SENTRI:
Jurnal Riset Ilmiah
ejournal.nusantaraglobal.ac.id/index.php/sentri

Vol. 5, No. 10, 2026
e-ISSN: 2963-1130
pp. 5323-5339

Klasifikasi Intensi Komunikatif Pada Gambar Reaksi Dalam Komentar Konten Komedi Tiktok Menggunakan Cnn Berbasis Transfer Learning Efficientnetb0

Kristi V. Mokodompit[1]*, Christian A. J. Soewoeh[2] , Tohap Manurung[3]
[123]Program Studi Sistem Informasi, Jurusan Matematika, Fakultas Matematika Dan Ilmu Pengetahuan Alam, Universitas Sam Ratulangi Manado

*Corresponding Author's e-mail: mktkristi12@gmail.com

Article History:
Received :August 19, 2026
Revised: September 22, 2026
Accepted: October 29, 2026

Keywords:
reaction images, communicative intentions, Literal Intention, Figurative Intention, CNN, EfficientNetB0, TikTok

Abstract: *Reaction images are a form of visual communication widely used in social media comments, including TikTok. They can convey direct responses or indirect communicative intentions, such as sarcasm, irony, evaluation, symbolism, and implicit humor. This study aimed to develop and validate a classification of communicative intentions into Literal and Figurative Intentions and implement a Convolutional Neural Network (CNN) using EfficientNetB0 transfer learning. A total of 339 images were annotated by three annotators using an expert-judged guideline. After revising the guideline and relabeling 124 disagreements, the average descriptive Cohen's Kappa increased from 0.389 to 0.886. Following duplicate removal, 334 images remained, comprising 245 Literal and 89 Figurative Intentions. The data were stratified into training, validation, and test sets. The baseline achieved 76.47% accuracy, 67.31% macro F1-score, 75.45% weighted F1-score, and 0.6933 AUC. Figurative Intentions remained more difficult to classify, indicating limitations of unimodal visual approaches for context-dependent communication.*

Copyright © 2026, The Author(s).
This is an open access article under the CC–BY-SA license

How to cite: Mokodompit, K. V., Christian A. J. Soewoeh, & Tohap Manurung. (2026). KLASIFIKASI INTENSI KOMUNIKATIF PADA GAMBAR REAKSI DALAM KOMENTAR KONTEN KOMEDI TIKTOK MENGGUNAKAN CNN BERBASIS TRANSFER LEARNING EFFICIENTNETB0. *SENTRI: Jurnal Riset Ilmiah*, *5*(10). https://doi.org/10.55681/sentri.v5i10.7257

PENDAHULUAN

Perkembangan media sosial telah memperluas bentuk komunikasi digital yang tidak lagi bergantung pada teks, tetapi juga memanfaatkan berbagai bentuk visual, seperti gambar, *emoticon*, meme, dan gambar reaksi (*reaction image*). Unsur visual tersebut memungkinkan pengguna menyampaikan respons terhadap suatu konten secara lebih ringkas dan ekspresif. Dalam perkembangan analisis komputasional, informasi visual mulai dimanfaatkan untuk memahami berbagai karakteristik komunikasi digital, salah satunya melalui *visual sentiment analysis* (VSA) yang menganalisis informasi visual untuk mengidentifikasi sentimen atau informasi afektif. Namun, komunikasi visual tidak selalu dapat direduksi menjadi sentimen positif atau negatif. Suatu gambar dapat digunakan untuk menyampaikan fungsi komunikasi tertentu yang maknanya bergantung pada cara dan konteks penggunaannya.

Salah satu bentuk komunikasi visual yang banyak digunakan pada media sosial adalah gambar reaksi. Pada platform TikTok, gambar reaksi dapat digunakan untuk memberikan respons terhadap konten video dan menyampaikan intensi komunikatif yang

beragam. Dalam penelitian ini, penggunaan gambar yang menunjukkan respons secara langsung atau sesuai dengan makna visual yang tampak dikategorikan sebagai **Intensi Literal**, sedangkan penggunaan gambar untuk menyampaikan respons secara tidak langsung melalui sindiran, ironi, sarkasme, evaluasi, simbolik, atau humor implisit dikategorikan sebagai **Intensi Kiasan**. Dengan demikian, tampilan visual suatu gambar tidak selalu secara langsung merepresentasikan fungsi komunikatif yang dimaksudkan oleh pengguna. Kondisi tersebut menunjukkan bahwa analisis gambar reaksi perlu mempertimbangkan intensi komunikatif, bukan hanya karakteristik visual atau sentimen yang terkandung di dalamnya.

Identifikasi intensi komunikatif pada gambar reaksi menjadi semakin kompleks karena interpretasinya dapat dipengaruhi oleh konteks konten yang dikomentari. Gambar yang sama berpotensi memiliki fungsi komunikasi yang berbeda apabila digunakan pada video dengan konteks yang berbeda. Permasalahan ini terutama relevan dalam komunikasi humor karena penyampaian makna dapat melibatkan inkongruitas, ambiguitas, konteks, serta isyarat visual yang menghasilkan fungsi komunikasi secara tidak langsung [1]. Oleh karena itu, penelitian ini menggunakan konteks video pada tahap anotasi sebagai dasar bagi anotator dalam menentukan label, tetapi tidak memasukkan konteks tersebut sebagai masukan model. Pendekatan ini memungkinkan penelitian menguji secara khusus sejauh mana informasi visual pada gambar reaksi dapat digunakan untuk membedakan Intensi Literal dan Intensi Kiasan melalui pendekatan *unimodal visual*.

Pemanfaatan pembelajaran mesin berbasis visual telah menunjukkan potensi dalam menganalisis komunikasi digital. Penelitian Boukes *et al.* menunjukkan bahwa *visual machine learning* dapat digunakan untuk mengidentifikasi dan menganalisis pola pada objek visual digital seperti meme [2]. Di sisi lain, penelitian terkait TikTok di Indonesia masih banyak berfokus pada analisis berbasis teks. Salah satunya adalah penelitian Setiawan *et al.* yang melakukan klasifikasi sentimen terhadap ulasan aplikasi TikTok [3]. Meskipun penelitian tersebut menunjukkan pemanfaatan metode komputasional untuk menganalisis komunikasi atau respons terkait TikTok, pendekatan berbasis teks belum secara langsung menjelaskan fungsi komunikatif yang disampaikan melalui gambar reaksi. Berdasarkan literatur yang ditinjau, penelitian yang secara khusus mengklasifikasikan intensi komunikatif gambar reaksi pada komentar TikTok ke dalam Intensi Literal dan Intensi Kiasan masih terbatas. Keterbatasan tersebut menunjukkan adanya ruang penelitian untuk mengkaji fungsi komunikatif gambar reaksi melalui pendekatan berbasis informasi visual.

Untuk menjawab permasalahan tersebut, penelitian ini menerapkan *Convolutional Neural Network* (CNN) dengan pendekatan *transfer learning* menggunakan arsitektur EfficientNetB0. CNN digunakan untuk mempelajari representasi fitur visual dari gambar, sedangkan *transfer learning* memungkinkan model memanfaatkan representasi yang telah dipelajari melalui bobot pralatih untuk mendukung proses klasifikasi citra. Penelitian difokuskan pada gambar reaksi yang berasal dari komentar konten komedi TikTok berbahasa Indonesia, dengan subgenre Darkjokes, *Relateable*, dan Slapstick. Pemilihan konteks komedi didasarkan pada karakteristik komunikasi humor yang dapat melibatkan respons tidak langsung serta hubungan yang kuat antara elemen visual dan konteks komunikasi [1].

Selain membangun model klasifikasi, penelitian ini juga memberikan perhatian pada validitas skema anotasi karena penentuan Intensi Literal dan Intensi Kiasan memiliki unsur interpretatif. Data dianotasi oleh tiga anotator menggunakan pedoman yang terlebih dahulu melalui *expert judgement*. Tingkat kesepakatan anotator kemudian digunakan untuk

mengevaluasi konsistensi penerapan skema klasifikasi. Tahapan tersebut penting karena model klasifikasi hanya dapat menghasilkan prediksi yang bermakna apabila label data yang digunakan sebagai dasar pembelajaran memiliki definisi dan konsistensi yang memadai.

Berdasarkan permasalahan dan kesenjangan penelitian tersebut, penelitian ini bertujuan untuk merancang dan memvalidasi skema klasifikasi intensi komunikatif gambar reaksi ke dalam Intensi Literal dan Intensi Kiasan serta mengimplementasikan CNN berbasis *transfer learning* dengan EfficientNetB0 untuk mengklasifikasikan kedua kategori tersebut. Penelitian ini diharapkan dapat memberikan kontribusi pada kajian komunikasi visual berbasis komputasi melalui penyediaan dataset beranotasi dan skema pelabelan intensi komunikatif dalam konteks komunikasi digital Indonesia. Selain itu, hasil evaluasi model dapat memberikan gambaran mengenai sejauh mana informasi visual secara *unimodal* mampu merepresentasikan intensi komunikatif, khususnya ketika makna gambar reaksi bergantung pada konteks penggunaan.

LANDASAN TEORI

Komunikasi Visual dan *Reaction Image*

Komunikasi digital berkembang dari penggunaan teks menuju pemanfaatan berbagai unsur visual, seperti *emoticon*, meme, dan *reaction image*. Unsur visual tersebut dapat membantu menyampaikan emosi, sikap, dan intensi secara lebih ekspresif [1]. *Reaction image* merupakan bentuk komunikasi visual yang digunakan untuk memberikan respons terhadap suatu konten dan dapat memuat ekspresi wajah, karakter, objek, teks, maupun kombinasi berbagai elemen visual. Maknanya tidak selalu ditentukan oleh tampilan visual, tetapi juga oleh hubungan gambar dengan konteks komunikasi yang direspons [2].

Intensi Komunikatif

Intensi komunikatif berkaitan dengan fungsi atau maksud yang ingin disampaikan melalui suatu pesan. Konsep ini dapat dikaitkan dengan *Speech Act Theory* yang memandang komunikasi sebagai tindakan untuk mencapai maksud tertentu [3]. Dalam penelitian ini, intensi komunikatif diklasifikasikan menjadi Intensi Literal dan Intensi Kiasan. Intensi Literal merupakan penggunaan *reaction image* yang maknanya dapat dipahami secara langsung dan selaras dengan konteks, sedangkan Intensi Kiasan merupakan penggunaan gambar yang mengandung makna tambahan atau tidak langsung, seperti sindiran, ironi, sarkasme, evaluasi, simbolik, dan humor implisit.

Humor dan Makna Kiasan

Pemaknaan *reaction image* dalam konteks humor dapat melibatkan inkongruitas, ambiguitas, konteks, dan isyarat visual [4]. Teori Inkongruitas menjelaskan humor melalui ketidakselarasan antara harapan dan situasi yang ditampilkan. Sementara itu, Teori Relevansi menjelaskan bahwa pemahaman terhadap maksud komunikatif dapat memerlukan proses inferensi berdasarkan konteks [5]. Dalam penelitian ini, kedua konsep tersebut menjadi dasar untuk memahami perbedaan antara respons literal dan respons kiasan. Konteks video digunakan oleh anotator dalam menentukan label, tetapi tidak digunakan sebagai input model sehingga klasifikasi tetap bersifat *unimodal visual* [6].

CNN, *Transfer Learning*, dan EfficientNetB0

Convolutional Neural Network (CNN) merupakan arsitektur *deep learning* yang mampu mempelajari representasi fitur visual secara hierarkis melalui operasi konvolusi [7]. Penelitian ini menggunakan *transfer learning*, yaitu pemanfaatan representasi dan

parameter dari model yang telah dilatih sebelumnya untuk tugas klasifikasi pada dataset baru. Pendekatan tersebut relevan untuk dataset berukuran terbatas karena model tidak perlu mempelajari seluruh representasi visual dari awal [8].

Arsitektur yang digunakan adalah EfficientNetB0, bagian dari keluarga EfficientNet yang dikembangkan menggunakan pendekatan *compound scaling* pada *depth*, *width*, dan resolusi input [9]. Dalam penelitian ini, EfficientNetB0 menggunakan bobot pralatih ImageNet sebagai *feature extractor*, kemudian lapisan klasifikasi disesuaikan untuk membedakan Intensi Literal dan Intensi Kiasan.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimental untuk mengklasifikasikan intensi komunikatif pada *reaction image* menggunakan CNN EfficientNetB0 berbasis *transfer learning*. Data berupa *reaction image* dari kolom komentar TikTok pada konten komedi kreator Indonesia yang dikumpulkan secara manual pada November 2025–Mei 2026 menggunakan *purposive sampling*. Data mencakup subgenre *Darkjokes*, *Relateable*, dan *Slapstick* serta diseleksi berdasarkan kesesuaian dengan karakteristik *reaction image*. Dari 339 gambar yang diperoleh, lima *exact duplicate* dihapus menggunakan *content hash*, sehingga diperoleh 334 data yang terdiri atas 245 Intensi Literal dan 89 Intensi Kiasan.

Pelabelan dilakukan secara independen oleh tiga anotator menggunakan pedoman yang disusun berdasarkan Teori Tindak Tutur, Teori Relevansi, teori inkongruitas, serta kajian humor dan informasi visual [3]–[6]. Pedoman divalidasi melalui *expert judgement*. Kesepakatan antar-anotator diukur menggunakan *Cohen's Kappa* secara berpasangan [10]. Dari 339 data awal, terdapat 124 *disagreement* yang kemudian dilabel ulang setelah revisi pedoman; 316 data mencapai kesepakatan dan 23 data diselesaikan menggunakan *majority vote*. Intensi Literal didefinisikan sebagai respons yang dapat dipahami secara langsung, sedangkan Intensi Kiasan mengandung makna tambahan atau tidak langsung berdasarkan konteks.

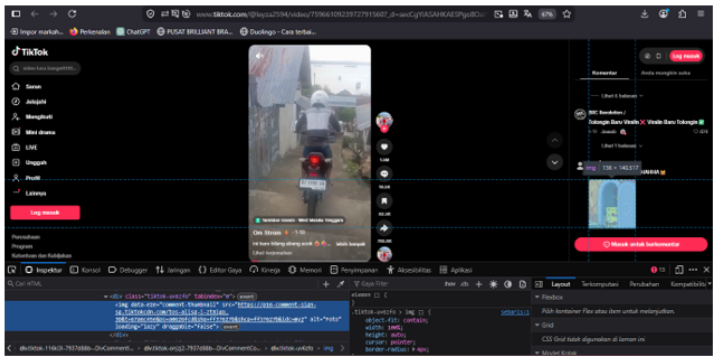
Model menggunakan EfficientNetB0 dengan bobot pralatih ImageNet sebagai *feature extractor*, seluruh *base model* dibekukan, dan lapisan klasifikasi disesuaikan menjadi *Global Average Pooling*, *Dropout* 0,3, serta *Dense* sigmoid untuk klasifikasi biner [9], [12]. Dataset dibagi secara *stratified split* menjadi 267 data *training*, 33 validasi, dan 34 *test*. Pelatihan dilakukan maksimal 40 *epoch* dengan *EarlyStopping* dan *ModelCheckpoint* (*patience* 8). Dua skenario diuji, yaitu *baseline* dan *class weighting*, kemudian dievaluasi menggunakan *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, ROC, dan AUC.Contoh Diagram:

HASIL DAN PEMBAHASAN

Hasil Penyusunan Dataset

Pengumpulan Data

Data penelitian berupa *reaction image* yang diperoleh dari kolom komentar konten komedi kreator Indonesia di TikTok. Pengumpulan dilakukan secara manual dengan menelusuri kolom komentar, kemudian sumber gambar diakses melalui fitur *Developer Tools*/*Inspect* pada peramban. Gambar yang sesuai dengan kriteria penelitian disimpan sebagai dataset, sedangkan gambar dengan teks dominan atau tidak sesuai karakteristik *reaction image* dikeluarkan. Proses pengambilan data melalui *Developer Tools* ditunjukkan pada Gambar 1.



Gambar 1 Proses Pengambilan Gambar Reaksi melalui Fitur Developer Tools

Beberapa gambar yang mengalami kendala akses atau sumber tidak tersedia dilakukan pengambilan ulang. Gambar yang sama tetap dipertahankan jika ditemukan pada video berbeda, sedangkan variasi ukuran, *crop*, dan *zoom* dipertahankan selama sesuai karakteristik data. Dari proses tersebut diperoleh 339 gambar reaksi yang selanjutnya diaudit dan diseleksi.

Audit dan Seleksi Data Akhir

Audit dilakukan terhadap 339 gambar reaksi untuk memastikan keterpakaian data dan mengidentifikasi duplikasi identik menggunakan *content hash*. Hasil pemeriksaan menemukan 5 *exact duplicate* yang dikeluarkan dari dataset pemodelan dengan mempertahankan satu data dari setiap kelompok duplikasi (*keep="first"*). Proses *cleaning* dilakukan setelah anotasi dan evaluasi reliabilitas, sehingga 339 data tetap menjadi basis pelabelan dan evaluasi kesepakatan antar-anotator. Gambar yang hanya memiliki kemiripan visual, tetapi berasal dari konteks berbeda, serta variasi ukuran, rasio, *cropping*, dan *zoom* tetap dipertahankan. Dengan demikian, diperoleh 334 gambar sebagai dataset pemodelan final, terdiri atas 245 Intensi Literal dan 89 Intensi Kiasan.

Tabel 1 Ringkasan Hasil Audit dan Seleksi Data

No	Keterangan	Jumlah
1	Data basis anotasi dan reliabilitas	339
2	*Exact duplicate*	5
3	Dataset pemodelan final	334

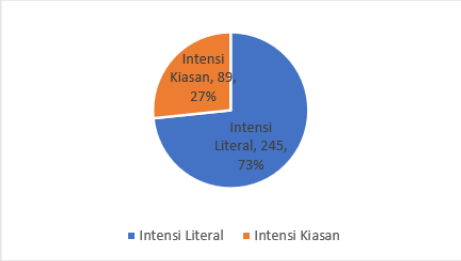
Distribusi Dataset Berdasarkan Label

Dataset pemodelan final terdiri atas 334 gambar setelah melalui proses anotasi, revisi panduan pelabelan, *relabeling* terhadap data *disagreement*, penentuan label menggunakan *majority vote*, dan penghapusan 5 *exact duplicate*. Sebanyak 245 data (73,35%) termasuk Intensi Literal, sedangkan 89 data (26,65%) termasuk Intensi Kiasan. Distribusi tersebut menunjukkan bahwa data Intensi Literal memiliki proporsi lebih besar dalam dataset pemodelan.

Tabel 2 Distribusi Dataset Berdasarkan Label

No	Label	Jumlah Data	Persentase (%)
1	Intensi Kiasan	89	26,65
2	Intensi Literal	245	73,35
Total		**334**	**100,00**

Setelah proses anotasi, evaluasi reliabilitas, revisi panduan, *relabeling*, dan *majority vote*, diperoleh 339 data sebagai basis anotasi. Pemeriksaan *content hash* menemukan 5 *exact duplicate* yang dikeluarkan untuk pemodelan, sehingga tersisa 334 data. Dataset tersebut terdiri atas 245 Intensi Literal (73,35%) dan 89 Intensi Kiasan (26,65%), dengan distribusi label akhir disajikan pada Tabel 2.



Gambar 2 Distribusi Label Berdasarkan Dataset

Pembagian Dataset (Training, Validation, dan Testing)

Setelah anotasi, penentuan label akhir, dan *cleaning*, 334 data dibagi menjadi *training set*, *validation set*, dan *test set* menggunakan *train_test_split* dengan rasio sekitar 80%:10%:10%, *random_state*=42, dan *stratify* berdasarkan label. Hasil pembagian terdiri atas 267 data *training*, 33 data *validation*, dan 34 data *test*. *Training set* digunakan untuk melatih model, *validation set* untuk memantau proses pelatihan, dan *test set* untuk evaluasi akhir. *Data augmentation* hanya diterapkan secara *on-the-fly* pada *training set*, sedangkan *validation set* dan *test set* menggunakan data asli.

Tabel 3 Pembagian Dataset Penelitian

No	Dataset	Intensi Literal	Intensi Kiasan	Jumlah Data	Persentase (%)
1	*Training Set*	196	71	267	79,94
2	*Validation Set*	24	9	33	10,03
3	*Test Set*	25	9	34	10,03
Total		**245**	**89**	**334**	**100,00**

Hasil Penyusunan dan Validasi Panduan Pelabelan

Panduan pelabelan disusun untuk menentukan Intensi Literal dan Intensi Kiasan dengan mempertimbangkan fungsi komunikatif gambar dan konteks video. Panduan divalidasi melalui *expert judgement* sebelum digunakan oleh tiga anotator. Setelah anotasi awal, reliabilitas diukur menggunakan *Cohen's Kappa*, kemudian panduan diperjelas berdasarkan data *disagreement*, dilakukan *relabeling*, dan *majority vote* digunakan untuk menentukan label akhir pada data yang masih berbeda.

Penyusunan dan Validasi Panduan Pelabelan

Panduan menetapkan dua kategori, yaitu Intensi Literal dan Intensi Kiasan, dengan penentuan label berdasarkan fungsi komunikasi gambar dalam konteks video. Intensi Literal menunjukkan respons langsung, sedangkan Intensi Kiasan menunjukkan makna tambahan atau tidak langsung, seperti sindiran, ironi, sarkasme, penilaian, simbolisasi, hiperbola, dan humor implisit. Hasil *expert judgement* terhadap delapan aspek memperoleh skor 31/32 (96,88%) dan menyatakan panduan layak digunakan dengan revisi minor.

Tabel 5 Hasil Validasi *Annotation Guideline* oleh Ahli Bahasa

No	Aspek yang Dinilai	Skor
1	Kejelasan definisi Intensi Literal	4
2	Kejelasan definisi Intensi Kiasan	4
3	Konsistensi penggunaan istilah	3
4	Kejelasan bentuk-bentuk Intensi	4
5	Kejelasan *decision rule*	4
6	Kesesuaian contoh dengan label	4
7	Kemudahan *annotation guideline* dipahami	4
8	Kelayakan *annotation guideline* digunakan	4

Total		**31/32**

Setelah anotasi awal ditemukan 124 data *disagreement*, panduan kemudian diperjelas dengan menekankan fungsi komunikatif, konteks video, hierarki penentuan label, dan penanganan kasus ambigu. Panduan hasil penyempurnaan digunakan sebagai acuan *relabeling* terhadap 124 data tersebut.

Hasil Pelabelan dan Reliabilitas Annotator

Anotasi awal dilakukan secara independen oleh tiga anotator terhadap 339 gambar dengan mempertimbangkan hubungan gambar, konteks video, dan fungsi komunikatif. Reliabilitas diukur menggunakan *Cohen's Kappa* secara berpasangan pada A1–A2, A1–A3, dan A2–A3. Hasil awal masing-masing sebesar 0,340; 0,456; dan 0,371, dengan rata-rata 0,389. Sebanyak 215 data memiliki label sama dari ketiga anotator, sedangkan 124 data mengalami *disagreement*.

Tabel 6 Hasil *Cohen's Kappa* pada Pelabelan Awal

No	Pasangan Annotator	Nilai Kappa
1	A1–A2	0,340
2	A1–A3	0,456
3	A2–A3	0,371
Rata-Rata		**0,389**

Penyelesaian *Disagreement* dan Revisi Panduan

Sebanyak 124 data *disagreement* dilabeli kembali secara independen setelah panduan diperjelas. Hasilnya, 316 data memperoleh label yang sama dari ketiga anotator dan 23 data masih berbeda. Nilai *Cohen's Kappa* meningkat menjadi 0,901 pada A1–A2, 0,875 pada A1–A3, dan 0,880 pada A2–A3, dengan rata-rata 0,886.

Tabel 7 Perbandingan *Cohen's Kappa* Sebelum dan Setelah *Relabeling*

No	Pasangan Annotator	Sebelum *Relabeling*	Sesudah *Relabeling*
1	A1–A2	0,340	0,901
2	A1–A3	0,456	0,875
3	A2–A3	0,371	0,880
Rata-Rata		**0,389**	**0,886**

Sebanyak 23 data yang masih berbeda diselesaikan dengan *majority vote*, yaitu label yang memperoleh minimal dua dari tiga suara ditetapkan sebagai label akhir. Setelah proses tersebut, seluruh 339 data memperoleh label final. Selanjutnya, *cleaning* menggunakan *content hash* mengeluarkan 5 *exact duplicate*, sehingga 334 data digunakan sebagai dataset pemodelan final.



Gambar 3 Contoh Penerapan Decision Rule pada Data Penelitian

Berdasarkan Gambar 3, RID034 dan RID035 berasal dari konteks video yang sama tetapi memperoleh label berbeda karena fungsi komunikatif gambar reaksinya berbeda. Pola serupa terlihat pada RID218 dan RID223, di mana RID218 menyampaikan respons langsung, sedangkan RID223 menyampaikan makna tambahan atau tidak langsung. Hal ini menunjukkan bahwa penentuan label didasarkan pada hubungan antara konteks video, gambar reaksi, dan fungsi komunikatif sesuai *decision rule*, bukan semata-mata pada jenis gambar atau konteks video.

Hasil Implementasi Model CNN

Model CNN diimplementasikan menggunakan pendekatan *transfer learning* dengan arsitektur EfficientNetB0 untuk mengklasifikasikan gambar reaksi ke dalam Intensi Literal dan Intensi Kiasan. Model *baseline* menggunakan bobot pralatih ImageNet dengan *base network* dibekukan (*frozen*) dan menjadi acuan utama evaluasi kinerja pada dataset pemodelan final.

Arsitektur Model EfficientNetB0

Model menggunakan EfficientNetB0 dengan bobot *pre-trained* ImageNet dan *include_top=False*. Citra masukan berukuran 224 × 224 piksel dengan tiga kanal RGB, sedangkan seluruh bobot *base model* dibekukan sebagai *feature extractor*. *Data augmentation* diterapkan secara *on-the-fly* hanya pada *training set*, meliputi *horizontal flip*, rotasi 0,10, *zoom* 0,10, dan perubahan kontras 0,10.

Keluaran EfficientNetB0 berukuran 7 × 7 × 1280 kemudian diproses melalui GlobalAveragePooling2D menjadi 1.280 fitur, dilanjutkan Dropout 0,3 dan satu neuron Dense dengan aktivasi *sigmoid* untuk klasifikasi biner. Model memiliki 4.050.852 parameter, terdiri atas 1.281 *trainable* dan 4.049.571 *non-trainable parameters*.

Tabel 8 Ringkasan Arsitektur Model EfficientNetB0

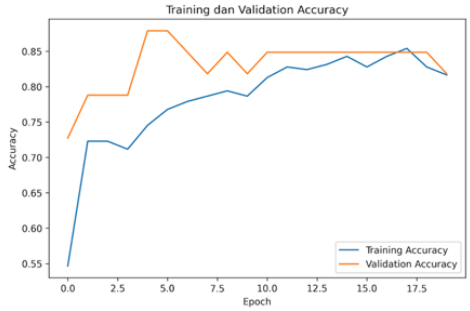
No	Lapisan	*Output Shape*	Parameter
1	Input Layer	(None, 224, 224, 3)	0
2	Data Augmentation	(None, 224, 224, 3)	0
3	EfficientNetB0	(None, 7, 7, 1280)	4.049.571
4	GlobalAveragePooling2D	(None, 1280)	0
5	Dropout (0,3)	(None, 1280)	0
6	Dense + Sigmoid	(None, 1)	1.281

Total			**4.050.852**

Konfigurasi Pelatihan Baseline

Pelatihan *baseline* menggunakan 267 data *training set* dan 33 data *validation set*, sedangkan 34 data *test set* hanya digunakan untuk evaluasi akhir. Model dilatih maksimal 40 *epoch* dengan pemantauan *validation loss* (*val_loss*). *ModelCheckpoint* digunakan untuk menyimpan model dengan *val_loss* terbaik, sedangkan *EarlyStopping* menghentikan pelatihan ketika tidak terjadi perbaikan sesuai *patience*. Dengan demikian, model evaluasi dipilih berdasarkan kinerja validasi terbaik dan *test set* tetap tidak terlibat dalam proses pelatihan maupun pemilihan model.

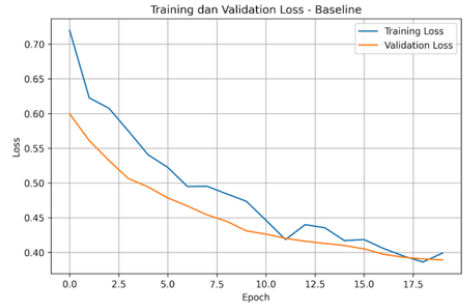
Hasil Pelatihan Baseline



Gambar 4 Grafik Training Accuracy dan Validation Accuracy

Berdasarkan Gambar 4, *training accuracy* meningkat dari sekitar 54,8% menjadi 82%, sedangkan *validation accuracy* meningkat dari sekitar 73% dan mencapai puncak sekitar 88% pada beberapa *epoch* awal, kemudian berfluktuasi pada kisaran 82%–88%. Hal ini menunjukkan peningkatan kemampuan model dalam mempelajari *training set* dengan kinerja *validation set* yang relatif terjaga.

Proses pelatihan berhenti sebelum mencapai maksimum 40 *epoch* sesuai mekanisme *EarlyStopping*. Meskipun *validation accuracy* sempat lebih tinggi daripada *training accuracy*, pemilihan *checkpoint* terbaik didasarkan pada *validation loss* (*val_loss*), bukan *accuracy*. Oleh karena itu, Gambar 8 digunakan untuk menunjukkan perkembangan pelatihan, sedangkan kinerja akhir model dievaluasi menggunakan *test set* terpisah.



Gambar 5 Grafik Training Loss dan Validation Loss

Gambar 5 menunjukkan *training loss* menurun dari sekitar 0,72 menjadi 0,39, sedangkan *validation loss* turun dari sekitar 0,60 menjadi 0,39. Meskipun terdapat fluktuasi pada beberapa *epoch*, kedua kurva secara umum menunjukkan penurunan kesalahan prediksi dengan nilai yang relatif berdekatan menjelang akhir pelatihan.

Validation loss menjadi dasar pemilihan model karena *ModelCheckpoint* menyimpan bobot dengan *val_loss* terbaik, sedangkan *EarlyStopping* menghentikan pelatihan ketika tidak terdapat perbaikan sesuai *patience*. *Checkpoint* terbaik kemudian digunakan untuk

mengevaluasi 34 data *test set* yang tidak terlibat dalam pelatihan maupun pemilihan model, sehingga hasilnya merepresentasikan kemampuan generalisasi model *baseline*.

Hasil Evaluasi Baseline Final

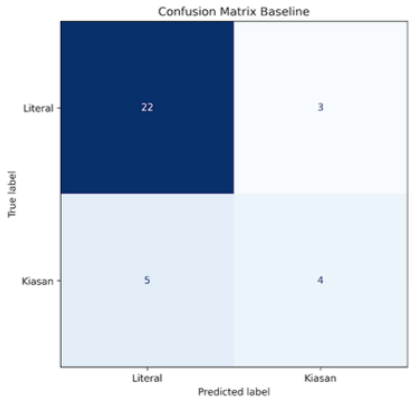
Classification Report

Tabel 9 Classification Report Model Baseline EfficientNetB0

No	Kelas	Precision	Recall	F1-Score	Support
1	Intensi Literal	0,8148	0,8800	0,8462	25
2	Intensi Kiasan	0,5714	0,4444	0,5000	9
3	Accuracy	–	–	0,7647	34
4	Macro Average	0,6931	0,6622	0,6731	34
5	Weighted Average	0,7504	0,7647	0,7545	34

Berdasarkan Tabel 9, model memperoleh *accuracy* sebesar 76,47%. Kinerja pada Intensi Literal lebih tinggi, dengan *precision* 0,8148, *recall* 0,8800, dan *F1-score* 0,8462, sedangkan Intensi Kiasan memperoleh *precision* 0,5714, *recall* 0,4444, dan *F1-score* 0,5000. *Macro average F1-score* sebesar 0,6731 menunjukkan kinerja rata-rata kedua kelas dengan bobot yang sama, sedangkan *weighted average F1-score* sebesar 0,7545 mempertimbangkan jumlah data tiap kelas.

Confusion Matrix



Gambar 6 Confusion Matrix Model Baseline EfficientNetB0

Berdasarkan *confusion matrix*, dari 25 data Intensi Literal, sebanyak 22 data diklasifikasikan benar sebagai Literal dan 3 data sebagai Kiasan. Dari 9 data Intensi Kiasan, 4 data diklasifikasikan benar sebagai Kiasan dan 5 data sebagai Literal. Secara keseluruhan, model menghasilkan 26 prediksi benar dan 8 kesalahan dari 34 data *test set*.

Tabel 10 Ringkasan Hasil Prediksi pada *Test Set*

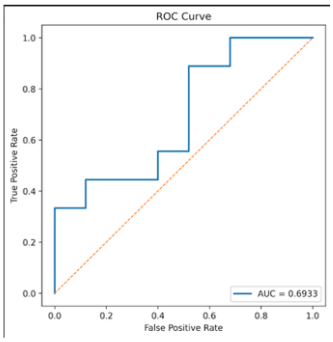
No	Label Aktual	Prediksi Literal	Prediksi Kiasan	Total
1	Literal	22	3	25
2	Kiasan	5	4	9
Total		**27**	**7**	**34**

Hasil tersebut menunjukkan bahwa prediksi benar lebih banyak pada Intensi Literal dibandingkan Intensi Kiasan, sejalan dengan *recall* masing-masing sebesar 88,00% dan 44,44%. Model mampu mengenali sebagian besar data Literal, tetapi masih terdapat data Kiasan yang diprediksi sebagai Literal.

ROC Curve dan Nilai AUC

ROC Curve digunakan untuk mengevaluasi kemampuan model membedakan dua kategori berdasarkan probabilitas prediksi pada berbagai *threshold*. Kurva menunjukkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR), sedangkan AUC menggambarkan kemampuan diskriminasi model.

Dalam pemodelan, Intensi Literal dikodekan sebagai kelas 0 dan Intensi Kiasan sebagai kelas 1. Dengan demikian, Intensi Kiasan digunakan sebagai kelas positif dalam perhitungan ROC.



Gambar 7 ROC Curve Model Baseline EfficientNetB0

Berdasarkan evaluasi final, model *baseline* memperoleh AUC sebesar 0,6933. Nilai tersebut berada di atas 0,5, sehingga menunjukkan kemampuan diskriminasi di atas klasifikasi acak pada *test set*. Secara keseluruhan, model memperoleh *accuracy* 76,47%, *macro F1-score* 0,6731, *weighted F1-score* 0,7545, dan AUC 0,6933. Pada tingkat kelas, Intensi Literal memperoleh *F1-score* 0,8462, sedangkan Intensi Kiasan 0,5000.

Eksperimen *Class Weighting*

Distribusi dataset menunjukkan ketidakseimbangan jumlah data antara Intensi Literal dan Intensi Kiasan. Oleh karena itu, dilakukan eksperimen *class weighting* untuk memberikan bobot lebih besar pada kelas minoritas dan menguji pengaruhnya terhadap kemampuan model mengenali Intensi Kiasan.

Penerapan *Class Weight*

Class weight dihitung berdasarkan distribusi *training set* yang terdiri atas 196 data Intensi Literal dan 71 data Intensi Kiasan. Hasil perhitungan menghasilkan bobot 0,6811 untuk Intensi Literal dan 1,8803 untuk Intensi Kiasan, sehingga kesalahan pada kelas Kiasan memperoleh bobot lebih besar selama pelatihan.

4.6.2 Hasil Eksperimen *Class Weighting*

Pada *test set* yang sama, model dengan *class weighting* memperoleh *accuracy* 70,59%, *macro precision*, *macro recall*, dan *macro F1-score* masing-masing 62,22%. Intensi Literal memperoleh *recall* dan *F1-score* 80,00%, sedangkan Intensi Kiasan memperoleh *recall* 44,44% dan *F1-score* 44,44%. AUC yang diperoleh sebesar 0,6844.

Tabel 11 Hasil Evaluasi Model dengan *Class Weighting*

No	Metrik	Hasil
1	Accuracy	70,59%
2	Macro Precision	62,22%
3	Macro Recall	62,22%
4	Macro F1	62,22%
5	F1 Literal	80,00%
6	Recall Kiasan	44,44%
7	F1 Kiasan	44,44%
8	AUC	0,6844

Perbandingan Baseline dan *Class Weighting*

Tabel 12 Perbandingan Hasil *Baseline* dan *Class Weighting*

No	Metrik	Baseline	Class Weighting
1	Accuracy	76,47%	70,59%
2	Macro Precision	69,31%	62,22%
3	Macro Recall	66,22%	62,22%
4	Macro F1	67,31%	62,22%
5	Recall Kiasan	44,44%	44,44%
6	F1 Kiasan	50,00%	44,44%
7	AUC	0,6933	0,6844
8	Accuracy	76,47%	70,59%

Berdasarkan Tabel 12, *class weighting* tidak meningkatkan kinerja pada *test set*. *Recall* Intensi Kiasan tetap 44,44%, sedangkan *F1-score* turun dari 50,00% menjadi 44,44%. *Accuracy*, *macro F1-score*, dan AUC juga mengalami penurunan. Dengan demikian, *class weighting* digunakan sebagai eksperimen pembanding, sedangkan *baseline* dipertahankan sebagai model utama.

Analisis Berdasarkan Subgenre

Analisis subgenre dilakukan secara deskriptif pada tiga kelompok konten, yaitu *Darkjokes*, *Relateable*, dan *Slapstick*. Analisis ini digunakan untuk memberikan gambaran kinerja model, dengan mempertimbangkan jumlah data pengujian pada setiap subgenre yang relatif terbatas.

Hasil Evaluasi Berdasarkan Subgenre

Tabel 13 Hasil Evaluasi Baseline Berdasarkan Subgenre

No	Subgenre	Jumlah Test	Accuracy	Recall Kiasan	F1 Kiasan
1	*Darkjokes*	17	76,47%	57,14%	66,67%
2	*Relateable*	7	71,43%	0%	0%
3	*Slapstick*	10	80,00%	0%	0%

Pada *Darkjokes*, model memperoleh *accuracy* 76,47%, *recall* Intensi Kiasan 57,14%, dan *F1-score* 66,67%. Pada *Relateable*, *accuracy* sebesar 71,43%, sedangkan *recall* dan *F1-score* Kiasan 0%. Sementara itu, *Slapstick* memperoleh *accuracy* 80,00%, dengan *recall* dan *F1-score* Kiasan juga 0%.

Hasil tersebut menunjukkan bahwa pada *test set* yang digunakan, model masih mampu mengenali sebagian data Kiasan pada *Darkjokes*, sedangkan sampel Kiasan pada *Relateable* dan *Slapstick* belum berhasil dikenali.

Pembahasan Hasil Berdasarkan Subgenre

Perbedaan hasil antar-subgenre dipengaruhi oleh jumlah sampel, terutama data Intensi Kiasan pada *Relateable* dan *Slapstick*. Oleh karena itu, analisis subgenre bersifat deskriptif dan digunakan untuk menggambarkan performa model pada data pengujian yang tersedia, bukan untuk generalisasi performa pada setiap subgenre. Hasil ini selanjutnya menjadi dasar tambahan dalam analisis kesalahan pada tingkat data individual.

Error Analysis

Error analysis dilakukan untuk memahami karakteristik kesalahan prediksi, dengan fokus pada Intensi Kiasan karena memiliki *recall* dan *F1-score* lebih rendah dibandingkan Intensi Literal.

Identifikasi Kesalahan Prediksi

Pada *test set*, terdapat 5 data Intensi Kiasan yang diprediksi sebagai Literal (*false negative*) dan 4 data yang berhasil diprediksi sebagai Kiasan (*true positive*). Untuk analisis kualitatif, dipilih seluruh 5 *false negative* dan 2 *true positive*, sehingga terdapat 7 kasus yang dianalisis untuk membandingkan karakteristik Kiasan yang berhasil dan tidak berhasil dikenali model.

Tabel 14 Ringkasan *Error Analysis* pada Intensi Kiasan

No	ID	Kondisi Prediksi	Karakteristik Kiasan	Ketergantungan Konteks
1	RID293	K → L	Evaluatif/implisit	Tinggi
2	RID265	K → K	Simbolik/asosiatif	Sedang
3	RID246	K → L	Hiperbola/berlebihan	Tinggi
4	RID226	K → L	Ironi/kontras	Tinggi
5	RID223	K → L	Sarkastik/evaluatif	Sedang
6	RID115	K → K	Simbolik/asosiatif	Sedang
7	RID025	K → L	Respons implisit/representatif	Sedang

Karakteristik Intensi Kiasan pada Hasil Prediksi

Error analysis menunjukkan bahwa Intensi Kiasan tidak ditandai oleh satu karakteristik visual tertentu, tetapi mencakup ironi/kontras, sarkasme/evaluasi, hiperbola, simbolik/asosiatif, dan respons implisit. Karakteristik tersebut hanya digunakan untuk interpretasi kesalahan dan bukan sebagai kelas tambahan.

1. RID293 menunjukkan respons evaluatif implisit yang bergantung pada konteks video, tetapi diprediksi sebagai Literal.
2. RID265 menggunakan pola simbolik/asosiatif dan berhasil diprediksi sebagai Kiasan, menunjukkan bahwa sebagian karakteristik visual dapat dikenali model.
3. RID246 menggunakan ekspresi hiperbolis, tetapi diprediksi sebagai Literal.
4. RID226 menggunakan hubungan kontras/ironi antara gambar dan konteks video, tetapi diprediksi sebagai Literal.
5. RID223 menggunakan tawa dan gestur sebagai respons evaluatif atau mengejek, tetapi diprediksi sebagai Literal.
6. RID115 menggunakan pola simbolik/asosiatif dan berhasil diprediksi sebagai Kiasan, sedangkan RID025 menggunakan respons implisit yang bergantung pada konteks video dan diprediksi sebagai Literal.
7. Karakteristik tersebut tetap merupakan aspek interpretatif; target klasifikasi model hanya Intensi Literal dan Intensi Kiasan.

Analisis Keterkaitan Konteks dengan Kesalahan Prediksi

Anotasi dilakukan dengan mempertimbangkan hubungan gambar reaksi dan konteks video, sedangkan model menggunakan pendekatan *unimodal visual* yang hanya menerima gambar sebagai masukan. Beberapa kasus salah klasifikasi, seperti RID293, RID246, dan RID226, memiliki ketergantungan konteks yang relatif tinggi.

Temuan tersebut tidak digunakan untuk menyatakan hubungan kausal bahwa ketiadaan konteks menyebabkan kesalahan prediksi. Hasil ini hanya menunjukkan bahwa pada beberapa kasus, informasi kontekstual relevan dalam menginterpretasikan intensi komunikatif dan menjadi salah satu keterbatasan pendekatan visual yang digunakan.

Pembahasan

Pencapaian Tujuan Penelitian

Tujuan pertama, yaitu menyusun dan memvalidasi skema klasifikasi Intensi Literal dan Intensi Kiasan, didukung oleh peningkatan rata-rata deskriptif *Cohen's Kappa* dari

0,389 menjadi 0,886 setelah peninjauan panduan dan *relabeling*. Jumlah *disagreement* juga berkurang dari 124 menjadi 23 data.

Tujuan kedua diwujudkan melalui penerapan CNN EfficientNetB0 dengan *transfer learning*. Model *baseline* memperoleh *accuracy* 76,47% pada *test set*, dengan *F1-score* 84,62% untuk Intensi Literal dan 50,00% untuk Intensi Kiasan.

Kemampuan Model dalam Membedakan Intensi Literal dan Intensi Kiasan

Kinerja model berbeda antar-kelas. Intensi Literal memperoleh *recall* 88,00%, sedangkan Intensi Kiasan 44,44%. Dengan demikian, pada *test set* yang digunakan, model lebih banyak mengenali sampel Literal. Perbedaan ini menunjukkan bahwa *accuracy* 76,47% perlu dibaca bersama metrik per kelas, khususnya *recall* dan *F1-score*.

Pengaruh Ketidakseimbangan Kelas

Dataset terdiri atas 245 Intensi Literal dan 89 Intensi Kiasan. Eksperimen *class weighting* dilakukan untuk menguji pengaruh pemberian bobot terhadap kelas minoritas. Hasilnya, *recall* Intensi Kiasan tetap 44,44%, sedangkan *F1-score* turun dari 50,00% menjadi 44,44% dan *accuracy* dari 76,47% menjadi 70,59%. Dengan demikian, pada konfigurasi penelitian ini, *class weighting* belum meningkatkan kinerja pada *test set*.

Keterbatasan Pendekatan Visual

Model memperoleh *accuracy* 76,47% pada *Darkjokes*, 71,43% pada *Relateable*, dan 80,00% pada *Slapstick*. Namun, *recall* Intensi Kiasan masing-masing sebesar 57,14%, 0%, dan 0%. Hasil ini tidak digunakan untuk menggeneralisasikan performa antar-subgenre karena jumlah sampel, khususnya kelas Kiasan, terbatas. Analisis hanya memberikan gambaran tambahan mengenai pola kinerja pada data pengujian.

Kontribusi Penelitian

Penelitian menggunakan pendekatan *unimodal visual*, sehingga model hanya menerima gambar reaksi, sedangkan anotator mempertimbangkan konteks video. Beberapa kasus *error analysis* menunjukkan bahwa fungsi Kiasan lebih mudah dipahami ketika gambar dikaitkan dengan konteks video. Temuan ini tidak membuktikan hubungan kausal antara ketiadaan konteks dan kesalahan model, tetapi menunjukkan relevansi informasi kontekstual pada sebagian kasus.

Selain keterbatasan konteks, jumlah data relatif terbatas, terutama pada kelas Intensi Kiasan dan beberapa subgenre. Oleh karena itu, hasil penelitian menggambarkan kinerja model pada dataset yang digunakan dan belum dapat digeneralisasikan secara luas.

Kontribusi Penelitian

Penelitian ini memberikan kontribusi melalui penyusunan skema klasifikasi Intensi Literal dan Intensi Kiasan, pengujian reliabilitas anotasi, penerapan EfficientNetB0, eksperimen *class weighting*, analisis subgenre, dan *error analysis*. Hasil menunjukkan bahwa gambar reaksi dapat menjadi objek klasifikasi intensi komunikatif berbasis visual, meskipun kinerja berbeda antar-kelas. Temuan mengenai ketergantungan konteks pada beberapa kasus dapat menjadi dasar pengembangan penelitian selanjutnya melalui pendekatan multimodal.

KESIMPULAN

Penelitian berhasil menyusun skema klasifikasi gambar reaksi pada komentar TikTok ke dalam Intensi Literal dan Intensi Kiasan melalui panduan berbasis konteks dan *expert judgement*. Pelabelan oleh tiga anotator menghasilkan rata-rata deskriptif *Cohen's Kappa* sebesar 0,389 dengan 124 *disagreement*. Setelah penegasan panduan dan *relabeling*,

nilai tersebut meningkat menjadi 0,886 dengan 23 *disagreement* yang diselesaikan melalui *majority vote*. Setelah penghapusan 5 *exact duplicate*, diperoleh 334 gambar, terdiri atas 245 Intensi Literal dan 89 Intensi Kiasan.

Klasifikasi dilakukan menggunakan EfficientNetB0 *transfer learning* dengan pembagian *stratified* 267 *training*, 33 *validation*, dan 34 *test*. Model *baseline* memperoleh *accuracy* 76,47%, *macro F1-score* 0,6731, *weighted F1-score* 0,7545, dan AUC 0,6933. Intensi Literal memperoleh *F1-score* 0,8462, sedangkan Intensi Kiasan 0,5000, menunjukkan perbedaan kinerja antar-kelas pada dataset penelitian.

Eksperimen *class weighting* tidak meningkatkan kinerja, dengan *accuracy* 70,59%, *macro F1-score* 0,6222, *F1-score* Literal 0,8000, *F1-score* Kiasan 0,4444, dan AUC 0,6844. *Recall* Intensi Kiasan tetap 44,44%, sehingga *baseline* dipertahankan sebagai model utama.

Analisis subgenre menunjukkan *accuracy* 76,47% pada *Darkjokes*, 71,43% pada *Relateable*, dan 80,00% pada *Slapstick*, dengan *recall* Intensi Kiasan masing-masing 57,14%, 0%, dan 0%. Hasil ini bersifat deskriptif karena jumlah sampel terbatas. *Error analysis* menunjukkan bahwa beberapa Intensi Kiasan bergantung pada hubungan gambar dan konteks video, sementara model hanya menggunakan gambar sebagai input. Dengan demikian, hasil penelitian menggambarkan kemampuan EfficientNetB0 pada dataset yang digunakan dan belum dapat digeneralisasikan secara luas.

Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar dan lebih seimbang, khususnya untuk Intensi Kiasan, serta mengintegrasikan gambar dan konteks video melalui pendekatan multimodal. Perbandingan EfficientNetB0 dengan arsitektur atau model visual lainnya juga dapat dilakukan. Selain itu, panduan pelabelan dapat diuji pada jumlah anotator, jenis konten, dan platform media sosial yang berbeda untuk mengevaluasi konsistensinya pada konteks komunikasi visual yang lebih beragam.

DAFTAR REFERENSI

1. J. R. Angelique and K. Kushartanti, "Analisis humor dalam film *Ngeri-Ngeri Sedap* dengan perspektif teori relevansi," *Multikultura*, vol. 3, no. 3, 2024, doi: 10.7454/multikultura.v3i3.1031.
2. A. Aryahiyyah and N. Solihati, "Kajian pragmatik tindak tutur ilokusi pada podcast Mata Najwa melalui kanal YouTube Najwa Shihab dan implikasinya terhadap pembelajaran Bahasa Indonesia di SMA," *Jurnal Onoma: Pendidikan, Bahasa, dan Sastra*, vol. 11, no. 3, pp. 3107–3117, 2025, doi: 10.30605/onoma.v11i3.6261.
3. E. Asmawati, A. Saikhu, and D. O. Siahaan, "Sentiment analysis of meme images using deep neural network based on keypoint representation," pp. 1–17, 2025.
4. J. L. Austin, *How to Do Things with Words*. Oxford, U.K.: Clarendon Press, 1962.
5. M. Boukes, G. Kuipers, Z. Lin, and S. Zannettou, "The faces and forms of pandemic humor: Exploring COVID-19 memes with visual machine learning," *Journal of Quantitative Description*, vol. 4, pp. 1–59, 2024.
6. G. Chandrasekaran, N. Antoanela, G. Andrei, M. Monica, and J. Hemanth, "Visual sentiment analysis using deep learning models with social media data," *Applied Sciences*, 2022.
7. M. R. Darwis, A. Ramadhan, D. N. Azila, S. Fatimah, A. Namar, R. A. Kasim, and M. R. Edy, "Analisis sentiment gambar pada media sosial dengan pendekatan deep learning," vol. 7, no. 2, pp. 91–95, 2024.

8. A. M. Davani, M. Díaz, and V. Prabhakaran, "Dealing with disagreements: Looking beyond the majority vote in subjective annotations," *Transactions of the Association for Computational Linguistics*, vol. 10, pp. 92–110, 2022, doi: 10.1162/tacl_a_00449.
9. S. Farabi, T. Ranasinghe, D. Kanojia, Y. Kong, and M. Zampieri, "A survey of multimodal sarcasm detection," in *Proc. IJCAI*, pp. 8020–8028, 2024, doi: 10.24963/ijcai.2024/887.
10. T. Feng, J. Liu, and J. Yang, "Probing sentiment-oriented pre-training inspired by human sentiment perception mechanism," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 2850–2860, doi: 10.1109/CVPR52729.2023.00279.
11. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org/
12. S. Z. Hassan, K. Ahmad, S. Hicks, P. Halvorsen, A. Al-Fuqaha, N. Conci, and M. Riegler, "Visual sentiment analysis from disaster images in social media," *Sensors*, vol. 22, no. 10, pp. 1–21, 2022, doi: 10.3390/s22103628.
13. M. Iman, H. R. Arabnia, and K. Rasheed, "A review of deep transfer learning and recent advancements," *Technologies*, vol. 11, no. 2, pp. 1–14, 2023, doi: 10.3390/technologies11020040.
14. J. Opitz, "A closer look at classification evaluation metrics and a critical reflection of common evaluation practice," *Transactions of the Association for Computational Linguistics*, vol. 12, pp. 820–836, 2024, doi: 10.1162/tacl_a_00675.
15. M. O. Kenneth, F. Khosmood, and A. Edalat, "Systematic literature review: Computational approaches for humour style classification," arXiv preprint arXiv:2402.01759, pp. 1–33, 2024. [Online]. Available: http://arxiv.org/abs/2402.01759
16. M. Li, Q. Gao, and T. Yu, "Kappa statistic considerations in evaluating inter-rater reliability between two raters: Which, when and context matters," *BMC Cancer*, vol. 23, pp. 1–5, 2023, doi: 10.1186/s12885-023-11325-z.
17. M. Nirmala, A. H. Gandomi, M. R. Babu, L. D. D. Babu, and R. Patan, "An emoticon-based novel sarcasm pattern detection strategy to identify sarcasm in microblogging social networks," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 4, pp. 5319–5326, 2024, doi: 10.1109/TCSS.2023.3306908.
18. N. H. Putri, M. Irsyad, S. Agustian, and F. Yanto, "Klasifikasi citra stroke menggunakan augmentasi dan convolutional neural network EfficientNet-B0," *Jurnal Media Informatika Budidarma*, vol. 7, no. 2, pp. 650–658, 2023, doi: 10.30865/mib.v7i2.5981.
19. O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, pp. 1–14, 2024, doi: 10.1038/s41598-024-56706-x.
20. J. C. Setiawan, K. M. Lhaksmana, and Bunyamin, "Sentiment analysis of Indonesian TikTok review using LSTM and IndoBERTweet algorithm," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 3, pp. 774–780, 2023.
21. M. A. Shalahuddin and L. Fajrianti, "Dampak emoticon dalam komunikasi di media sosial," *Propaganda*, vol. 4, no. 2, pp. 72–78, 2024, doi: 10.37010/prop.v4i2.1632.

22. D. Sukmana and D. Claretta, "Analisis semiotika dalam meme politik yang diunggah oleh akun TikTok @bem.ui," *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 21, pp. 215–223, 2023, doi: 10.5281/zenodo.10075171.
23. M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019.
24. S. Urolagin, J. Nayak, and U. R. Acharya, "Gabor CNN based intelligent system for visual sentiment analysis of social media data on cloud environment," *IEEE Access*, vol. 10, pp. 132455–132471, 2022, doi: 10.1109/ACCESS.2022.3228263.
25. B. S. Wiguna, C. V. Hudiyanti, A. Rausanfita, A. Z. Arifin, and R. W. Sholikah, "Sarcasm detection engine for Twitter sentiment analysis using textual and emoji feature," vol. 1, pp. 1–8, 2021.
26. H. Zhang, Y. Liu, Z. Xiong, Z. Wu, and D. Xu, "Visual sentiment analysis with semantic correlation enhancement," *Complex & Intelligent Systems*, vol. 10, no. 2, pp. 2869–2881, 2024, doi: 10.1007/s40747-023-01296-w.
27. M. Zidan, "Analisis sentimen kenaikan harga bahan bakar minyak (BBM) berdasarkan respon pengguna media sosial Twitter di Indonesia menggunakan metode Naive Bayes," Skripsi, UIN Walisongo, Semarang, Indonesia, 2022.