kelengkapan teks. Pemanfaatan ulasan publik sebagai sumber data utama dalam analisis sentimen telah banyak diterapkan dalam penelitian sebelumnya dan terbukti relevan untuk merepresentasikan persepsi pengguna terhadap aplikasi, khususnya dalam konteks pemodelan sentimen pengguna [7,8].

### 2.3 Preprocessing Data

Tahap preprocessing bertujuan untuk menormalkan teks agar siap diproses oleh algoritma pembelajaran mesin. Di antara tindakan yang dilakukan meliputi:

1. *Cleaning text*, yaitu menghapus URL, emoji, simbol, dan mengubah seluruh huruf menjadi *lowercase*;
2. *Stopword removal* menggunakan daftar *stopword* Bahasa Indonesia;
3. *Tokenisasi*, memecah kalimat menjadi kata-kata terpisah;
4. *Padding* untuk menyesuaikan panjang input pada model *deep learning*;
5. *Representasi fitur*, di mana data klasik seperti SVM dan *Logistic Regression* menggunakan TF-IDF, sedangkan model CNN dan LSTM menggunakan *word embedding*.

Pembobotan TF-IDF tetap menjadi metode yang efisien untuk merepresentasikan teks yang tidak terstruktur, sementara penggunaan *Word2Vec* terbukti mampu meningkatkan akurasi model *deep learning*. Selain itu, teknik TF-IDF juga menunjukkan hasil yang optimal untuk pemrosesan bahasa Indonesia ketika dikombinasikan dengan SVM atau *Logistic Regression* [9,10,11].

## 2.4 Sentiment Labeling

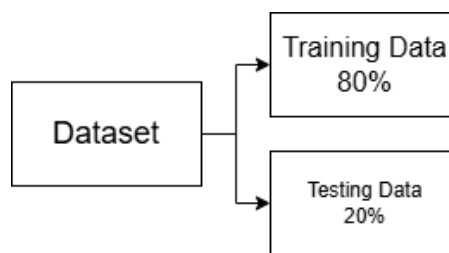
Proses pelabelan sentimen dilakukan dengan menggunakan pendekatan *semi-rule-based* dengan mengombinasikan leksikon sentimen Bahasa Indonesia dan pemeriksaan manual guna meningkatkan akurasi pelabelan. Setiap ulasan diklasifikasikan menjadi tiga kategori utama, yakni positif, netral, dan negatif. Pendekatan ini sejalan dengan praktik umum dalam penelitian analisis sentimen, di mana kombinasi metode *rule-based* dan *manual checking* terbukti mampu menghasilkan dataset dengan tingkat validitas yang lebih tinggi [12,13].

## 2.5 Auto Labeling Kategori Ulasan

Selain pelabelan sentimen, sistem juga menerapkan *auto-labeling* terhadap topik ulasan berdasarkan pola kata kunci tertentu. Kategori yang digunakan meliputi *bug*, *ads*, *update*, dan *general*. Sebagai contoh, kata kunci seperti “error”, “lag”, atau “crash” diklasifikasikan ke dalam kategori *bug*, sedangkan kata “iklan”, “ads”, dan “promosi” dimasukkan ke dalam kategori *ads*. Pendekatan ini meniru metode klasifikasi berbasis konteks yang mengelompokkan ulasan pengguna aplikasi berdasarkan tema fungsional melalui teknik *text mining* [7].

## 2.6 Pembagian Data

Dataset yang telah melalui tahap *preprocessing* dipisah menjadi data latih sebesar 80% dan data uji sebesar 20% menggunakan fungsi *train\_test\_split*. Proporsi pembagian tersebut dipilih karena mampu memberikan keseimbangan antara proses pembelajaran model dan tahap pengujian, sehingga membantu meminimalkan risiko *overfitting* serta bias pada hasil evaluasi model [14].



Gambar 2 Data Split 80:20

## 2.7 Pembangunan Model

### 2.7.1 Model CNN

Model *Convolutional Neural Network* (CNN) digunakan untuk menangkap pola lokal antar kata melalui lapisan konvolusi dan *max pooling*. Arsitektur CNN untuk pemrosesan teks terbukti efektif dalam mendeteksi *n-gram* penting. Dalam penelitian ini, setiap teks diubah menjadi vektor *embedding* dan diproses melalui lapisan konvolusi dengan beberapa filter serta fungsi aktivasi ReLU, kemudian dilanjutkan dengan *global max pooling* dan *dense layer* untuk proses klasifikasi [15,16].

### 2.7.2 Model LSTM

Model *Long Short-Term Memory* (LSTM) digunakan untuk mempelajari hubungan jangka panjang antar kata dalam teks ulasan. Input berupa matriks *embedding* yang kemudian diproses melalui satu lapisan LSTM dengan mekanisme *dropout* untuk mengurangi risiko *overfitting*. Perancangan model ini mengikuti desain umum yang banyak digunakan dalam penelitian analisis sentimen ulasan aplikasi, khususnya dalam konteks pemodelan urutan kata pada opini pengguna, karena kemampuan LSTM dalam memahami konteks sekuensial teks secara efektif [13].

### 2.7.3 Model SVM

Pada model *Support Vector Machine* (SVM), fitur teks direpresentasikan menggunakan pembobotan TF-IDF dengan pemilihan *kernel* linear karena efisien untuk dataset berukuran besar dan berdimensi tinggi. Pendekatan ini telah terbukti efektif dalam berbagai penelitian analisis sentimen pada media sosial dan ulasan daring.

Dalam penelitian ini, SVM digunakan sebagai model pembanding untuk mengevaluasi kinerja metode klasifikasi klasik terhadap model berbasis *deep learning* [11,12].

#### 2.7.4 Model Logistic Regression

Model *Logistic Regression* (LR) diterapkan sebagai *baseline* dalam penelitian ini. Pendekatan ini dinilai efisien untuk menangani data teks yang bersifat linier dengan representasi fitur TF-IDF serta tetap relevan karena kemudahan interpretasi hasil dan kestabilan performanya. Selain itu, LR dimanfaatkan untuk membandingkan efektivitas representasi fitur konvensional dengan pendekatan *embedding* yang lebih kompleks dalam konteks perkembangan teknik representasi teks modern [11,17].

### 2.8 Evaluasi Model

Tiga metrik utama digunakan dalam pendekatan evaluasi kinerja, yaitu: *Accuracy*, *Macro-F1*, dan *Weighted-F1*. *Accuracy* digunakan untuk mengukur persentase prediksi yang benar secara keseluruhan, sementara *Macro-F1* mengevaluasi kinerja model dengan memperlakukan setiap kelas secara seimbang tanpa mempertimbangkan ukuran kelas. Adapun *Weighted-F1* mempertimbangkan proporsi masing-masing kelas sehingga lebih representatif untuk dataset dengan distribusi kelas yang tidak seimbang. Penggunaan kombinasi ketiga metrik ini dinilai mampu memberikan evaluasi yang lebih adil dan komprehensif dibandingkan penggunaan akurasi semata, khususnya pada tugas klasifikasi teks multikategori [14].

#### 2.8.1 Accuracy

Accuracy mengukur seberapa banyak prediksi model yang benar dari seluruh data uji

$$Accuracy = \frac{\text{Jumlah prediksi benar}}{\text{total data}}$$

#### 2.8.2 Macro-F1 (Macro Average F1-Score)

Tanpa memperhitungkan kuantitas data per kelas, Macro-F1 adalah skor F1 rata-rata untuk setiap kelas.

$$Macro - F1 = \frac{F1_{class1} + F1_{class2} + \dots}{\text{Jumlah kelas}}$$

#### 2.8.3 Weighted-F1 (Weighted Average F1-Score)

Weighted-F1 menghitung rata-rata F1-score berdasarkan proporsi jumlah data tiap kelas

$$Weighted - F1 = \sum(F1_{classi} \times \text{proporsi data kelas itu})$$

## 3. Hasil dan Diskusi

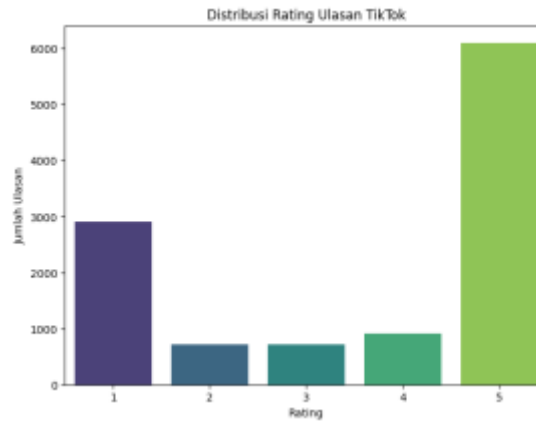
### 3.1 Hasil Preprocessing

Proses preprocessing pada dataset dilakukan untuk menghilangkan gangguan yang tidak perlu dari data.. Data awal yang berjumlah 19.569 ulasan TikTok mengalami pengurangan setelah tahap pembersihan, yang menghasilkan 11.368 ulasan yang siap untuk dianalisis. Penghapusan data duplikat, karakter non-alfanumerik, serta teks yang tidak lengkap merupakan langkah-langkah utama dalam proses ini. Hasil *preprocessing* ini sangat penting untuk menjamin data yang digunakan dalam analisis adalah data yang relevan dan berkualitas tinggi.

### 3.2 Statistik Dasar Dataset

#### 3.2.1 Distribusi Rating

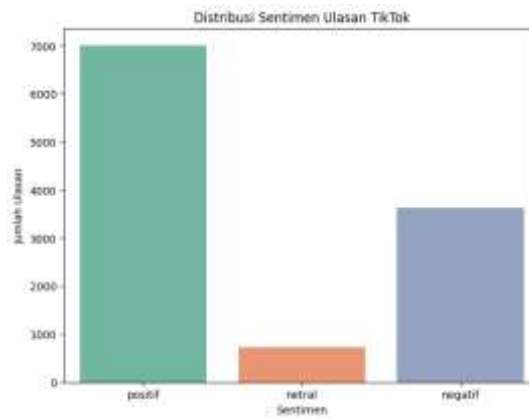
Distribusi rating dari 11.368 ulasan menunjukkan bahwa mayoritas penonton memberikan rating positif terhadap video TikTok yang mereka tonton. Grafik distribusi rating pada Gambar 3 menggambarkan sebagian besar ulasan diberi rating tinggi, yang menunjukkan tingkat kepuasan yang tinggi terhadap konten yang diproduksi di platform tersebut. Data ini menggambarkan dominasi ulasan positif yang mungkin mencerminkan kualitas konten TikTok yang tinggi.



Gambar 3 Distribusi Rating Ulasan Tiktok

### 3.2.2 Distribusi Sentimen

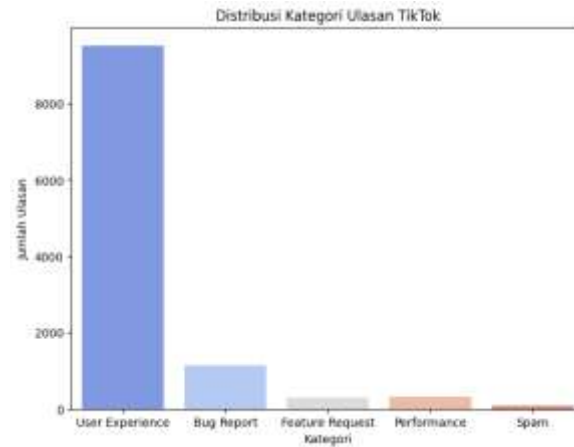
Distribusi sentimen dalam dataset mengklasifikasikan ulasan berdasarkan tiga kategori utama: positif, negatif, dan netral. Grafik yang ditampilkan pada Gambar 4 menunjukkan bahwa sebagian besar sentimen yang terkandung dalam ulasan adalah positif. Hal ini menunjukkan bahwa mayoritas pengguna TikTok memiliki pandangan yang baik terhadap video yang mereka tonton. Walaupun sebagian besar sentimen adalah positif, terdapat pula sejumlah sentimen negatif dan netral yang memberikan perspektif yang lebih beragam tentang penerimaan konten di platform tersebut.



Gambar 4 Distribusi Sentimen Ulasan Tiktok

### 3.2.3 Distribusi Kategori

Distribusi kategori ulasan TikTok pada Gambar 5 menunjukkan bahwa kategori “User Experience” mendominasi jumlah ulasan dengan selisih yang sangat signifikan dibandingkan kategori lainnya. Hal ini menandakan bahwa sebagian besar pengguna lebih banyak memberikan tanggapan terkait pengalaman penggunaan aplikasi secara keseluruhan, baik dari segi tampilan antarmuka, kemudahan navigasi, maupun kepuasan terhadap fitur-fitur utama. Sementara itu, kategori “Bug Report” berada di urutan kedua, yang menunjukkan masih adanya keluhan pengguna terkait kesalahan teknis atau gangguan sistem. Adapun kategori “Feature Request” dan “Performance” memiliki jumlah ulasan yang relatif kecil, menandakan bahwa sebagian pengguna memberikan saran terhadap peningkatan fitur dan kinerja aplikasi. Kategori “Spam” menjadi yang paling sedikit, menunjukkan bahwa ulasan yang tidak relevan atau bersifat promosi tidak mendominasi dalam dataset ini. Secara keseluruhan, distribusi ini mengindikasikan bahwa fokus utama pengguna TikTok lebih banyak pada pengalaman penggunaan daripada aspek teknis aplikasi.



Gambar 5 Distribusi Kategori Ulasan Tiktok

### 3.3 Hasil Pelatihan Model

Setelah model dilatih menggunakan berbagai metode, hasil evaluasi kinerja setiap model dirangkum dalam Tabel 1, yang memuat nilai Akurasi, Macro-F1, dan Weighted-F1 untuk CNN, LSTM, SVM, dan Logistic Regression. Tabel ini memberikan ringkasan umum tentang performa model dalam mengklasifikasikan sentimen ulasan TikTok.

Tabel 1 Hasil Evaluasi Kinerja Setiap Model

Model	Akurasi	Macro-F1	Weighted-F1
CNN	0.989526	0.940703	0.989705
LSTM	0.984750	0.879517	0.984215
SVM	0.978281	0.887202	0.977457
Logistic Regression	0.946858	0.656916	0.939063

Tabel 1 digunakan sebagai ringkasan untuk memberikan gambaran cepat tentang perbandingan kinerja model-model yang diuji. Dengan melihat nilai Akurasi, Macro-F1, dan Weighted-F1, kita bisa memperoleh insight tentang seberapa efektif masing-masing model ketika mengklasifikasikan sentimen dari ulasan TikTok.

#### 3.3.1 Tabel Akurasi

Setelah model dilatih menggunakan berbagai metode, hasil evaluasi kinerja setiap model dirangkum dalam Tabel 1 yang memuat nilai akurasi, macro-F1, dan weighted-F1 untuk CNN, LSTM, SVM, dan Logistic Regression. Berdasarkan kolom Akurasi pada Tabel 1, terlihat bahwa Convolutional Neural Network (CNN) mencapai akurasi tertinggi sebesar 0,989526, diikuti oleh LSTM sebesar 0,984750, SVM sebesar 0,978281, dan Logistic Regression sebesar 0,946858. Hasil ini menunjukkan bahwa CNN merupakan model yang paling efektif dalam mengklasifikasikan sentimen ulasan TikTok secara keseluruhan, karena mampu memprediksi label dengan benar pada hampir seluruh data uji dibandingkan model lainnya.

#### 3.3.2 Tabel Macro-F1

Macro-F1 digunakan untuk mengevaluasi kinerja model dalam menangani sentimen dengan distribusi kelas yang tidak seimbang. Nilai Macro-F1 dihitung dengan menghitung rata-rata F1-score dari setiap kelas, tanpa memperhatikan jumlah data di setiap kelas. Berdasarkan Tabel 4.3, CNN juga unggul dalam hal Macro-F1, dengan nilai 0.940703, yang berarti model ini dapat mengklasifikasikan sentimen positif, negatif, dan netral secara seimbang, meskipun kelas-kelas tersebut tidak tersebar merata. LSTM mengikuti dengan nilai 0.879517, sedangkan SVM dan Logistic Regression menunjukkan nilai yang lebih rendah, masing-masing 0.887202 dan 0.656916.

### 3.3.3 Tabel Weighted-F1

Weighted-F1 dihitung dengan cara yang mirip dengan Macro-F1, tetapi dengan mempertimbangkan ukuran setiap kelas dalam dataset. Berdasarkan Tabel 4.3, CNN menunjukkan performa terbaik pada nilai Weighted-F1, dengan nilai 0.989705, yang mengindikasikan bahwa model ini efektif dalam menangani ketidakseimbangan antara kelas-kelas sentimen. LSTM juga menunjukkan hasil yang baik, tetapi CNN tetap unggul dalam hal kinerja pada kelas minoritas, seperti yang tercermin pada nilai weighted-F1 yang lebih tinggi. SVM dan Logistic Regression menunjukkan nilai Weighted-F1 yang lebih rendah, masing-masing 0.977457 dan 0.939063.

### 3.4 Perbandingan Model

Untuk mengevaluasi kinerja berbagai model dalam analisis sentimen pada ulasan TikTok, digunakan classification report yang menampilkan metrik precision, recall, dan f1-score untuk setiap kelas, serta nilai *macro average* dan *weighted average*. Hasil dari classification report memberikan wawasan mendalam tentang bagaimana setiap model menangani berbagai kelas dalam dataset, serta seberapa efektif model dalam mengidentifikasi Bug Report, Feature Request, Performance, Spam, dan User Experience.

#### 3.4.1 CNN Classification Report

	precision	recall	f1-score	support
Bug Report	0.99	0.97	0.98	171
Feature Request	1.00	0.95	0.98	44
Performance	0.98	0.98	0.98	48
Spam	0.71	0.86	0.77	14
User Experience	0.99	0.99	0.99	1346
accuracy			0.99	1623
macro avg	0.93	0.95	0.94	1623
weighted avg	0.99	0.99	0.99	1623

Gambar 6 CNN Classification Report

Model Convolutional Neural Network (CNN) menunjukkan hasil luar biasa dalam mengklasifikasikan sentimen. *Precision* tertinggi dicapai oleh kelas Feature Request (1.00), yang artinya model sangat akurat dalam mengklasifikasikan kelas ini. User Experience juga memiliki precision yang tinggi (0.99). Recall yang sangat tinggi untuk User Experience (1.00) mengindikasikan bahwa model ini mampu mengidentifikasi hampir semua data User Experience dengan sangat baik. *F1-score* yang tinggi pada Performance dan Bug Report (0.98) menunjukkan keseimbangan yang efektif antara precision dan recall dalam mengklasifikasikan kedua kelas tersebut.

#### 3.4.2 LSTM Classification Report

=== LSTM Classification Report ===				
	precision	recall	f1-score	support
Bug Report	0.97	0.96	0.96	228
Feature Request	0.96	0.88	0.92	59
Performance	0.91	0.97	0.94	64
Spam	0.69	0.50	0.58	18
User Experience	0.99	1.00	0.99	1795
accuracy			0.98	2164
macro avg	0.91	0.86	0.88	2164
weighted avg	0.98	0.98	0.98	2164

Gambar 7 LSTM Classification Report

Model *Long Short-Term Memory* (LSTM) juga menunjukkan hasil yang baik dengan precision tinggi untuk Bug Report (0.97) dan User Experience (0.99). Recall untuk User Experience adalah 1.00, menunjukkan kemampuan model untuk mengidentifikasi seluruh data kelas ini dengan luar biasa. Namun, *precision* dan *recall* untuk Spam (0.69 dan 0.50) masing-masing menunjukkan bahwa model tersebut kesulitan dalam mengklasifikasikan kelas ini dengan tepat. F1-score untuk Performance dan Feature Request adalah 0.94 dan 0.92, menunjukkan keseimbangan yang baik.

### 3.4.3 SVM Classification Report

```
=== SVM Classification Report ===
              precision    recall  f1-score   support

   Bug Report         1.00      0.91      0.95     228
  Feature Request     0.94      0.83      0.88      59
    Performance     0.92      0.84      0.88      64
         Spam        0.92      0.61      0.73      18
  User Experience     0.98      1.00      0.99    1795

 accuracy                   0.98     2164
  macro avg                 0.95     2164
  weighted avg              0.98     2164
```

Gambar 8 SVM Classification Report

Model *Support Vector Machine* (SVM) menyajikan precision yang sangat baik, dengan nilai 1.00 untuk Bug Report dan User Experience, serta nilai 0.92 untuk Performance. Recall yang baik untuk User Experience (1.00) dan Performance (0.84) menunjukkan kemampuan model untuk mengidentifikasi kelas-kelas ini dengan baik. F1-score untuk Performance dan Feature Request adalah 0.88, yang menunjukkan keseimbangan yang sangat baik antara recall dan precision, meskipun kelas Spam menunjukkan recall yang lebih rendah (0.61).

### 3.4.4 Logistic Regression Classification Report

```
=== Logistic Regression Classification Report ===
              precision    recall  f1-score   support

   Bug Report         0.99      0.81      0.89     228
  Feature Request     0.97      0.61      0.75      59
    Performance     0.92      0.53      0.67      64
         Spam        0.00      0.00      0.00      18
  User Experience     0.94      1.00      0.97    1795

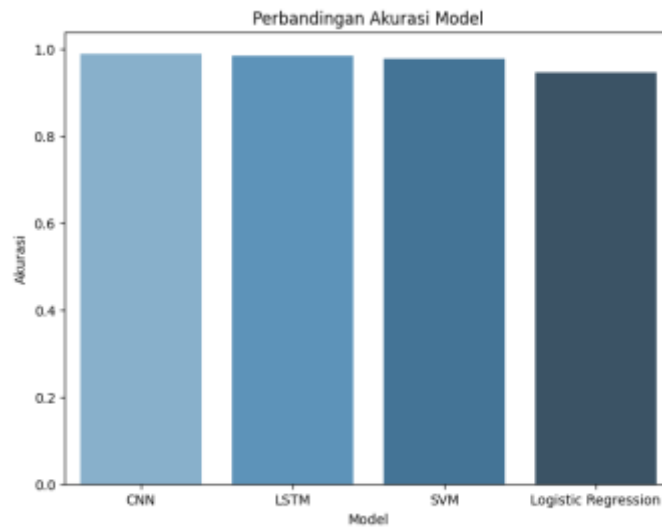
 accuracy                   0.95     2164
  macro avg                 0.77     2164
  weighted avg              0.94     2164
```

Gambar 9 LR Classification Report

Model *Logistic Regression* menunjukkan hasil yang beragam. Precision sangat tinggi untuk Bug Report (0.99) dan User Experience (0.94), namun recall untuk Feature Request (0.61) dan Performance (0.53) cukup rendah, yang menunjukkan bahwa model tersebut mengalami kesulitan dalam mengidentifikasi masing-masing kelas dengan baik. F1-score untuk Spam adalah 0.00, yang menunjukkan bahwa model ini hampir tidak dapat mengidentifikasi kelas tersebut. Nilai accuracy untuk *Logistic Regression* adalah 0.95, yang menunjukkan performa yang cukup baik meskipun ada kesulitan dalam beberapa kelas.

### 3.5 Visualisasi

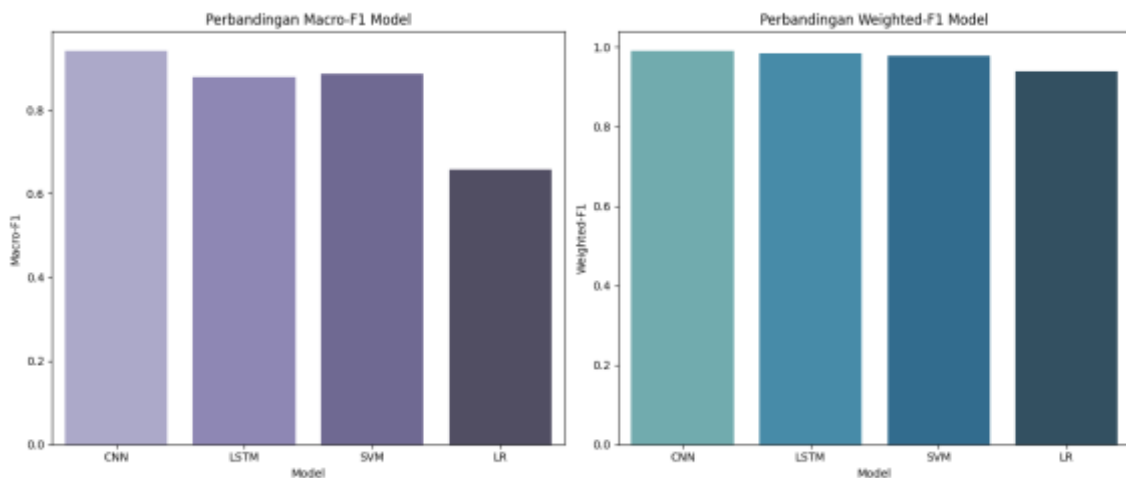
#### 3.5.1 Grafik Akurasi Semua Model



Gambar 10 Grafik Akurasi Semua Model

Grafik yang ditampilkan pada Gambar 10 memberikan gambaran perbandingan akurasi dari semua model yang diuji. CNN menunjukkan akurasi tertinggi, diikuti oleh model-model lainnya, yang memberikan wawasan tentang seberapa efektif setiap model dalam menganalisis sentimen pada ulasan TikTok.

#### 3.5.2 Grafik Macro-F1 dan Weighted-F1



Gambar 11 Grafik Macro-F1 dan Weighted-F1

Gambar 11 menunjukkan perbandingan antara nilai macro-F1 dan weighted-F1 dari setiap model. CNN menunjukkan performa terbaik pada kedua metrik ini, yang menunjukkan bahwa model ini tidak hanya akurat, tetapi juga seimbang dalam mengklasifikasikan setiap kelas sentimen.

## 4. Kesimpulan

Hasil analisis menunjukkan bahwa Convolutional Neural Network (CNN) adalah sebuah model yang paling efektif dalam analisis sentimen pada data ulasan TikTok. Model CNN unggul dalam hal akurasi, macro-F1, dan weighted-F1, serta mampu menangani teks pendek dengan baik, yang merupakan karakteristik utama data TikTok. Sementara itu, model LSTM, SVM, dan Logistic Regression menunjukkan performa yang lebih rendah karena kesulitan dalam menangani urutan yang tidak dominan dan hubungan non-linear pada data. Dengan menggunakan

data sebanyak 11.368 ulasan setelah preprocessing dari total 19.569 ulasan, hasil ini memberikan gambaran yang jelas bahwa CNN adalah pilihan terbaik untuk klasifikasi sentimen pada platform media sosial seperti TikTok, yang memiliki banyak teks pendek dan berbagai macam sentimen. Untuk penelitian lebih lanjut, disarankan untuk mengumpulkan dataset yang lebih beragam yang mencakup ulasan dari beberapa platform media sosial lainnya, seperti Instagram, Twitter, atau YouTube, guna meningkatkan generalisasi model. Selain itu, dapat dipertimbangkan penerapan transfer learning dengan model pre-trained seperti BERT atau RoBERTa, yang dapat meningkatkan kinerja model, terutama pada teks yang lebih panjang dan variatif. Mengingat adanya ketidakseimbangan kelas dalam data ulasan, penelitian selanjutnya sebaiknya menggunakan teknik seperti oversampling, undersampling, atau cost-sensitive learning untuk mengatasi masalah ini, sehingga model dapat mengklasifikasikan kelas minoritas dengan lebih baik. Selain itu, penelitian lebih lanjut dapat meningkatkan kualitas preprocessing teks dengan menggunakan word embeddings seperti Word2Vec atau GloVe untuk memberikan representasi kata yang lebih kaya, serta mengevaluasi model menggunakan metrik lain seperti precision-recall curves dan ROC-AUC untuk mengukur kinerja pada kelas minoritas. Penelitian juga dapat mengembangkan model multimodal yang menggabungkan teks dengan elemen visual (gambar/video) dari ulasan TikTok untuk memperoleh hasil yang lebih holistik dan akurat dalam analisis sentimen.

## Referensi

- [1] M. A. H. Ahmad Asmawi, P. Isawasan, and L. Shamugam, "Exploring sentiment trends in TikTok comments using GPT for influencer content strategy," *e-Academia J.*, 2025, [Online]. Available: <https://ir.uitm.edu.my/id/eprint/118867/1/118867.pdf>
- [2] R. Kansal and C. Diwaker, "Efficiency determination of various machine learning techniques for sentiment analysis on social media platforms," *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 1, pp. 1234–1240, 2025, [Online]. Available: <https://etasr.com/index.php/ETASR/article/download/11158/5339>
- [3] D. D. N. Cahyo, R. Handayani, and V. B. Lestari, "Sentiment analysis of public opinion on online gambling through social media using convolutional neural network," *J. Informatics Telecommun. Eng.*, vol. 9, no. 2, pp. 115–124, 2025, [Online]. Available: <https://ojs.uma.ac.id/index.php/jite/article/download/15024/6502>
- [4] S. Kumiawan, A. S. Pramayoga, and Y. F. Ashari, "Sentiment analysis in coastal tourism: A systematic literature review," *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 3, pp. 2311–2321, 2025, [Online]. Available: <https://etasr.com/index.php/ETASR/article/download/14644/5986>
- [5] M. Dixit, D. Sharma, and V. Singh, "Video sentiment analysis on social media using an advanced VADER technique," *J. Data Sci. Anal.*, 2025, [Online]. Available: <https://doi.org/10.1007/s41060-025-00849-1>
- [6] L. Zhiming, H. Huijian, and L. Zongwei, "Predicting the Shanghai Composite Index using Chinese TikTok self-media data and machine learning model in China," *Discret. Dyn. Nat. Soc.*, vol. 2024, pp. 1–12, 2024, [Online]. Available: <https://doi.org/10.1155/ddns/7201831>
- [7] Q. Shen, S. Han, Y. Han, and X. Chen, "User review analysis of dating apps based on text mining," *PLoS One*, vol. 18, no. 4, p. e0283896, 2023.
- [8] J. Dąbrowski, E. Letier, A. Perini, and A. Susi, "Analysing app reviews for software engineering: A systematic literature review," *Empir. Softw. Eng.*, vol. 27, no. 2, p. 43, 2022.
- [9] M. Das and P. J. A. Alphonse, "A comparative study on TF-IDF feature weighting method and its analysis using unstructured dataset." 2023. [Online]. Available: [arxiv:2308.04037](https://arxiv.org/abs/2308.04037)
- [10] A. H. Dani, E. Y. Puspaningrum, and R. Mumpuni, "Studi performa TF-IDF dan Word2Vec pada analisis sentimen cyberbullying," *Router J. Tek. Inform. dan Terap.*, vol. 2, no. 2, pp. 94–106, 2024.
- [11] V. B. Lestari and C. A. Hutagalung, "Evaluation of TF-IDF extraction techniques in sentiment analysis of Indonesian-language marketplaces using SVM, logistic regression, and naive Bayes," *J-KOMA J. Comput. Sci. Appl.*, vol. 8, no. 1, pp. 36–44, 2025.
- [12] A. Hermawan, I. Jowensen, and J. Junaedi, "Implementasi text-mining untuk analisis sentimen pada Twitter dengan algoritma support vector machine," *JST (Jurnal Sains dan Teknol.)*, vol. 12, no. 1, pp. 129–137, 2023.
- [13] M. Prasetya, M. Wulandari, and S. A. Nikmah, "Implementasi NLP (Natural Language Processing) Dasar pada Analisis Sentiment Review Spotify BT - Seminar Nasional Teknologi & Sains," 2024, pp. 145–153.
- [14] J. Opitz, "A closer look at classification evaluation metrics and a critical reflection of common evaluation practice," *Trans. Assoc. Comput. Linguist.*, vol. 12, pp. 820–836, 2024.

- [15] W. Zhao, R. Singh, T. Joshi, A. Sudjianto, and V. N. Nair, "Self-interpretable convolutional neural networks for text classification." 2021. [Online]. Available: arxiv:2105.08589
- [16] S. Soni, S. S. Chouhan, and S. S. Rathore, "TextConvoNet: A convolutional neural network-based architecture for text classification," *Appl. Intell.*, vol. 53, no. 11, pp. 14249–14268, 2023.
- [17] H. Cao, "Recent advances in text embedding: A comprehensive review of top-performing methods on the MTEB benchmark." 2024. [Online]. Available: arxiv:2406.01607