ISSN 2085-4552

Multiclass Emotion Detection on YouTube Comments Using IndoBERT

A Web-Based Incremental Learning System with Multiple Data Split Evaluation

Naufal Syarifuddin[1], Nurirwan Saputra[2]
[1,2] Prodi Informatika, Fakultas Sains dan Teknologi, Universitas PGRI Yogyakarta
[1] naufalsrfdn@gmail.com, [2] nurirwan@upy.ac.id

Accepted 05 January 2026
Approved 23 January 2026

***Abstract*— YouTube comment sections provide rich textual data that reflect users' emotional responses to various social, political, and entertainment-related issues. However, the large volume of user-generated comments makes manual emotion analysis inefficient and impractical. This study proposes a multiclass emotion classification approach for Indonesian YouTube comments using the IndoBERT model integrated with a database-driven incremental learning system. Comment data were collected through the YouTube Data API and manually labeled into six emotion categories: anger, sadness, happiness, fear, surprise, and neutral. Text preprocessing included lowercasing, text cleaning, and normalization of informal Indonesian words. The IndoBERT model was fine-tuned using three training–testing split scenarios (60:40, 70:30, and 80:20) to evaluate model performance and stability. Experimental results indicate that the 80:20 split achieved the best performance with an accuracy of 68%. This performance is influenced by a highly imbalanced class distribution, where minority emotion classes such as fear (2%) and surprise (3%) are significantly underrepresented compared to dominant classes. In addition to classification performance, the proposed system incorporates a database to continuously store newly collected and validated data and supports incremental retraining, enabling the model to learn from new data without discarding previously acquired knowledge. This adaptive learning mechanism allows the system to improve over time and represents a key advantage over conventional static emotion classification approaches.**

Index Terms*— emotion classification; IndoBERT; incremental learning; social media analysis; YouTube comments.

I. Introduction

YouTube is one of the largest video-sharing platforms and has evolved into a major space for public interaction through its comment sections. Users actively express opinions, attitudes, and emotional reactions toward various social, political, and entertainment-related issues in textual form[1]. These emotional expressions reflect public perception and collective responses, making YouTube comments an important data source for large-scale emotion analysis and social media research [2], [3].

Despite their analytical potential, the massive and continuously growing volume of YouTube comments makes manual emotion analysis inefficient, time-consuming, and difficult to maintain consistently[4]. In addition, the language used in YouTube comments is highly informal, frequently containing slang, abbreviations, spelling variations, and rapidly evolving expressions[5]. These characteristics increase the complexity of emotion classification and pose significant challenges for automated text analysis systems [6].

Recent advances in Natural Language Processing (NLP) have enabled the automation of emotion analysis through machine learning techniques. Early studies commonly employed traditional classifiers such as Naïve Bayes and Support Vector Machine [7]. Although these approaches achieved moderate performance, they rely heavily on handcrafted features and have limited ability to capture contextual and semantic relationships within text, particularly in informal social media environments [8].

Transformer-based models have emerged as a more effective solution by leveraging self-attention mechanisms to capture long-range contextual dependencies. Bidirectional Encoder Representations from Transformers (BERT) has demonstrated strong performance across various text classification tasks, including sentiment and emotion analysis [9]. For the Indonesian language, IndoBERT was developed as a pre-trained transformer model using large-scale Indonesian corpora and has shown superior results compared to conventional machine learning approaches in multiple NLP tasks [10].

Although IndoBERT and other transformer-based models have achieved promising results, most existing studies on emotion classification still adopt a static learning paradigm. In this setting, models are trained once using a fixed dataset and deployed without further updates. Such an approach is less suitable for social media platforms like YouTube, where language usage, slang expressions, and emotional patterns continuously evolve over time[11], [12]. Models that are not updated with new data risk becoming less relevant and experiencing performance degradation as data characteristics change [13].

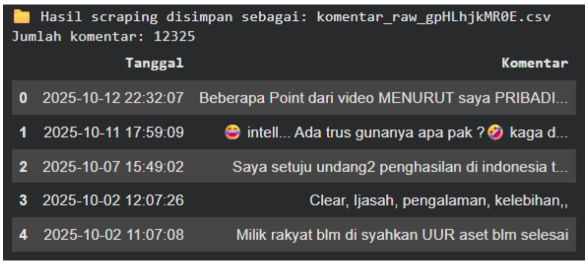
Based on the above discussion, several key problems can be identified. First, the continuously increasing volume of YouTube comments makes manual emotion analysis impractical and inconsistent. Second, the informal and dynamic nature of YouTube comment language poses significant challenges for conventional and static machine learning models. Third, most existing emotion classification studies do not address the need for adaptive learning mechanisms that allow models to evolve as new data become available. These problems indicate the necessity of an emotion classification system that is not only accurate but also adaptive to the dynamic nature of social media data.

To address the identified problems, this study aims to achieve the following objectives. First, to evaluate the performance of the IndoBERT model in multiclass emotion classification of Indonesian YouTube comments using different training–testing data split scenarios. Second, to analyze the impact of class imbalance on classification performance, particularly for minority emotion categories. Third, to develop and implement a web-based emotion classification system that integrates a database-driven incremental learning mechanism, enabling the model to continuously learn from newly validated data without discarding previously acquired knowledge.

II. METHOD

A. Data Collection

This study utilized public comments collected from a YouTube video on the Rakyat Bersuara channel. The data were obtained using the YouTube Data API v3, which provides structured access to publicly available comment data on YouTube[14]. Only comments written in Indonesian were retained, while spam content, hyperlinks, and emoji-only comments were removed to ensure data relevance and quality. This filtering process resulted in a clean dataset suitable for emotion classification tasks.



📁 Hasil scraping disimpan sebagai: komentar_raw_gpHLhjkMR0E.csv
Jumlah komentar: 12325

	Tanggal	Komentar
0	2025-10-12 22:32:07	Beberapa Point dari video MENURUT saya PRIBADI...
1	2025-10-11 17:59:09	😂 intell... Ada trus gunanya apa pak ?😂 kaga d...
2	2025-10-07 15:49:02	Saya setuju undang2 penghasilan di indonesia t...
3	2025-10-02 12:07:26	Clear, Ijasah, pengalaman, kelebihan,,
4	2025-10-02 11:07:08	Milik rakyat blm di syahkan UUR aset blm selesai

Fig 1. YouTube Comment Collection Process

B. Data Labeling

The collected comments were manually annotated using a supervised labeling approach. Each comment was assigned to one of six predefined emotion categories, namely anger, sadness, happiness, fear, surprise, and neutral. The labeling process was conducted by considering the dominant emotional expression conveyed in each comment, following emotion classification schemes commonly adopted in previous studies[15]. The labeled dataset served as ground truth data for model training and evaluation.

TABLE I. EMOTION CATEGORIES

Text	Emotion
Terus gimana? Coba lu jelasin, jgn nyocot doang	Anger
Ferry di pojokin terus. Bahkan host juga pojokin ferry. kasihan.	Sadness
Diskusi mantulll	Happiness
Indonesia hancur karna di adudomba	Fear
Daginggg semuaa gelooo	Surprise
Banyakk yang dipotong videonyaa	Neutral

C. Text Preprocessing

Text Text preprocessing was applied to reduce noise and standardize raw comment data prior to model training. The preprocessing steps included text lowercasing, removal of punctuation, numbers, URLs, and emojis, as well as normalization of informal Indonesian words using a slang dictionary[16], [17]. These steps are necessary due to the informal and unstructured nature of social media text.

Tokenization was performed using the WordPiece tokenizer provided by IndoBERT. This tokenizer represents words at the subword level and effectively handles out-of-vocabulary terms that frequently appear in user-generated content[9].

TABLE II. EXAMPLES OF TEXT PREPROCESSING RESULTS

Text	Cleaned Text
Terus gimana? Coba lu jelasin, jgn nyocot doang	terus gimana coba kamu jelasin jangan bicara doang
Ferry di pojokin terus. Bahkan host juga pojokin ferry. kasihan.	ferry di pojokin terus bahkan host juga pojokin ferry kasihan
Diskusi mantulll	diskusi mantul
Indonesia hancur karna di adudomba	indonesia hancur karena di adu domba
Daginggg semuaa gelooo	daging semua gila
Banyakk yang dipotong videonyaa	banyak yang dipotong videonya

D. IndoBERT Fine-Tuning

The IndoBERT-base model was employed as the core architecture for multiclass emotion classification. IndoBERT is a pre-trained transformer model specifically designed for the Indonesian language and trained on large-scale Indonesian corpora, enabling it to capture linguistic characteristics unique to Indonesian text [10]. Fine-tuning was performed by adding a fully connected classification layer on top of the [CLS] token representation, followed by a Softmax activation function to predict six emotion classes.

To evaluate model robustness and generalization capability, three training–testing data split scenarios were applied, namely 60:40, 70:30, and 80:20. These split configurations were selected to examine the impact of training data proportion on classification performance while maintaining a consistent evaluation framework. All experiments were conducted using identical preprocessing steps and hyperparameter settings to ensure fair comparison across scenarios.

Model optimization was carried out using the AdamW optimizer with categorical cross-entropy as the loss function, which is widely adopted for multiclass text classification tasks involving transformer-based models [18], [19].The learning rate was set to 2×10^{-5}, as this value is commonly recommended for fine-tuning BERT-based architectures to allow gradual parameter updates without disrupting the pre-trained representations. A batch size of 16 was selected to balance gradient stability and computational efficiency, particularly under hardware memory constraints typical in transformer fine-tuning.

The model was trained for three epochs, which was considered sufficient to adapt the pre-trained IndoBERT model to the emotion classification task while minimizing the risk of overfitting. Training for a larger number of epochs may lead to performance degradation on unseen data, especially in social media text classification tasks where class imbalance is present [20]. Overall, this hyperparameter configuration was chosen to provide stable convergence, efficient training, and reliable performance comparison across different data split scenarios.

TABLE III. FINE-TUNING HYPERPARAMETERS

Hyperparameter	Value
Epoch	3
Batch Size	16
Learning Rate	2e-5
Optimizer	AdamW
Max Sequence Length	256
Loss Function	Categorical Cross-Entropy

E. Evaluation Metrics

Model performance was evaluated using a confusion matrix and standard classification metrics, including accuracy, precision, recall, and F1-score. Since the classification task involved multiple emotion categories with imbalanced class distribution, macro-averaged metrics were used to provide a balanced evaluation across all classes[21]. The best-performing model was selected based on overall accuracy and macro F1-score[22].

F. Database-Driven Incremental Retraining Strategy

To accommodate the dynamic characteristics of social media data, this study implements a database-driven incremental retraining strategy as a key system contribution. A relational database (MySQL) is employed to store YouTube comments together with their predicted emotion labels, confidence scores, validation status, and timestamps. This structured data management enables the continuous accumulation of newly collected and human-validated samples, ensuring that training data can be systematically reused over time [23].

The use of a relational database is motivated by the structured nature of the stored data and the need for data consistency in the learning process. Unlike document-oriented NoSQL databases such as MongoDB, which prioritize schema flexibility, MySQL provides predefined schemas and relational constraints that ensure data integrity and traceability. These characteristics are particularly important in incremental learning scenarios, where training data must be accurately filtered, validated, and retrieved based on specific conditions.

Incremental retraining is performed by reusing the previously fine-tuned IndoBERT model as the initial model state and further fine-tuning it using newly validated data retrieved from the database. This approach allows the model to adapt to evolving linguistic patterns and emerging emotional expressions without discarding previously acquired knowledge. Compared to training from scratch, incremental retraining reduces computational cost and mitigates the risk of catastrophic forgetting, making the proposed system more suitable for long-term deployment in real-world emotion analysis applications [24], [25].

III. RESULT AND DISCUSSIONS

A. Dataset Distribution

The final dataset consisted of 5,500 Indonesian YouTube comments that were manually labeled into six emotion categories: anger, sadness, happiness, fear, surprise, and neutral. The distribution of emotion classes was imbalanced, with anger and happiness dominating the dataset, while fear and surprise appeared less frequently. Such imbalance is a common characteristic of user-generated social media data and may influence classification performance, particularly for minority emotion classes.

TABLE IV. EMOTION DISTRIBUTION

Emotion	Count	Percentage
Anger	2091	38%
Happiness	1927	35%
Fear	112	2%
Surprise	164	3%
Sadness	216	4%
Neutral	990	18%

Total	5.500	100%

B. Performance Comparison Across Data Split Scenarios

To evaluate model robustness, the IndoBERT model was trained and tested using three different data split scenarios: 60:40, 70:30, and 80:20. All experiments were conducted using identical preprocessing steps and hyperparameter settings to ensure a fair comparison.

The experimental results indicate a consistent improvement in model performance as the proportion of training data increased. The 80:20 split achieved the highest accuracy, followed by the 70:30 and 60:40 splits. This finding suggests that IndoBERT benefits from larger labeled datasets, enabling the model to better capture contextual information and emotional patterns within YouTube comments.

TABLE V. PERFORMANCE COMPARISON UNDER DIFFERENT DATA SPLIT SCENARIOS

Data Split	Accuracy
60 : 40	63.0%
70 : 30	65.7%
80 : 20	68.0%

C. Classification Report Analysis

A detailed classification report was generated to analyze model performance across individual emotion categories. The results show that the anger and happiness classes consistently achieved higher precision, recall, and F1-score values compared to other classes. This performance can be attributed to the larger number of training samples and more distinctive linguistic patterns associated with these emotions.

In contrast, the fear and surprise classes exhibited lower F1-scores due to their limited sample size and semantic overlap with other emotion categories. Similar observations have been reported in previous emotion classification studies on Indonesian social media text. Among the evaluated scenarios, the 80:20 split produced the most balanced performance, as reflected by the highest macro-averaged F1-score.

TABLE VI. CLASSIFICATION REPORT FOR THE 80:20 DATA SPLIT

Emotion	Precision	Recall	F1-Score	Support
Anger	0.72	0.76	0.74	418
Sadness	0.52	0.35	0.42	43
Happiness	0.75	0.74	0.74	385
Fear	0.69	0.48	0.56	23
Surprise	0.58	0.21	0.31	33
Neutral	0.51	0.58	0.54	198
Accuracy	—	—	**0.68**	**1100**
Macro Average	0.63	0.52	0.55	1100
Weighted Average	0.68	0.68	0.67	1100

D. Confusion Matrix Analysis

The confusion matrix for the best-performing 80:20 data split reveals that the IndoBERT model performs well primarily on two dominant emotion classes, namely anger and happiness. Most predictions for these classes are correctly located along the diagonal, indicating that the model is able to capture their distinctive linguistic patterns. In contrast, minority emotion classes such as fear, surprise, and sadness exhibit noticeably higher misclassification rates, suggesting limited predictive capability for these categories.

This imbalance in performance is mainly attributed to the highly skewed distribution of emotion classes in the dataset. While anger and happiness account for a large proportion of the training data, fear and surprise represent only a small fraction of the samples. As a result, the model is exposed to insufficient examples of minority emotions during training, which restricts its ability to learn robust and discriminative representations for these classes. When encountering ambiguous expressions, the model tends to favor dominant classes that it has learned more confidently.

Additionally, misclassifications frequently occur between semantically related emotion pairs. For instance, comments expressing fear are often confused with sadness, while surprise is sometimes misclassified as neutral. This phenomenon reflects the linguistic characteristics of informal YouTube comments, where emotional cues are often subtle, context-dependent, and expressed through shared vocabulary or short phrases. Even transformer-based models such as IndoBERT may struggle to distinguish fine-grained emotional differences under such conditions, particularly when training data are limited.

Although the overall accuracy of 68% indicates reasonable performance, the confusion matrix highlights the limitations of relying solely on aggregate metrics in the presence of class imbalance. The strong performance on dominant classes masks weaker recognition of minority emotions. These findings suggest that future improvements may be achieved through data balancing strategies or by leveraging the proposed database-driven incremental learning mechanism to gradually enrich underrepresented emotion classes over time, thereby improving classification robustness and fairness.

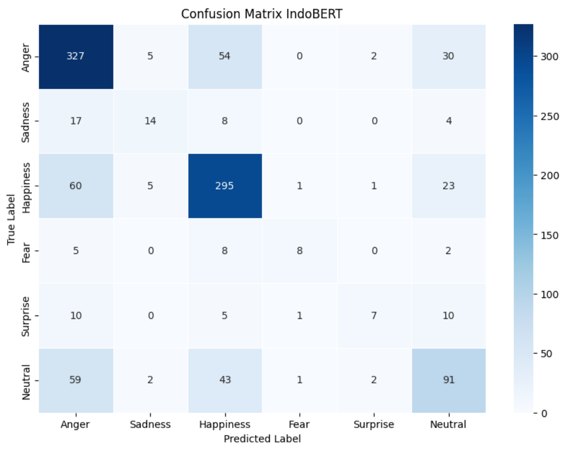


Fig. 2. Confusion Matrix of IndoBERT for the 80:20 Split

E. Discussion of Model Performance

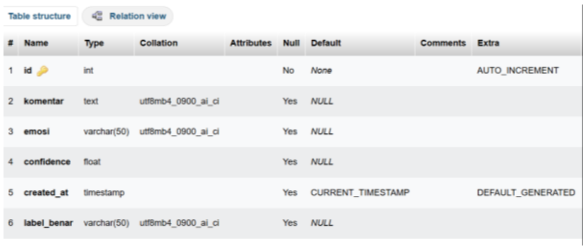
Overall, the experimental results demonstrate that IndoBERT is effective for multiclass emotion classification on Indonesian YouTube comments. The observed performance improvement as the proportion of training data increases, particularly under the 80:20 split scenario, is consistent with previous studies on transformer-based models, which rely on sufficient contextual examples to learn robust text representations.

Nevertheless, the evaluation results also reveal a noticeable performance gap between majority and minority emotion classes. As shown in Table IV, emotion categories such as fear (2%) and surprise (3%) are significantly underrepresented compared to dominant classes such as anger (38%). This extreme class imbalance limits the model's ability to learn discriminative patterns for minority classes, leading to lower recall and F1-score values despite strong overall accuracy.

Although this study does not apply data balancing techniques, the findings indicate that future performance improvements may be achieved by incorporating data augmentation or class rebalancing strategies. Techniques such as Random Oversampling or Synthetic Minority Over-sampling Technique (SMOTE) could be explored to increase the representation of minority emotion classes. In addition, the continuous data accumulation enabled by the proposed database-driven incremental learning system provides a practical pathway to gradually reduce class imbalance as more labeled data become available over time.

F. Database-Driven Incremental Retraining Analysis

Beyond conventional model evaluation, this study introduces a database-driven system architecture that enables continuous data management and incremental learning for emotion classification. The trained IndoBERT model is deployed within a web-based environment that serves as an interface for data collection, human validation, and analysis. A relational database is employed to store YouTube comments along with their predicted emotion labels, confidence scores, and validation status. This structured storage mechanism allows newly collected comments to be accumulated systematically and prepared for subsequent learning processes.



#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra
1	id	int			No	None		AUTO_INCREMENT
2	komentar	text	utf8mb4_0900_ai_ci		Yes	NULL		
3	emosi	varchar(50)	utf8mb4_0900_ai_ci		Yes	NULL		
4	confidence	float			Yes	NULL		
5	created_at	timestamp			Yes	CURRENT_TIMESTAMP		DEFAULT_GENERATED
6	label_benar	varchar(50)	utf8mb4_0900_ai_ci		Yes	NULL		

Fig. 3. Database Structure

The database plays a central role in supporting system adaptability and long-term learning. By maintaining both historical data and newly validated samples, the system ensures data consistency, traceability, and controlled data growth throughout the learning lifecycle. Unlike static emotion classification approaches that rely on a fixed dataset, the proposed system is designed to evolve over time, reflecting changes in language usage, emerging slang, and shifting emotional expressions commonly observed in YouTube comments.

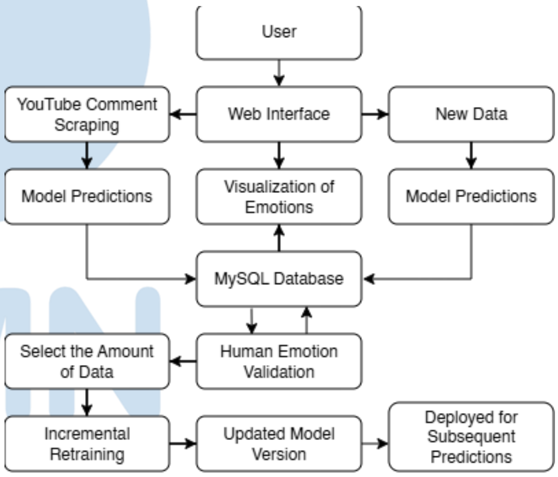


Fig. 4. Integrated System Architecture and Database-Driven Incremental Retraining Workflow

Incremental retraining is performed by reusing the previously fine-tuned IndoBERT model as the initial model state. Instead of reinitializing the model from scratch, the retraining process refines existing model parameters using newly validated data retrieved from the database. This approach enables the model to adapt to new emotional expressions while preserving previously acquired contextual knowledge.

A critical challenge in incremental learning is catastrophic forgetting, where a model loses previously learned knowledge when trained on new data. In the proposed system, this issue is mitigated by retaining the previously learned model parameters and performing controlled fine-tuning using incremental data batches. Rather than replacing existing knowledge, the retraining process refines and extends the learned representations. As a result, the incremental retraining mechanism functions as a knowledge accumulation process rather than a knowledge replacement process, enabling stable long-term learning.

In the proposed system, the retraining process is triggered after a predefined number of newly validated comments have been accumulated in the database. Specifically, retraining is performed when at least 100 new validated samples are available, ensuring that sufficient new information is provided to meaningfully update the model. The retraining procedure can be executed manually by an administrator to maintain data quality control. This design choice balances learning efficiency and computational cost, preventing excessive retraining while enabling periodic model updates as new data are collected.

Furthermore, the incremental retraining mechanism offers significant computational advantages compared to full retraining. As the volume of stored data increases—particularly for underrepresented emotion classes—the model can progressively improve its classification performance. This adaptive learning capability distinguishes the proposed system from prior emotion classification studies that employ static models and demonstrates its suitability for scalable, real-world emotion analysis on dynamic social media platforms such as YouTube.

IV. Conclusions

This study presents an adaptive multiclass emotion classification system for Indonesian YouTube comments using the IndoBERT model integrated with a database-driven incremental learning framework. Experimental results demonstrate that the best performance is achieved under the 80:20 training–testing split, with an overall accuracy of 68%, confirming that a larger proportion of labeled training data improves contextual understanding in informal social media text.

Beyond classification accuracy, the proposed system introduces a human-in-the-loop validation mechanism that selectively targets only incorrect or uncertain model predictions. Rather than validating all predictions, human annotators correct misclassified emotion labels, and these validated samples are stored back into the database. This selective validation strategy ensures that subsequent learning focuses on informative errors, improving data quality and training efficiency.

Incremental retraining is then performed by reusing the previously fine-tuned IndoBERT model as the initial state and updating it with the validated samples. This targeted retraining process allows the model to refine its decision boundaries and progressively align its predictions with human judgment, particularly for ambiguous emotional expressions. As a result, the updated model becomes more accurate and human-aligned without discarding previously acquired knowledge or incurring the high computational cost of full retraining.

Overall, the experimental findings validate that combining transformer-based language models with selective human validation and database-supported incremental learning provides an effective and scalable solution for emotion classification on YouTube comments. The proposed framework supports continuous performance improvement, mitigates catastrophic forgetting, and is suitable for long-term deployment in dynamic social media environments.

References

[1] H. Ahuja, N. Kaur, P. Kumar, and A. Hafiz, "Machine Learning based Sentiment Analysis of YouTube Video Comments," *2023 1st International Conference on Advances in Electrical, Electronics and Computational Intelligence, ICAEECI 2023*, 2023, doi: 10.1109/ICAEECI58247.2023.10370907.

[2] S. Yang, D. Brossard, D. A. Scheufele, and M. A. Xenos, "The science of YouTube: What factors influence user engagement with online science videos?," *PLoS One*, vol. 17, no. 5 May, May 2022, doi: 10.1371/JOURNAL.PONE.0267697.

[3] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, "Text-based emotion detection: Advances, challenges, and opportunities," *Engineering Reports*, vol. 2, no. 7, p. e12189, Jul. 2020, doi: 10.1002/ENG2.12189;PAGEGROUP:STRING:PUBLICATION.

[4] T. Padma, R. Visweshvar, K. Tamilarasan, and C. J. Bhadrinath, "Dynamic YouTube Comment Sentiment Analysis with Supervised Fine-Tuned BERT," *2024 International Conference on Cognitive Robotics and Intelligent Systems, ICC - ROBINS 2024*, pp. 663–669, 2024, doi: 10.1109/ICC-ROBINS60238.2024.10533926.

[5] U. Krishna, "YouTube Comments Sentiments Analysis," *Int J Res Appl Sci Eng Technol*, vol. 13, no. 1, pp. 875–880, Jan. 2025, doi: 10.22214/IJRASET.2025.66475.

[6] T. Lonkar, T. Katkar, M. Karajgar, G. Lonkar, and S. Shelar, "YouTube Comments Analyzer Using Natural Language Processing And Artificial Intelligence," *International Journal of Computer Sciences and Engineering*, vol. 12, no. 12, pp. 1–14, Dec. 2024, doi: 10.26438/IJCSE/V12I12.114.

[7] M. Z. Asghar, S. Ahmad, A. Marwat, and F. M. Kundi, "Sentiment Analysis on YouTube: A Brief Survey," Nov. 2015, Accessed: Jan. 04, 2026. [Online]. Available: https://arxiv.org/pdf/1511.09142

[8] O. El Azzouzy, T. Chanyour, and S. J. Andaloussi, "Transformer-based models for sentiment analysis of YouTube video comments," *Sci Afr*, vol. 29, p. e02836, Sep. 2025, doi: 10.1016/J.SCIAF.2025.E02836.

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of the 2019 Conference of the North*, pp. 4171–4186, Jun. 2019, doi: 10.18653/V1/N19-1423.

[10] B. Wilie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," Association for Computational Linguistics (ACL), Nov. 2020, pp. 843–857. doi: 10.18653/V1/2020.AACL-MAIN.85.

[11] M. Yeza Baihaqi, E. Halawa, R. Asyahira, S. Syah, A. Nurrahma, and W. Wijaya, "Emotion Classification in Indonesian Language: A CNN Approach with Hyperband Tuning," *Jurnal Buana Informatika*,

vol. 14, no. 02, pp. 137–146, Oct. 2023, doi: 10.24002/JBI.V14I02.7558.
[12] N. Hilmiaji, K. M. Lhaksmana, and M. D. Purbolaksono, "Identifying Emotion on Indonesian Tweets using Convolutional Neural Networks," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 3, pp. 584–593, Jun. 2021, doi: 10.29207/RESTI.V5I3.3137.
[13] M. Masana, X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov, and J. Van De Weijer, "Class-Incremental Learning: Survey and Performance Evaluation on Image Classification," *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 5, pp. 5513–5533, May 2023, doi: 10.1109/TPAMI.2022.3213473;PAGE:STRING:ARTICLE/CHAPTER.
[14] S. (santi) thomas, Y. (Yuliana) Yuliana, and N. (Noviyanti) P, "Study Analisis Metode Analisis Sentimen pada YouTube," *Journal of Information Technology*, vol. 1, no. 1, pp. 1–7, Mar. 2021, doi: 10.46229/JIFOTECH.V1I1.201.
[15] M. H. Algifari and E. D. Nugroho, "Emotion Classification of Indonesian Tweets using BERT Embedding," *Journal of Applied Informatics and Computing*, vol. 7, no. 2, pp. 172–176, Nov. 2023, doi: 10.30871/JAIC.V7I2.6528.
[16] S. Jadhav and V. Milosavljevic, "Sentiment Analysis of User comments for a YouTube Educational videos MSc Research Project Masters of Science in Data Analytics (MSCDAD_A_JAN24I)".
[17] S. Shetty, S. Shetty, P. Kamath B, and D. Shetty, "SENTIMENT PATTERNS IN YOUTUBE COMMENTS: A COMPREHENSIVE ANALYSIS," *Computer Science & Engineering: An International Journal (CSEIJ)*, vol. 15, no. 1, 2025, doi: 10.5121/cseij.2025.15119.
[18] Y. Shao *et al.*, "An Improved BGE-Adam Optimization Algorithm Based on Entropy Weighting and Adaptive Gradient Strategy," *Symmetry (Basel)*, vol. 16, no. 5, p. 623, May 2024, doi: 10.3390/sym16050623.
[19] A. Mao, M. Mohri, and Y. Zhong, "Cross-Entropy Loss Functions: Theoretical Analysis and Applications," 2023.
[20] Y. O. Sihombing, R. Fuad Rachmadi, S. Sumpeno, and M. J. Mubarok, "Optimizing IndoRoBERTa Model for Multi-Class Classification of Sentiment & Emotion on Indonesian Twitter," *Proceeding - IEEE 10th Information Technology International Seminar, ITIS 2024*, pp. 12–17, 2024, doi: 10.1109/ITIS64716.2024.10845566.
[21] I. Markoulidakis and G. Markoulidakis, "Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis," *Technologies (Basel)*, vol. 12, no. 7, p. 113, Jul. 2024, doi: 10.3390/technologies12070113.
[22] J. H. Cabot and E. G. Ross, "Evaluating prediction model performance," *Surgery (United States)*, vol. 174, no. 3, pp. 723–726, Sep. 2023, doi: 10.1016/j.surg.2023.05.023.
[23] H. A. Mumtahana, "Optimization of Transaction Database Design with MySQL and MongoDB," *SinkrOn*, vol. 7, no. 3, pp. 883–890, Jul. 2022, doi: 10.33395/sinkron.v7i3.11528.
[24] R. Wang, J. Fei, R. Zhang, M. Guo, Z. Qi, and X. Li, "DRnet: Dynamic Retraining for Malicious Traffic Small-Sample Incremental Learning," *Electronics 2023, Vol. 12, Page 2668*, vol. 12, no. 12, p. 2668, Jun. 2023, doi: 10.3390/ELECTRONICS12122668.
[25] M. Masana, X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov, and J. Van De Weijer, "Class-incremental learning: survey and performance evaluation on image classification," *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 5, pp. 5513–5533, Oct. 2020, doi: 10.1109/TPAMI.2022.3213473