

ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG PRODUK EIGER MENGUNAKAN MODEL DEEP LEARNING RNN, CNN, DAN LSTM

SENTIMENT ANALYSIS OF YOUTUBE COMMENTS ON EIGER PRODUCT USING RNN, CNN, and LSTM DEEP LEARNING MODELS

Maichal Jem's Hard Aderzon Putra¹, Magdalena A. Ineke Pakereng²

Teknologi dan Informasi, Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga, Indonesia¹

Teknologi dan Informasi, Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga, Indonesia²

672022335@student.uksw.edu¹

ABSTRACT

This study analyzes public sentiment toward Eiger products based on comments from an independent third-party YouTube channel using three Deep Learning models: Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM). A dataset of 2,015 unique comments was collected via the YouTube Data API v3, preprocessed, and labeled using an automatic lexicon-based method. The results revealed high data imbalance: Neutral at 74.74%, Positive at 19.50%, and Negative at 5.76%. The dataset was randomly split into an 80:20 ratio for training and testing. Experimental results demonstrate that the 1D CNN model achieved the best performance with an accuracy of 94.79% and balanced F1-Scores across all classes (Positive: 0.97, Negative: 0.87, Neutral: 0.88). Conversely, the RNN and LSTM models stagnated at a pseudo-accuracy of 75% with F1-Scores of 0.00 for minority classes. Confusion Matrix analysis proves that RNN and LSTM suffered from classification paralysis due to the majority class bias phenomenon, blindly predicting all samples into the testing data's local majority class (the Positive class, which accounted for 74.69% of test samples). This study concludes that the 1D CNN architecture is significantly more robust in extracting local sentiment features from short, imbalanced social media texts compared to temporal sequential-based architectures.

Keywords: Sentiment Analysis, YouTube, Eiger, Deep Learning, CNN, RNN, LSTM, Majority Class Bias

ABSTRAK

Penelitian ini menganalisis sentimen publik terhadap produk Eiger berdasarkan komentar pada kanal YouTube pihak ketiga independen menggunakan tiga model Deep Learning: Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), dan Long Short-Term Memory (LSTM). Dataset sebanyak 2.015 komentar unik dikumpulkan melalui YouTube Data API v3, lalu diproses dan diberi label berbasis kamus (lexicon-based). Hasil pelabelan menunjukkan ketidakseimbangan data yang tinggi: Netral 74,74%, Positif 19,50%, dan Negatif 5,76%. Dataset dibagi acak dengan proporsi 80:20 untuk data latih dan uji. Hasil eksperimen menunjukkan model CNN 1D menghasilkan performa terbaik dengan akurasi mencapai 94,79% serta nilai F1-Score yang seimbang di seluruh kelas (Positif: 0,97, Negatif: 0,87, Netral: 0,88). Sebaliknya, model RNN dan LSTM tertahan di angka akurasi semu sebesar 75% dengan F1-Score 0,00 pada kelas minoritas. Analisis Confusion Matrix membuktikan RNN dan LSTM mengalami kelumpuhan klasifikasi akibat fenomena majority class bias, di mana model memprediksi seluruh sampel ke kelas mayoritas lokal data uji (kelas Positif berproporsi 74,69%). Penelitian ini menyimpulkan bahwa arsitektur CNN 1D jauh lebih tangguh dalam mengekstraksi fitur lokal penentu sentimen pada teks media sosial yang pendek dan tidak seimbang dibandingkan arsitektur berbasis sekuensial temporal.

Kata kunci: Analisis Sentimen, YouTube, Eiger, Deep Learning, CNN, RNN, LSTM, Majority Class

PENDAHULUAN

Platform YouTube kini berperan sebagai salah satu ruang utama bagi pengguna untuk memberikan opini mengenai berbagai produk, termasuk produk perlengkapan *outdoor* seperti Eiger. Komentar-komentar yang dihasilkan oleh pengguna dapat menjadi sumber informasi penting bagi perusahaan untuk memahami persepsi, kepuasan, serta respons publik

secara langsung dan *real-time*. Analisis sentimen terhadap komentar-komentar tersebut memungkinkan identifikasi kecenderungan opini yang muncul, sehingga dapat dimanfaatkan sebagai bahan evaluasi kualitas produk maupun strategi pemasaran. Seiring meningkatnya jumlah komentar yang dihasilkan pengguna, dibutuhkan metode analisis yang mampu bekerja secara sistematis dan akurat

terhadap data berukuran besar Ahuja, H., Kaur, N., Kumar, P. et al. 2023.

Selain itu, Surana, P., Vispute, S., Kalbhor, M., Vairagar, S. et al. (2025) menunjukkan bahwa perkembangan teknik analisis sentimen berbasis *machine learning* dan *deep learning* memberikan peluang yang lebih kuat dalam mengekstraksi pola opini dari teks. Penelitian sebelumnya menunjukkan bahwa komentar YouTube merupakan sumber data yang kaya untuk memahami reaksi masyarakat terhadap video, isu, maupun produk tertentu. Menurut Liyih, A., Anagaw, S., Yibeyin, M. & Tehone, Y. 2024, Metode *deep learning* secara khusus menawarkan kemampuan dalam menangkap struktur bahasa alami dan konteks kata dalam kalimat, sehingga meningkatkan kualitas prediksi sentimen dibandingkan pendekatan tradisional. Dengan demikian, analisis sentimen berbasis deep learning semakin relevan untuk diterapkan pada data komentar YouTube yang bersifat padat dan beragam.

Sebuah studi lain yang dilakukan oleh Bindhumol, M., Singh, T. & Patra, P. 2024 menunjukkan bahwa beberapa *model deep learning* telah digunakan secara luas dalam penelitian sentimen, seperti *Recurrent Neural Network* (RNN), *Convolutional Neural Network* (CNN), dan *Long Short-Term Memory* (LSTM). Penggunaan ketiga model ini di berbagai penelitian menunjukkan performa yang kompetitif dalam memproses data komentar pengguna, termasuk komentar pada platform YouTube.

Penelitian mengenai analisis sentimen pada komentar YouTube telah dilakukan dalam berbagai konteks, seperti ulasan video edukasi, hiburan, film, hingga isu sosial. Penggunaan *Natural Language Processing* (NLP) dan model klasifikasi telah menghasilkan temuan yang memperlihatkan bahwa karakteristik komentar di YouTube cenderung dipengaruhi oleh konteks video serta pengalaman pengguna (Visariya, D., Prajapati, P. & Bhatt, N. 2025). Selain itu,

Ercikan, Y. S. & Elmasry, W. (2025) menunjukkan bahwa terdapat pendekatan yang memanfaatkan model NLP kustom untuk meningkatkan akurasi analisis, terutama ketika data yang digunakan memiliki pola bahasa yang kompleks dan informal sebagaimana umum ditemui pada komentar YouTube. Di sisi lain, Furqoni, W., Sindura, I. P., Rahayuningtyas, I. P. et al. (2025) menerapkan model CNN dan RNN secara bersamaan untuk menghasilkan kombinasi fitur yang lebih kaya dan akurat.

Untuk memetakan posisi penelitian, sejumlah kajian terdahulu telah menerapkan model-model ini pada berbagai domain platform. Sementara itu, penelitian yang dilakukan oleh Utami, H. 2022 membahas analisis sentimen pada ulasan aplikasi Shopee menggunakan metode *Recurrent Neural Network* (RNN) menunjukkan bahwa model berbasis jaringan berulang cukup efektif dalam mempelajari urutan kata sehingga akurat dalam mengklasifikasikan sentimen ulasan pengguna. Irfani, M. & Khomsah, S. menegaskan pentingnya pemodelan urutan dalam data teks untuk memaksimalkan performa analisis sentimen. Pendekatan serupa digunakan dalam penelitian lain yang menerapkan metode CNN dan *Word2Vec* pada penilaian pembelajaran EDOM, menunjukkan kemampuan CNN dalam mengekstraksi fitur penting dari *embedding* kata dan menghasilkan klasifikasi sentimen yang baik. Model LSTM juga telah banyak digunakan dalam berbagai analisis sentimen. Sebuah penelitian yang menerapkan algoritma LSTM pada media sosial menunjukkan bahwa model ini unggul dalam menangkap dependensi jangka panjang antar kata sehingga menghasilkan prediksi sentimen yang stabil meskipun data mengandung struktur kalimat yang kompleks (Mudding, A. A. & Karim, A. A. A. 2022). Kinerja LSTM dalam memproses teks berkualitas rendah atau informal menjadikan model ini sangat relevan untuk diterapkan pada

komentar YouTube yang sering kali tidak mengikuti kaidah penulisan formal.

Selain itu, Berliana, A. S. & Mustikasari, M. 2024 yang menganalisis sentimen ulasan aplikasi Jakartanotebook menggunakan metode RNN menunjukkan peningkatan akurasi ketika model dilengkapi dengan pra-pemrosesan data yang baik, seperti pembersihan teks, penghapusan *stop words*, dan normalisasi kata. Hal ini memperkuat bahwa kualitas tahap pra-pemrosesan memegang peranan penting dalam keberhasilan model deep learning dalam klasifikasi sentimen. Metode LSTM juga telah digunakan dalam penelitian mengenai opini publik terhadap isu sosial, seperti penundaan Pemilu 2024. Hasil penelitian tersebut menunjukkan bahwa model LSTM memberikan performa yang tinggi dalam mengelompokkan opini yang berasal dari teks pendek seperti tweet (Prasetyo, P. A. P. 2024). Sifat komentar YouTube yang cenderung singkat menjadikan hasil penelitian tersebut relevan sebagai acuan.

Penelitian lain yang menganalisis komentar aplikasi Enterkomputer di Google Play menggunakan metode RNN menunjukkan bahwa model tersebut cocok digunakan untuk mendeteksi pola emosi pengguna dalam ulasan produk teknologi Apriliani, S. F. & Khorunnisa, N. L. 2024.

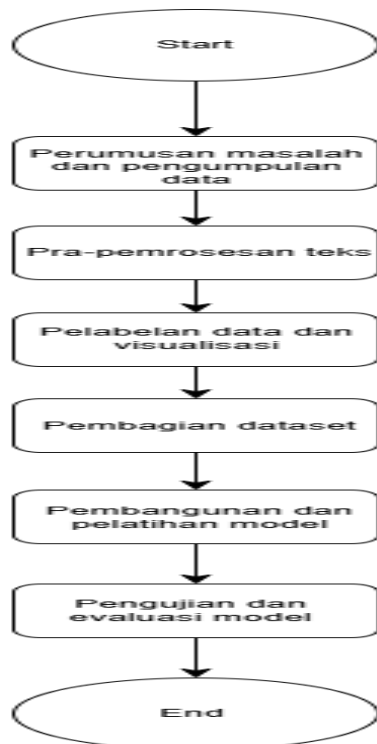
Meskipun ketiga model tersebut memiliki performa yang baik pada dataset standar yang seimbang, tantangan utama dalam klasifikasi teks komentar media sosial adalah tingginya tingkat ketidakseimbangan data (*imbalanced data*) dan penggunaan bahasa yang sangat informal serta pendek. Model berbasis urutan seperti RNN dan LSTM sering kali mengalami bias kelas mayoritas (*majority class bias*) ketika dilatih menggunakan dataset yang didominasi oleh satu sentimen (misalnya sentimen positif), sehingga cenderung mengalami kegagalan total dalam mengenali sentimen minoritas (negatif dan netral). Evaluasi performa model melalui metrik komprehensif (*Precision*, *Recall*, *F1-Score*) dan

Confusion Matrix sangat diperlukan untuk melihat apakah model benar-benar mampu mengenali pola teks secara objektif atau hanya terjebak pada akurasi semu (*pseudo-accuracy*).

Berdasarkan permasalahan tersebut, penelitian ini berfokus pada analisis sentimen komentar YouTube mengenai produk Eiger dengan menerapkan tiga model deep learning, yaitu RNN, CNN, dan LSTM. Penelitian ini menguji efektivitas ketiga model melalui tahapan pelaksanaan yang detail mulai dari ekstraksi data menggunakan YouTube API, pemrosesan teks (*preprocessing*), ekstraksi fitur berbasis *embedding*, hingga pengujian performa arsitektur. Hasil eksperimen dievaluasi secara ketat menggunakan fungsi matematis metrik akurasi dan klasifikasi. Penelitian ini diharapkan dapat memberikan gambaran empiris mengenai kecenderungan opini terhadap produk Eiger, serta mengidentifikasi ketahanan masing-masing arsitektur *deep learning* dalam menghadapi karakteristik data media sosial yang tidak seimbang.

METODE

Metodologi dalam penelitian ini disusun secara sistematis untuk memastikan integritas data dan validitas model yang dibangun, di mana alur penelitian terbagi ke dalam enam tahapan utama yang saling berkesinambungan.



Gambar 1. Flowchart Penelitian

2.1 Pengumpulan Data (Data Crawling)

Tahap pertama adalah pengumpulan data komentar dari platform YouTube pada kanal pihak ketiga independen bernama "Bang Ikiw TV" yang mengulas produk perlengkapan outdoor Eiger. Proses penambahan data dilakukan secara otomatis memanfaatkan dokumen API resmi via YouTube Data API v3 menggunakan pustaka (library) google-api-python-client pada lingkungan Python.

Kriteria pencarian difokuskan pada video ulasan produk Eiger dengan mengekstraksi parameter textDisplay dari objek komentar. Melalui proses crawling ini, berhasil dikumpulkan sebanyak 2.017 komentar mentah (raw data) dalam format JSON yang kemudian dikonversi menjadi berkas tabulasi (DataFrame).

2.2 Pra-Pemrosesan Teks (Text Preprocessing)

Tahap kedua mencakup pra-pemrosesan teks untuk mengubah data teks komentar YouTube yang bersifat tidak terstruktur dan informal menjadi format bersih yang siap diproses oleh lapisan *embedding*. Berdasarkan implementasi

kode program, langkah-langkah *preprocessing* yang dilakukan secara berurutan adalah sebagai berikut:

1. Cleaning (Pembersih Teks)

Menghapus elemen naskah yang tidak memiliki bobot informatif bagi analisis sentimen menggunakan fungsi ekspresi reguler (*re.sub*).

2. Case Folding

Mengubah seluruh karakter huruf yang tersisa di dalam teks komentar menjadi huruf kecil (*lowercase*) agar seragam dan menghindari duplikasi makna akibat perbedaan kapitalisasi kata.

3. Stemming Bahasa Indonesia

Mengubah kata-kata berimbuhan menjadi kata dasarnya. Proses ini sangat krusial untuk teks berbahasa Indonesia dan diimplementasikan memanfaatkan pustaka Sastrawi melalui fungsi *StemmerFactory()*. Contohnya, kata "mengecewakan" akan direduksi menjadi kata dasar "kecewa".

4. Tokenization dan Padding

Memecah untaian kalimat komentar menjadi token unit kata tunggal (*list of words*). Selanjutnya, diterapkan teknik padding menggunakan nilai nol (*zero-padding*) di akhir urutan (*post-padding*) untuk memastikan seluruh sampel komentar memiliki panjang dimensi input (N_{max}) yang seragam sebelum memasuki arsitektur *Deep Learning*.

2.3 Pelabelan Berbasis Kamus (Lexicon-Based Labeling)

Mengingat besarnya volume data tanpa label, tahap ketiga menerapkan pendekatan otomatis berbasis kamus kata kunci (*lexicon-based*) untuk menentukan polaritas sentimen. Setiap kata dalam token komentar yang telah diubah ke kata dasar dicocokkan dengan basis data kamus sentimen bahasa Indonesia yang memiliki bobot nilai polaritas. Akumulasi skor polaritas (S) dihitung untuk menentukan kelas akhir komentar:

$$\text{Sentimen} = \begin{cases} \text{Positif}, & \text{Jika } S > 0 \\ \text{Negatif}, & \text{Jika } S < 0 \\ \text{Netral}, & \text{Jika } S = 0 \end{cases}$$

Setelah label terbentuk, visualisasi kata menggunakan *Word Cloud* dilakukan untuk mengidentifikasi kata kunci dominan pada masing-masing kelas guna memvalidasi relevansi data secara visual sebelum fase pemodelan.

2.4 Pembagian Data (Data Splitting)

Tahap keempat adalah pembagian dataset untuk menghindari *overfitting* dan menjamin objektivitas pengujian. Sebanyak 2.017 data komentar yang telah bersih dan berlabel dibagi menggunakan proporsi 80:20 menggunakan metode *stratified sampling* untuk menjaga distribusi representasi kelas yang riil.

1. Data Latih (Training Data)

Sebanyak 80% (1.614 sampel) dialokasikan untuk memperbarui bobot (*weights*) internal arsitektur jaringan selama proses *training*.

2. Data Uji (Testing Data)

Sebanyak 20% (403 sampel) diisolasi sepenuhnya dan hanya digunakan pada fase evaluasi akhir performa generalisasi model.

2.5 Arsitektur dan Pembangunan Model Deep Learning

Tahap kelima adalah mengumpulkan representasi teks bersih ke dalam tiga arsitektur *deep learning* yang diuji secara paralel: CNN, RNN, dan LSTM. Ketiganya dikonfigurasi dengan parameter seragam demi menjamin perbandingan yang adil: lapisan Embedding berdimensi 128, 128 *unit hidden layer*, *optimizer Adam*, dan dilatih selama 10 epoch.

Secara matematis, mekanisme pemrosesan internal inti dari masing-masing model yang diuji didefinisikan sebagai berikut:

2.5.1 Convolutional Neural Network (CNN) 1D

Lapisan konvolusi 1D mengekstraksi fitur spasial lokal (*n-gram*) dari vektor embedding teks tanpa terlalu bergantung pada urutan waktu keseluruhan kalimat. Operasi konvolusi untuk menghasilkan

representasi fitur c_i dari jendela kata (*window*) berukuran k dirumuskan sebagai:

$$c_i = f(\mathbf{w} \cdot \mathbf{x}_{i:i+k-1} + b)$$

Di mana $\mathbf{w}$ melambangkan matriks bobot filter (*kernel weight*), $\mathbf{x}_{i:i+k-1}$ mewakili konkatenasi vektor *embedding* kata berturutan, b adalah nilai bias, dan f menyatakan fungsi aktivasi *Rectified Linear Unit* (ReLU).

2.5.2 Recurrent Neural Network (RNN) Standard

RNN memproses data secara sekuensial langkah demi langkah waktu (*time-step t*) dengan mempertahankan memori masa lalu berupa *hidden state* (h_t). Persamaan matematis untuk memperbarui *hidden state* pada waktu t didasarkan pada input saat ini (x_t) dan *hidden state* sebelumnya (h_{t-1}):

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t + b_h)$$

Di mana W_{hh} dan W_{xh} adalah matriks bobot pengali, dan b_h merupakan vektor bias.

2.5.3 Long Short-Term Memory (LSTM)

Untuk mengatasi masalah hilangnya gradien (*vanishing gradient*) pada RNN sekuensial, LSTM mengenalkan *komponen cell state* (C_t) yang diatur oleh tiga gerbang (*gates*): *forget gate* (f_t), *input gate* (i_t), dan *output gate* (o_t). Persamaan matematis internal unit LSTM adalah:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \odot \tanh(C_t)$$

Keterangan: σ melambangkan fungsi aktivasi sigmoid, $\odot$ adalah operasi perkalian elemen (*element-wise multiplication*), dan h_t adalah *hidden state* akhir yang diteruskan ke lapisan klasifikasi *Softmax*.

2.6 Metode Pengujian Validasi dan Evaluasi Performa

Tahap keenam ditutup dengan menguji ketahanan model menggunakan 403 sampel data uji yang telah diisolasi. Metrik evaluasi dihitung secara teoritis berdasarkan visualisasi tabel *Confusion Matrix* yang menghasilkan nilai elemen *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

Formulasi matematika standar yang digunakan untuk mengukur nilai tingkat akurasi, presisi, *recall*, dan *F1-score* didefinisikan melalui persamaan berikut:

1. Accuracy

Mengukur total prediksi benar dari keseluruhan data uji

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision

Mengukur ketepatan antara total prediksi kelas positif yang benar dengan seluruh hasil yang diprediksi positif oleh model.

$$\text{Precision} = \frac{TP}{TP + FP}$$

3. Recall

Mengukur kemampuan model dalam menemukan kembali seluruh sampel data yang sejatinya termasuk kelas positif.

$$\text{Recall} = \frac{TP}{TP + FN}$$

4. F1-Score

Nilai rata-rata harmonis antara metrik Precision dan Recall untuk mengukur keseimbangan performa model, terutama pada kondisi dataset tidak seimbang (*imbalanced data*).

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

HASIL DAN PEMBAHASAN

3.1 Deskripsi dan Visualisasi Dataset

Tahap awal dalam memaparkan hasil penelitian ini adalah mendeskripsikan secara komprehensif karakteristik dataset yang digunakan. Data primer dalam penelitian ini diperoleh melalui proses *crawling* dan *extraction* komentar pada platform sosial media YouTube, dengan fokus spesifik pada video ulasan (*review*) produk perlengkapan *outdoor* bermerek Eiger pada kanal YouTube pihak ketiga independen, yaitu "Bang Ikiw TV". Pemilihan sumber data ini bertujuan untuk menjaga objektivitas akademis karena komentar-komentar dari audiens cenderung organik, kritis, dan jujur dibandingkan kanal resmi perusahaan yang bersifat promosional. Melalui parameter pemrosesan tersebut, berhasil dihimpun sebanyak 2.015 data komentar unik setelah dibersihkan dari data duplikat dan kosong. Setelah melalui tahapan pra-pemrosesan teks dan pelabelan otomatis berbasis kamus (*lexicon-based labeling*), ditemukan bahwa karakteristik sebaran data memiliki tingkat ketimpangan kelas (*imbalanced data*) yang cukup signifikan. Sebaran frekuensi dan persentase untuk masing-masing kelas sentimen disajikan secara mendalam pada Tabel 1.

Tabel 1. Distribusi Frekuensi Label Sentimen Komentar Eiger

Kelas Sentimen	Jumlah Data Latih (80%)	Jumlah Data Uji (20%)	Total Data (100%)	Persentase (%)
Netral	1.427	79	1.506	74,74%
Positif	92	301	393	19,50%
Negatif	93	23	116	5,76%
Total	1.612	403	2.015	100%

Berdasarkan Tabel 1, terlihat jelas bahwa opini publik didominasi secara mutlak oleh kelas sentimen Netral, yakni sebanyak 1.506 komentar atau setara dengan 74,74% dari keseluruhan dataset. Sentimen Positif menempati posisi kedua dengan representasi sebesar 19,50% (393 komentar), sementara sentimen Negatif bertindak sebagai kelas minoritas ekstrem dengan kuantitas hanya sebesar 5,76% (116 komentar). Ketimpangan sebaran kelas secara global ini menjadi data konkret yang memicu tantangan tersendiri bagi model

putaran dan *Batch Size* sebesar 32. Berdasarkan hasil eksperimen pelatihan yang dilakukan pada tahap ini, didapatkan nilai akurasi dan *loss* awal untuk masing-masing arsitektur sebagaimana dirangkum dalam Tabel 3.

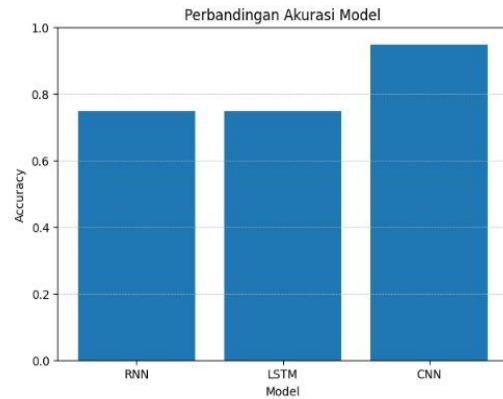
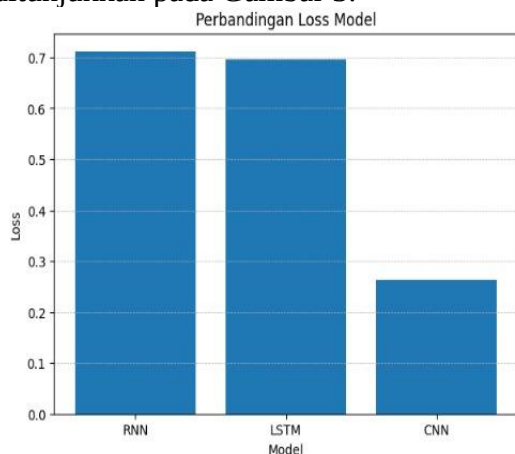
Tabel 3. Proses Pelatihan

Arsitektur Model	Akurasi Pelatihan	Loss (Tingkat Kesalahan)
RNN	0.7469	0.7112
LSTM	0.7469	0.6957
CNN	0.9479	0.2628

Berdasarkan Tabel 3, terlihat perbedaan performa yang cukup signifikan pada tahap pelatihan awal. Model CNN menunjukkan performa yang paling unggul dengan tingkat akurasi mencapai 0.9479 dan nilai *loss* terendah sebesar 0.2628. Sementara itu, model RNN dan LSTM memiliki nilai akurasi yang identik yaitu 0.7469, namun LSTM memiliki nilai *loss* (0.6957) yang sedikit lebih rendah dibandingkan RNN (0.7112), yang menunjukkan kemampuan LSTM dalam memproses data dengan tingkat kesalahan yang lebih kecil pada dataset komentar Eiger ini.

3.4 Analisis Performa Pelatihan Model (*Loss* dan *Accuracy*)

Setelah dilakukan proses pelatihan terhadap ketiga model, dilakukan visualisasi perbandingan hasil akhir untuk mempermudah analisis performa. Visualisasi ini disajikan dalam dua grafik batang, yaitu perbandingan akurasi dan perbandingan *loss*, sebagaimana ditunjukkan pada Gambar 3.



Gambar 3. Perbandingan Akurasi dan Loss Model

Berdasarkan visualisasi pada Gambar 3, dapat ditarik beberapa analisis penting mengenai kinerja masing-masing arsitektur dalam mengolah data komentar YouTube Eiger. Pada grafik Perbandingan Akurasi Model, terlihat secara jelas bahwa model CNN mendominasi dengan tingkat akurasi yang paling tinggi, mendekati angka 1.0 (94,79%). Sementara itu, model RNN dan LSTM memiliki performa yang setara (74,69%), yang menunjukkan bahwa pada dataset ini, keduanya memiliki kemampuan klasifikasi yang identik dalam memproses teks."

Analisis selanjutnya terlihat pada grafik Perbandingan *Loss* Model. Nilai *loss* mencerminkan tingkat kesalahan prediksi model; semakin rendah nilai *loss*, maka semakin baik model dalam mengenali pola data. Model CNN kembali menunjukkan keunggulannya dengan nilai *loss* yang sangat rendah (0,2628) dibandingkan model lainnya. Menariknya, meskipun RNN dan LSTM memiliki akurasi yang sama, grafik menunjukkan bahwa LSTM memiliki nilai *loss* yang sedikit lebih rendah daripada RNN. Hal ini membuktikan bahwa struktur gate pada LSTM bekerja sedikit lebih efektif dalam meminimalkan kesalahan prediksi dibandingkan struktur RNN konvensional.

Secara keseluruhan, hasil visualisasi ini mengonfirmasi bahwa arsitektur CNN adalah yang paling adaptif dan efisien dalam mengekstraksi informasi penting dari komentar-komentar pendek yang ada pada dataset penelitian ini.

3.5 Evaluasi dan Perbandingan Kinerja Model (Classification Report)

Tahap evaluasi mendalam dilakukan untuk melihat performa model pada setiap kelas sentimen (Negatif, Positif, dan Netral). Evaluasi ini menggunakan testing data sebanyak 403 sampel. Metrik yang diukur meliputi *Precision* (ketepatan prediksi), *Recall* (keberhasilan model menemukan kembali informasi), dan *F1-Score* (keseimbangan antara *Precision* dan *Recall*). Hasil evaluasi ketiga model disajikan secara mendalam pada Tabel 4.

Tabel 4. Hasil Classification report model RNN, LSTM, dan CNN

Arsitektur	Sentimen	Precision	Recall
RNN	Negatif (0)	0.00	0.00
	Positif (1)	0.75	1.00
	Netral (2)	0.00	0.00
LSTM	Negatif (0)	0.00	0.00
	Positif (1)	0.75	1.00
	Netral (2)	0.00	0.00
CNN	Negatif (0)	0.87	0.87
	Positif (1)	0.96	0.98
	Netral (2)	0.92	0.85

Berdasarkan Tabel 4, terdapat perbedaan kinerja yang sangat kontras antara model berbasis *recurrent* (RNN & LSTM) dengan model berbasis *convolutional* (CNN). Model RNN dan LSTM menunjukkan kegagalan total dalam mengklasifikasikan sentimen negatif (0) dan netral (2), yang ditunjukkan dengan nilai 0.00 pada seluruh metrik di kelas tersebut. Kedua model ini hanya mampu mengenali sentimen positif (1), sehingga nilai akurasi akhir sebesar 75% yang didapat semata-mata berasal dari dominasi kelas positif pada subset data uji.

Perbedaan kecenderungan bias model yang mengarah pada kelas positif (bukan kelas netral seperti pada sebaran global dataset awal) dapat dijelaskan melalui analisis nilai *support* pada data pengujian. Melalui proses pembagian data acak, sebanyak 301 dari total 403 sampel data uji secara kebetulan diisi oleh sentimen positif, yang setara dengan 74,69% dari

keseluruhan data uji. Sebaliknya, kelas netral hanya terwakili oleh 79 sampel (19,6%) dan kelas negatif sebanyak 23 sampel (5,7%). Ketimpangan lokal yang ekstrem pada subset data uji inilah yang memicu terjadinya fenomena *majority class bias* pada arsitektur RNN standar dan LSTM.

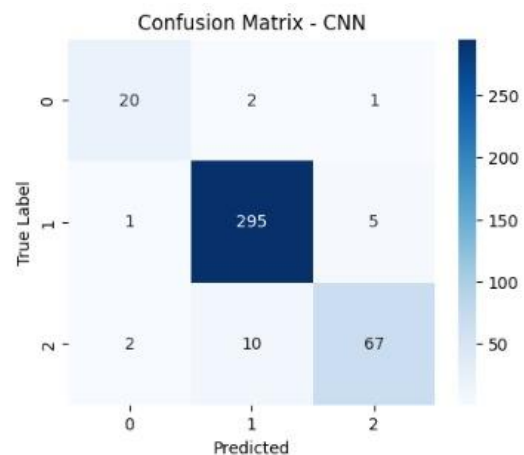
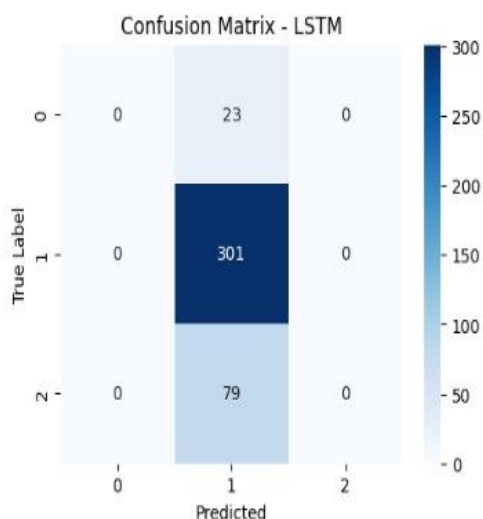
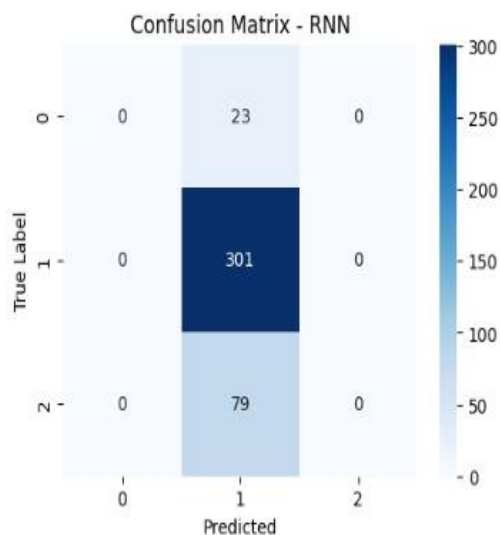
Dalam kondisi dataset pengujian yang tidak seimbang (*imbalanced data*), arsitektur berbasis urutan sekuensial temporal seperti RNN dan LSTM mengandalkan ketergantungan konteks kata dari waktu ke waktu. Ketika dihadapkan pada teks pendek dan timpang, fungsi penyesuaian bobot (*backpropagation through time*) mengalami gejala *vanishing gradient* pada kelas minoritas (0 dan 2). Akibatnya, demi meminimalkan nilai kerugian (*loss*) secara cepat selama 10 epoch pelatihan, jalan pintas optimasi terbaik yang diambil model adalah melakukan generalisasi berlebih (*over-generalization*) dengan selalu memprediksi seluruh sampel ke kelas mayoritas lokal data uji, yaitu kelas positif. Oleh karena itu, tingkat akurasi 75% pada RNN dan LSTM merupakan akurasi semu (*pseudo-accuracy*) karena model sama sekali tidak memiliki kemampuan diskriminasi dan klasifikasi riil antar-kelas.

Sebaliknya, model CNN menunjukkan performa yang sangat impresif, stabil, dan seimbang di seluruh kelas dengan akurasi keseluruhan mencapai 94,79% (tercatat 0.95 pada pembulatan sistem). Pada kelas positif, CNN mencapai *F1-Score* sebesar 0.97, sementara pada kelas negatif dan netral tetap mampu menghasilkan *F1-Score* yang tinggi masing-masing sebesar 0.87 dan 0.88. Hal ini membuktikan bahwa arsitektur CNN memiliki ketahanan (*resilience*) yang jauh lebih baik terhadap kendala ketidakseimbangan sebaran data. Melalui mekanisme filter spasial pada lapisan konvolusi 1D, CNN mampu mengekstraksi fitur lokal berupa kombinasi kata kunci (*n-grams*) spesifik yang mengandung bobot sentimen kuat tanpa bergantung pada

panjang pendeknya urutan ataupun dominasi jumlah kalimat. Keunggulan ekstraksi fitur lokal inilah yang membuat CNN tetap adaptif, presisi, dan terhindar dari bias kelas mayoritas meskipun bekerja pada subset dataset media sosial yang tidak seimbang.

3.6 Analisis Kesalahan (Confusion Matrix)

Untuk mendalami sejauh mana ketepatan prediksi masing-masing model terhadap setiap kelas sentimen, dilakukan analisis menggunakan *Confusion Matrix*. Melalui matriks ini, dapat diketahui distribusi data yang diprediksi dengan benar (terletak pada garis diagonal) serta data yang mengalami salah klasifikasi. Visualisasi *Confusion Matrix* untuk model RNN, LSTM, dan CNN disajikan pada Gambar 4.



Gambar 4. Confusion Matrix

Berdasarkan Gambar 4, terlihat fenomena yang sangat jelas pada model RNN dan LSTM. Kedua model tersebut mengalami kegagalan total dalam mengklasifikasikan kelas 0 (Negatif) dan kelas 2 (Netral). Seluruh data pada kelas Negatif (23 sampel) dan Netral (79 sampel) justru diprediksi sebagai kelas 1 (Positif). Hal ini mengonfirmasi adanya masalah *majority class bias*, di mana model cenderung 'menebak aman' dengan memilih kelas mayoritas (Positif) untuk semua data guna mendapatkan akurasi yang tampak tinggi secara semu. Berbeda secara signifikan, model CNN menunjukkan persebaran prediksi yang sangat akurat pada diagonal utama matriks.

- Pada kelas negatif (0), CNN memprediksi dengan benar sebanyak 20 dari 23 data.
- Pada kelas positif (1), CNN memprediksi dengan tepat 295 dari 301 data.
- Pada kelas netral (2), CNN berhasil mengidentifikasi 67 dari 79 data.

Terdapat sedikit kesalahan klasifikasi pada model CNN, di mana 10 data netral terdeteksi sebagai positif. Hal ini cukup wajar dalam analisis sentimen komentar YouTube, karena seringkali terdapat kalimat pertanyaan atau pernyataan netral yang menggunakan kosakata yang mirip dengan komentar positif. Secara keseluruhan, *Confusion Matrix* membuktikan bahwa arsitektur CNN

memiliki kemampuan diskriminasi kelas yang jauh lebih superior dibandingkan RNN dan LSTM dalam menangani dataset sentimen produk Eiger ini.

3.7 Pembahasan Model Terbaik

Berdasarkan seluruh rangkaian pengujian yang telah dilakukan, mulai dari analisis metrik evaluasi hingga visualisasi matriks kebingungan, dapat disimpulkan bahwa *Convolutional Neural Network* (CNN) merupakan arsitektur terbaik untuk melakukan klasifikasi sentimen pada komentar YouTube produk Eiger dalam penelitian ini. Model CNN konsisten mengungguli arsitektur RNN dan LSTM dengan perbedaan performa yang sangat signifikan. Keunggulan CNN dalam penelitian ini didorong oleh beberapa faktor teknis :

1. Ekstraksi fitur lokal

Arsitektur CNN melalui lapisan Conv1D sangat efektif dalam mendeteksi pola kata kunci (*n-grams*) yang menentukan sentimen. Hal ini sangat cocok untuk data komentar YouTube yang cenderung pendek dan langsung pada inti opini.

2. Resiliensi terhadap ketidakseimbangan data

Meskipun dataset didominasi oleh sentimen positif, CNN terbukti tetap mampu mengenali kelas minoritas (negatif dan netral) dengan presisi yang tinggi, sementara RNN dan LSTM terjebak pada bias kelas mayoritas.

3. Efisiensi pelatihan

CNN mencapai nilai loss terendah (0,2628), yang menunjukkan bahwa model ini paling cepat beradaptasi dengan karakteristik bahasa informal dan istilah-istilah produk outdoor yang ada pada dataset.

Sebaliknya, kegagalan RNN dan LSTM dalam mengenali sentimen negatif dan netral menunjukkan bahwa pendekatan berbasis urutan (*sequential*) standar kurang efektif untuk dataset yang memiliki ketimpangan jumlah label yang besar tanpa dilakukan teknik penyeimbangan data tambahan. Dengan demikian, penggunaan

CNN direkomendasikan sebagai model otomatisasi analisis sentimen untuk membantu brand seperti Eiger dalam memantau persepsi konsumen secara *real-time* melalui media sosial.

SIMPULAN

Berdasarkan hasil analisis dan eksperimen yang telah dilakukan terhadap klasifikasi sentimen komentar produk Eiger di YouTube, dapat disimpulkan bahwa penggunaan metode *Deep Learning* memberikan hasil yang sangat signifikan dalam memetakan persepsi konsumen. Penelitian ini membuktikan bahwa arsitektur CNN (*Convolutional Neural Network*) merupakan model yang paling unggul dengan tingkat akurasi mencapai 94,79%, jauh melampaui arsitektur RNN dan LSTM yang tertahan di angka 74,69% (atau 75%). Keunggulan CNN terletak pada kemampuannya dalam mengekstraksi fitur kata kunci secara efektif serta ketangguhannya dalam mengklasifikasikan sentimen negatif dan netral meskipun jumlah datanya tidak seimbang. Dengan demikian, model CNN yang dikembangkan dalam penelitian ini sangat layak diimplementasikan sebagai alat analisis sentimen otomatis untuk memantau reputasi merek di media sosial secara akurat dan efisien.

DAFTAR PUSTAKA

- Ahuja, H., Kaur, N., Kumar, P., et al., "Machine Learning based Sentiment Analysis of YouTube Video Comments," 2023 First International ..., 2023.
- Surana, P., Vispute, S., Kalbhor, M., Vairagar, S., et al., "Opinion Mining of YouTube Video Comments Using Machine Learning and Deep Learning," Conference on ICT for ..., Springer, 2025.
- Liyih, A., Anagaw, S., Yibeyin, M., & Tehone, Y., "Sentiment analysis of the Hamas-Israel war on YouTube comments using deep learning," Scientific Reports, 2024.

- Ashtagi, R., Bidwe, R. V., Fukate, A., et al., "Sentiment analysis on YouTube comments using long short-term memory (LSTM) networks," 2025 6th ..., 2025.
- Singha, R., Das, S., Das, K., et al., "Sentiment Analysis of YouTube Comments Using Machine Learning Algorithm," 2024 IEEE International ..., 2024.
- Bindhumol, M., Singh, T., & Patra, P., "Sentiment Analysis using YouTube Comments," 2024 15th International ..., 2024.
- Visariya, D., Prajapati, P., & Bhatt, N., "YouTube comment sentiment analysis using NLP approach," AIP Conference Proceedings, 2025.
- Ercikan, Y. S., & Elmasry, W., "YouTube Video Comments Sentiment Analysis Using Custom NLP Model," 2025 7th International Congress on ..., 2025.
- Furqoni, W., Sindura, I. P., Rahayuningtyas, I. P., et al., "Analisis Sentimen Ulasan Film Moana 2 di Website Letterboxd Menggunakan Algoritma CNN dan RNN," *Journal of Informatics* ..., 2025.
- Utami, H., "Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network," *Indonesian Journal of Applied Statistics*, 2022.
- Irfani, M., & Khomsah, S., "Analisis Sentimen Berbasis Aspek pada EDOM Pembelajaran Menggunakan Metode CNN dan Word2vec," *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*.
- Mudding, A. A., & Karim, A. A. A., "Analisis Sentimen Menggunakan Algoritma LSTM Pada Media Sosial," *Jurnal* ..., 2022.
- Berliana, A. S., & Mustikasari, M., "Analisis Sentimen pada Ulasan Aplikasi Jakartanotebook di Google Play Menggunakan Metode Recurrent Neural Network (RNN)," *Jurnal Informatika dan Teknik* ..., 2024.
- Prasetyo, P. A. P., "Analisis Sentimen Pada Twitter terhadap Isu Penundaan Pemilu 2024 Menggunakan Metode Long Short-Term Memory (LSTM)," 2024.
- Aprilyani, S. F., & Khorunnisa, N. L., "Sentimen Analisis Komentar Aplikasi Enterkomputer di Google Play Menggunakan Metode Recurrent Neural Network (RNN)," *Jurnal Jawara* ..., 2024.