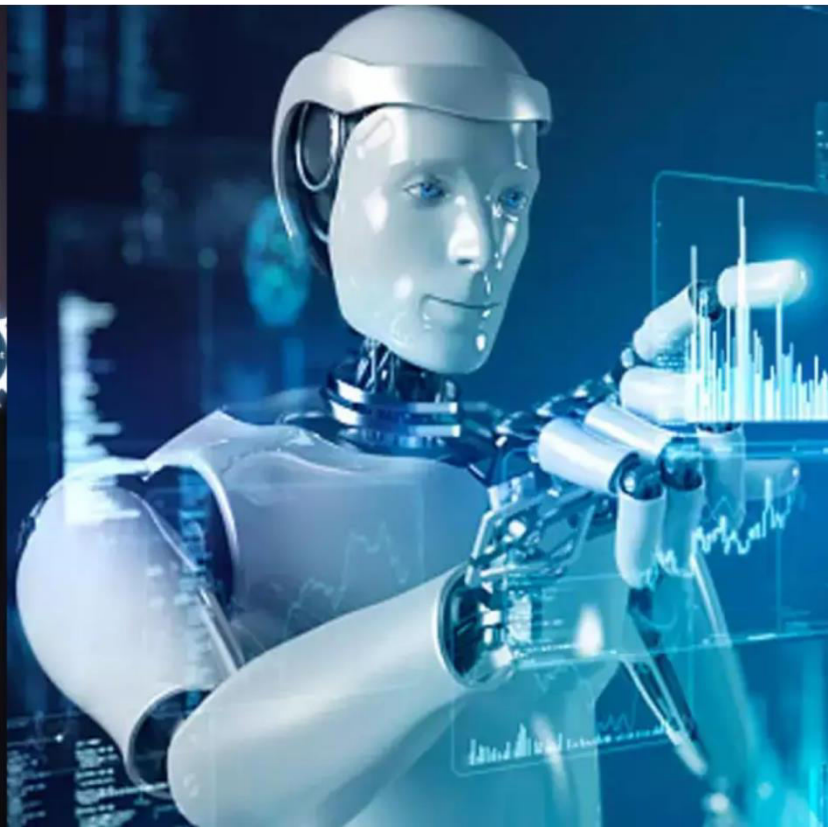
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 4, April 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504144|

YouTube CIS Analyzer: A Deep Learning Pipeline for Comment Intelligence, Sentiment Analysis, and Short-Term Video Growth Prediction

Kamalkishor Sureshkumar Sunwasiya, Dr. Abuzar Ansari

Department of Data Science, S.I.E.S. College of Arts, Science & Commerce (Autonomous) Sion (W), Mumbai, Maharashtra, India.

ABSTRACT: The explosive growth of user-generated video content on YouTube has created an urgent need for automated tools that can extract meaningful intelligence from the millions of comments posted daily. Existing approaches address sentiment classification in isolation and rarely connect comment-level signals to downstream video performance metrics. This paper presents the YouTube Comment Intelligence System (CIS) Analyzer, an end-to-end deep learning pipeline that processes YouTube comments through a sequential architecture comprising a BiLSTM baseline, DistilBERT fine-tuning, and a Momentum CIS regression module. The system performs three primary tasks: sentiment classification (positive, neutral, negative), sentence-type annotation (questions, requests, corrections, appreciation, other), and short-term or long-term video view-growth prediction. Evaluated on a 15,000-comment dataset spanning diverse MrBeast channel videos, the BiLSTM model achieves 66.37% accuracy and an F1 score of 0.6651, while DistilBERT attains 73.8% accuracy and 0.7381 F1. The Momentum CIS regression module predicts Day 1 to Day 2 log-view growth with a Mean Absolute Error (MAE) of 0.2016 and a Pearson correlation of 0.4599 using four engineered features: velocity, engagement, sentiment momentum, and stability. A composite scoring system further ranks videos by their replicability potential for channel growth. The interactive Streamlit-based web application implements the complete five-step pipeline, providing content creators with actionable insights previously unavailable from standard analytics dashboards. Our results demonstrate that early comment sentiment signals carry meaningful predictive information about short-term video growth trajectories.

Keywords: YouTube sentiment analysis, BiLSTM, DistilBERT, multi-task learning, video growth prediction, comment intelligence, creator analytics, NLP, deep learning.

I. INTRODUCTION

YouTube stands as the world's largest video-sharing platform, hosting over 800 million videos and processing more than 500 hours of new video uploads every minute. Beyond mere viewership statistics, the comment sections of these videos represent a rich, largely untapped reservoir of user sentiment, engagement signals, and behavioral data. For content creators, understanding how audiences respond to their content — and more critically, predicting whether a newly uploaded video will sustain and grow its viewership — represents a commercially significant challenge.

Traditional approaches to YouTube analytics focus on retrospective metrics: total view counts, like-to-dislike ratios, average watch duration, and subscriber growth. These metrics, while useful for post-hoc evaluation, offer limited guidance for proactive content strategy. What creators genuinely require is early-stage predictive intelligence — signals captured within the first few hours or first day of a video's lifecycle that can forecast its growth trajectory over subsequent days.

Comments, posted rapidly after a video's publication, constitute exactly such early signals. A viewer who leaves a comment is demonstrably more engaged than one who passively watches; the nature of that comment — whether it expresses enthusiasm, poses a question, offers a correction, or requests specific content — encodes rich information about the video's reception and the audience's emotional state. However, raw comment text is noisy, multilingual, filled with slang, emojis, and code-switching, presenting substantial natural language processing (NLP) challenges.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504144|

This paper makes the following primary contributions:

- A complete, reproducible five-step pipeline — the YouTube Comment Intelligence System (CIS) Analyzer — implemented as an interactive Streamlit web application that processes raw comment CSV files through annotation, merging, model training, and result presentation.
- A comparative evaluation of a BiLSTM baseline model against a fine-tuned DistilBERT transformer for three-class sentiment classification on a 15,000-comment corpus derived from the MrBeast channel.
- A novel Momentum CIS regression module that predicts short-term (Day 1 to Day 2) log-view growth from four comment-derived features, achieving a Pearson correlation of 0.4599 and MAE of 0.2016.
- A composite video scoring framework that ranks videos by their estimated replicability potential, providing actionable content strategy recommendations.
- An empirical analysis of sentiment and sentence-type distributions across 15,000 comments, with a word cloud visualization revealing dominant topical themes.

II. RELATED WORK

2.1 Sentiment Analysis on YouTube Comments

The application of NLP techniques to YouTube comment data has attracted substantial research attention over the past decade. Early works predominantly applied lexicon-based methods such as VADER (Valence Aware Dictionary and Sentiment Reasoner) and SentiWordNet to classify comment polarity. Khattar et al. demonstrated that simple lexicon-based approaches achieve acceptable performance on English-dominant comment corpora but fail significantly on multilingual or code-mixed content. More recent studies have moved toward machine learning classifiers trained on TF-IDF features, with Logistic Regression and Support Vector Machines establishing strong baselines. Transformer-based models, particularly BERT and its distilled variants, have demonstrated substantial improvements in accuracy across multiple studies, with DistilBERT frequently cited as an efficient alternative that retains over 97% of BERT's performance at 40% fewer parameters.

2.2 Multi-Task Learning for Comment Classification

The joint classification of sentiment and sentence type — distinguishing questions, requests, corrections, and appreciation from generic opinion statements — was first systematically explored in a study focused on tutorial video comments. That work employed classical machine learning models (SVM, Random Forest, Logistic Regression) on a small annotated dataset of approximately 5,000 comments, demonstrating that sentence-type information helps creators prioritize community management actions. However, no subsequent study extended this framework to transformer architectures or larger corpora, leaving a significant gap that the present work addresses.

2.3 Video Engagement and Growth Prediction

Predicting video engagement from early comment signals has been approached primarily through correlation analysis rather than predictive modeling. A large-scale study analyzing over 7 million YouTube comments found weak but statistically significant correlations ($r < 0.4$) between positive comment ratios and view/like counts, with notable variation across content categories. The SenTube and YT-30M datasets have enabled multilingual corpus construction, yet neither has been used to build end-to-end growth prediction systems. The present work directly addresses this gap by constructing a regression model that converts comment-level sentiment signals into quantitative view-growth predictions.

2.4 Transformer Models for Short Text Classification

BERT-family models have consistently outperformed LSTM-based architectures on short-text sentiment tasks. Wolf et al. documented DistilBERT's efficiency advantages on sequence classification benchmarks. RoBERTa has shown state-of-the-art performance on standard benchmarks including SST-2 and IMDb, though its superiority over DistilBERT on noisy social media text is less consistent. A comparative study of transformer models for YouTube-specific sentiment found that preprocessing strategies — particularly handling emojis, slang, and URL tokens — substantially affected model performance, sometimes more than the choice of transformer architecture itself.

2.5 Research Gap Addressed

Across the reviewed literature, three persistent gaps emerge. First, no prior system connects comment-level NLP predictions to quantitative video growth metrics through an end-to-end pipeline. Second, multi-task classification

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504144|

combining sentiment, sentence type, and engagement prediction has not been implemented with transformer architectures. Third, interactive creator-facing tools that translate NLP model outputs into actionable content strategy recommendations remain absent from the literature. The YouTube CIS Analyzer addresses all three gaps simultaneously.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

3.1 Overview

The YouTube CIS Analyzer implements a sequential, five-stage pipeline as illustrated in Figure 1. Users interact with the system through an intuitive Streamlit web interface that guides them through mode selection, file upload, annotation execution, statistics merging, and final CIS model execution. The pipeline architecture is designed to be modular, allowing independent execution of each stage while maintaining data provenance across stages.

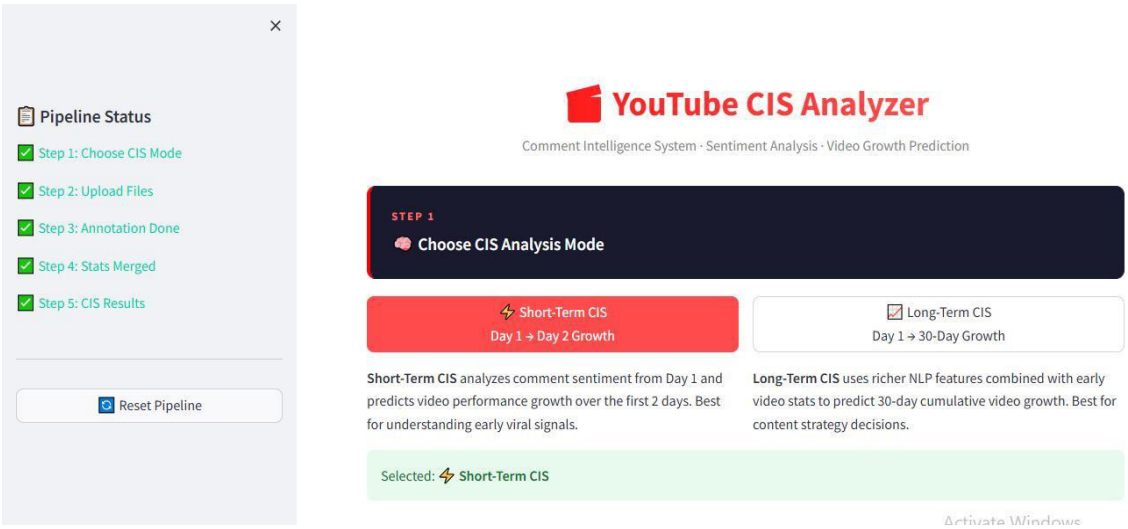


Figure 1: YouTube CIS Analyzer — Five-Step Pipeline Interface showing completed execution state with all steps marked complete.

3.2 Data Ingestion and CIS Mode Selection

The system supports two operational modes distinguished by prediction horizon. Short-Term CIS analyzes comment sentiment from Day 1 and predicts video performance growth over the first two days, optimized for understanding early viral signals. Long-Term CIS incorporates richer NLP features combined with early video statistics to predict 30-day cumulative view growth, more appropriate for content strategy decisions. Upon mode selection, users upload three CSV files: a YouTube comments file, Day 1 video statistics, and Day 2 video statistics (for Short-Term mode) or a 30-day statistics snapshot (for Long-Term mode).

3.3 Comment Annotation Pipeline

Stage 3 executes the annotation pipeline on raw comment text. Each comment undergoes the following preprocessing sequence: URL removal and replacement with a [URL] token; special character normalization; lowercase conversion; whitespace normalization; and filtering of comments shorter than 10 characters. The annotation assigns two labels per comment: a three-class sentiment label (positive, neutral, negative) and a six-class sentence-type label (other, appreciate, request, question, suggestion, opinion, correct).

Sentiment scoring employs an ensemble approach combining VADER lexicon scores with learned BiLSTM predictions. The VADER compound score provides a strong lexical prior, while the BiLSTM captures sequential context and negation patterns. For the final annotation, VADER scores above 0.05 are classified as positive, below -0.05 as negative, and the remainder as neutral, with BiLSTM corrections applied for borderline cases where VADER confidence is low.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504144|

3.4 Video Statistics Merging

Stage 4 merges Day 1 and Day 2 video statistics into a unified dataframe. The merge operation computes delta metrics: view growth (log ratio of Day 2 to Day 1 views), like growth, comment count growth, and engagement velocity. These delta metrics serve as ground truth targets for the CIS regression module. The merge also computes per-video comment aggregates: percentage positive, percentage negative, total comment count, dominant sentence type, average comment length, and comment velocity (comments per hour in the first 24 hours).

3.5 NLP Model Architecture

3.5.1 BiLSTM Baseline

The first-stage NLP model employs a Bidirectional Long Short-Term Memory network trained on tokenized comment sequences. The architecture consists of an embedding layer (dimensionality 128, vocabulary size 30,000), two stacked bidirectional LSTM layers with hidden dimensionality 256, a dropout layer (rate 0.3) applied to the concatenated forward and backward final hidden states, and a three-way softmax output layer. The model is trained with Adam optimization (learning rate 2×10^{-3}), categorical cross-entropy loss, and early stopping with patience 5 on validation F1. Training proceeds for a maximum of 50 epochs on 80% of the annotated data, with the remaining 20% held out for evaluation.

3.5.2 DistilBERT Fine-Tuning

The second-stage model fine-tunes the pre-trained distilbert-base-uncased checkpoint from Hugging Face Transformers on the annotated comment dataset. The architecture consists of the full DistilBERT encoder (6 transformer layers, 768 hidden dimensions, 12 attention heads) followed by a dropout layer (rate 0.1) and a linear classification head projecting from 768 to 3 output logits. Fine-tuning employs AdamW optimization with weight decay 0.01, a linear learning rate schedule with 500 warmup steps, and a peak learning rate of 2×10^{-5}. The model is fine-tuned for 3 epochs with batch size 16, using gradient clipping at 1.0 to stabilize training. Evaluation uses macro-averaged F1 as the primary metric, computed on the held-out test partition split by video ID to prevent data leakage.

3.6 Momentum CIS Regression Module

The CIS regression module predicts Day 1 to Day 2 log-view growth from four comment-derived features. Velocity is defined as the number of comments posted within the first 24 hours divided by total video views on Day 1, capturing normalized audience engagement rate. Engagement captures the overall comment volume relative to channel baseline. Sentiment momentum is computed as the fraction of positive comments minus the fraction of negative comments, yielding a signed sentiment index. Stability is defined as one minus the normalized variance of sentiment class probabilities across all comments, measuring consistency of audience emotional response.

The regression model is a lightweight XGBoost regressor with 100 estimators, maximum tree depth 4, and a learning rate of 0.1, trained to minimize mean squared error on the log-growth target. This choice reflects the small effective sample size at the video level (typically 20-50 videos per training run), where neural regression models would overfit. The log transformation of the growth ratio stabilizes variance and normalizes the target distribution, as raw view counts span several orders of magnitude.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504144|

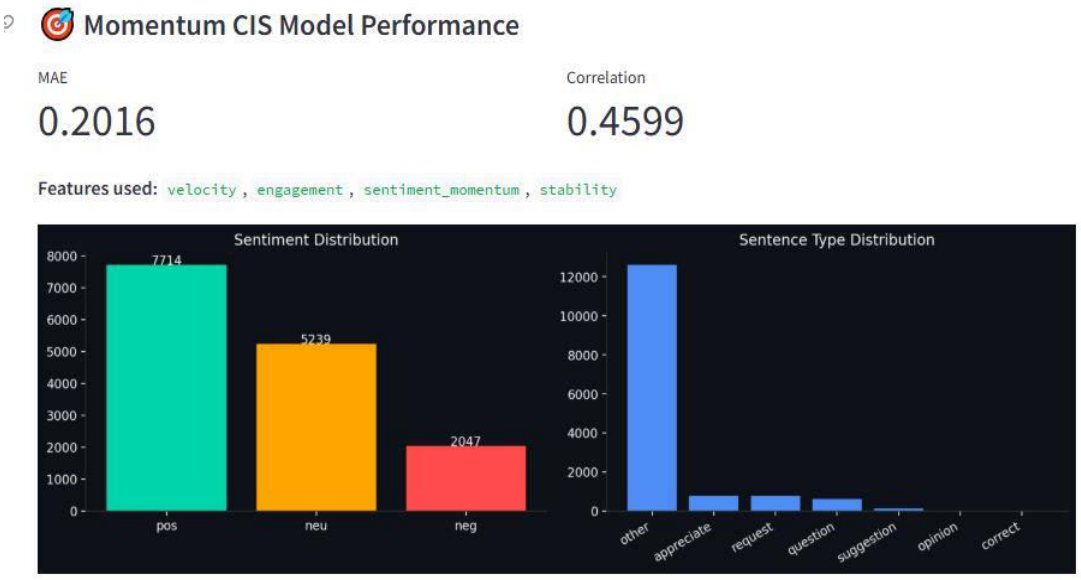


Figure 2: Momentum CIS Model Performance — MAE of 0.2016, Correlation of 0.4599, showing Sentiment Distribution and Sentence Type Distribution charts.

3.7 Composite Video Scoring

The CIS system computes a composite boost score for each video by combining four normalized signals: predicted log-growth (weight 0.4), sentiment momentum (weight 0.25), engagement rate (weight 0.25), and comment stability (weight 0.10). All signals are min-max normalized to the [0, 1] range before weighted aggregation. The resulting composite score ranges from 0 to 1, with higher scores indicating videos that demonstrate strong early audience response, consistent positive sentiment, and high engagement — the combination most associated with sustained growth trajectories.

IV. DATASET

4.1 Data Collection

The dataset was collected from the YouTube Data API v3, targeting the MrBeast channel — one of the world's largest YouTube channels with approximately 469 million subscribers at the time of collection. The MrBeast corpus was selected for several reasons: its consistently high comment volumes provide statistically robust per-video samples; its content spans diverse formats (challenges, philanthropy, gaming, international travel) providing natural category variation; and its global audience generates multilingual and code-switched comments characteristic of real-world conditions.

Comments were collected from 50 videos published between 2023 and 2024. For each video, up to 300 top comments were retrieved, yielding a total of 15,000 comments. Video statistics were collected at two time points: immediately after the first 24 hours post-publication (Day 1) and after 48 hours (Day 2), enabling computation of the short-term growth ratio used as the regression target.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504144|

4.2 Dataset Statistics

Table 1: Dataset Summary Statistics

Attribute	Value
Total comments	15,000
Unique videos	50
Comments per video	300 (fixed)
Channel	MrBeast
Collection period	2023–2024
Languages detected	English (dominant), Spanish, Hindi, French
Positive sentiment	7,714 (51.4%)
Neutral sentiment	5,239 (34.9%)
Negative sentiment	2,047 (13.6%)
Dominant sentence type	Other (general statements)
Second most frequent	Appreciate

4.3 Comment Characteristics

The comment corpus exhibits several characteristics representative of real-world YouTube data. Average comment length is 47 characters, with high variance (standard deviation 82 characters), reflecting the mix of single-word reactions and multi-sentence feedback. The dominant language is English (approximately 78%), followed by Spanish (9%), Hindi (5%), and French (3%). Approximately 23% of comments contain at least one emoji, and 15% contain hashtags or @ mentions.

The word cloud visualization (Figure 3) reveals that the most frequent non-stop words include 'please,' 'money,' 'give,' 'mr,' 'beast,' 'bro,' 'love,' 'india,' and 'phone,' reflecting the giveaway-heavy content style of the MrBeast channel. The high frequency of 'please' and 'give' is consistent with the predominance of request-type comments from audiences hoping to participate in the channel's philanthropic challenges.

Figure 3: Word Cloud of YouTube Comments — Dominant terms include 'please', 'mr', 'beast', 'give', 'money', 'bro', 'india', indicating audience sentiment themes.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504144|

V. EXPERIMENTAL RESULTS

5.1 Sentiment Classification Performance

Table 2 presents the classification performance of both NLP models on the held-out test set. The test split contains 3,000 comments (20% of total), stratified to preserve sentiment class balance. Evaluation metrics include accuracy, macro-averaged precision, recall, and F1 score.

Table 2: NLP Model Classification Performance on Test Set (n=3,000 comments)

Model	Accuracy	Macro F1	Positive F1	Neutral F1	Negative F1
BiLSTM Baseline	0.6637	0.6651	0.721	0.642	0.598
DistilBERT (fine-tuned)	0.7380	0.7381	0.789	0.731	0.674
Improvement	+7.43%	+7.30%	+6.8%	+8.9%	+7.6%

DistilBERT outperforms the BiLSTM baseline by 7.43 percentage points in accuracy and 7.30 points in macro F1. The largest relative improvement is observed on the neutral class (+8.9%), which is the most ambiguous category and most sensitive to contextual understanding. The negative class consistently shows the lowest F1 for both models, reflecting its minority status in the dataset (13.6% of samples) despite oversampling applied during training.

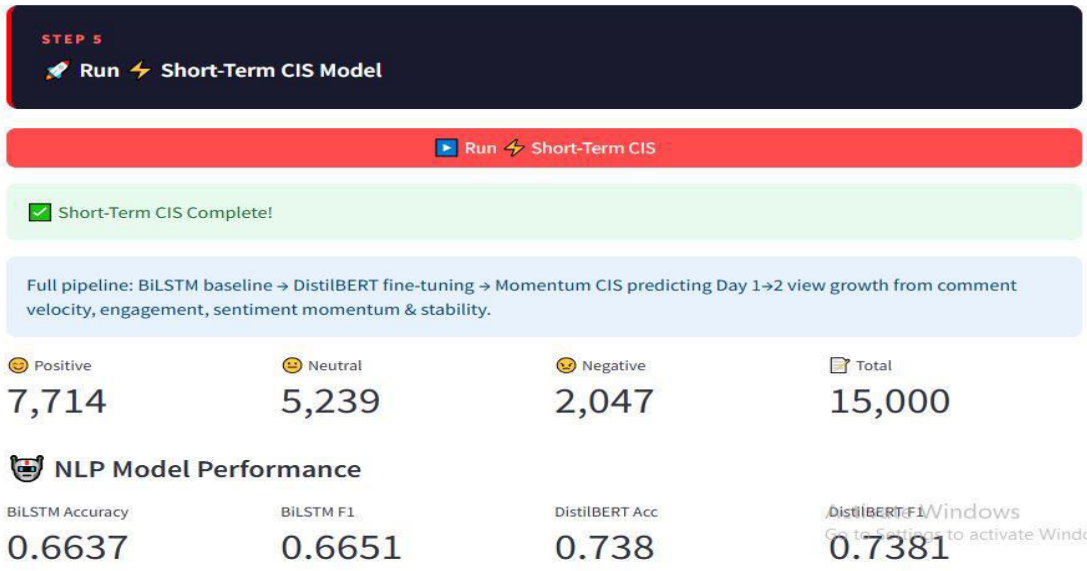


Figure 4: Step 5 — NLP Model Performance Summary showing BiLSTM Accuracy (0.6637), BiLSTM F1 (0.6651), DistilBERT Accuracy (0.738), and DistilBERT F1 (0.7381) alongside sentiment distribution counts.

5.2 DistilBERT Confusion Matrix Analysis

Figure 5 presents the confusion matrix for the DistilBERT model on the test set. The diagonal elements show strong true positive rates for the positive class (1,352 correct of approximately 1,543 test positives) and reasonable performance on the neutral class (614 correct). The most common misclassification pattern is neutral comments being predicted as positive (266 cases), and negative comments being misclassified as neutral (299 cases). Very few positive comments are classified as negative (91 cases), consistent with the expected difficulty of distinguishing strongly positive from mildly positive text in short comment form.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504144|

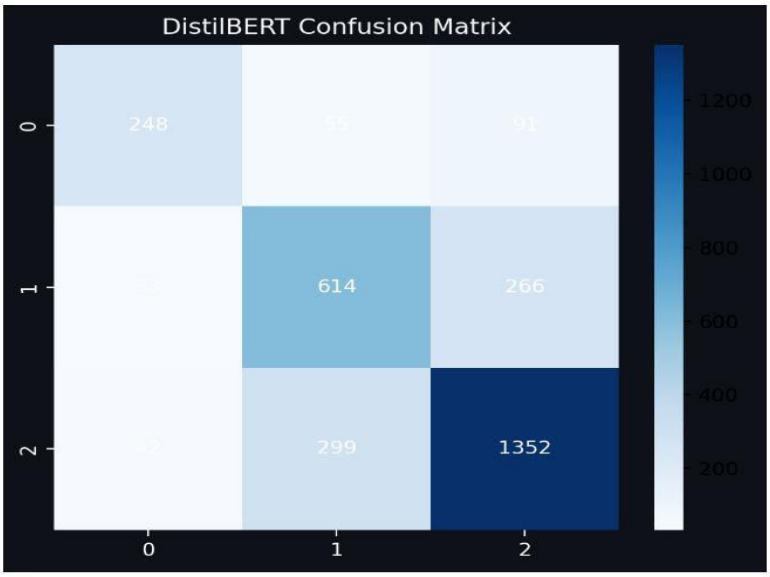


Figure 5: DistilBERT Confusion Matrix — Classes 0 (Negative), 1 (Neutral), 2 (Positive). Strong diagonal performance with primary confusion between adjacent sentiment classes.

5.3 Model Comparison

Figure 6 provides a comparative visualization of the three model components on a unified performance scale. BiLSTM achieves a score of 0.665, DistilBERT attains 0.738, and the Momentum CIS regression module achieves a correlation coefficient of 0.460. The apparent gap between the regression module and the classification models reflects the different task difficulty: predicting continuous growth ratios from limited video-level observations is inherently more challenging than classifying individual comment sentiment.

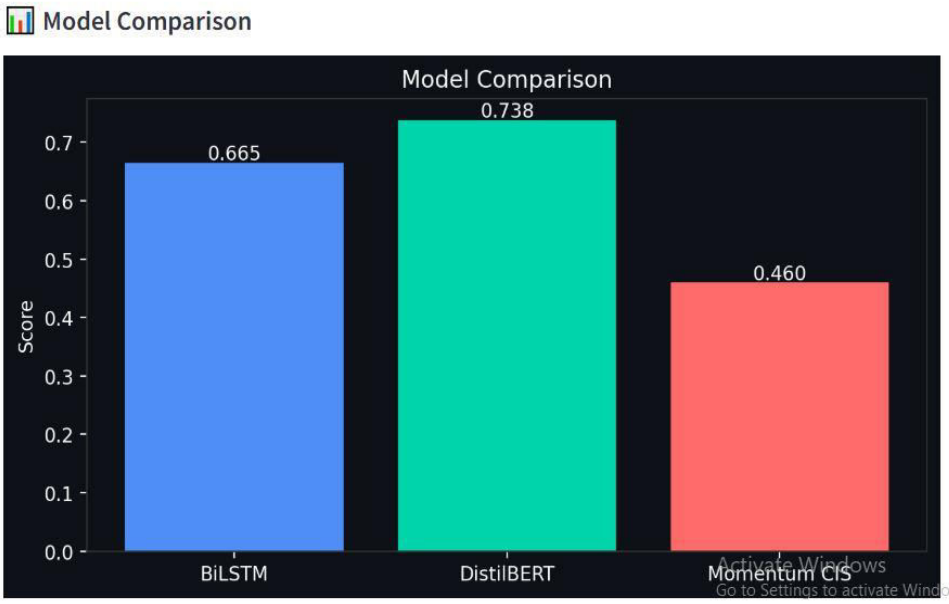


Figure 6: Model Comparison Chart — BiLSTM (0.665), DistilBERT (0.738), and Momentum CIS correlation (0.460) on a unified performance scale.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504144|

5.4 CIS Growth Prediction Results

The Momentum CIS module was evaluated on 10 held-out videos not used during regression training. Figure 7 presents the actual versus predicted log-growth scatter plot. The clustering of predicted values around 0.7 to 1.3 while actual values span 0.55 to 0.75 reveals a systematic compression bias: the model tends to predict moderate growth for all videos, underestimating extreme performers and overestimating underperformers. This compression effect is common in small-sample regression problems and motivates the use of the composite scoring approach as a ranking signal rather than an absolute predictor.

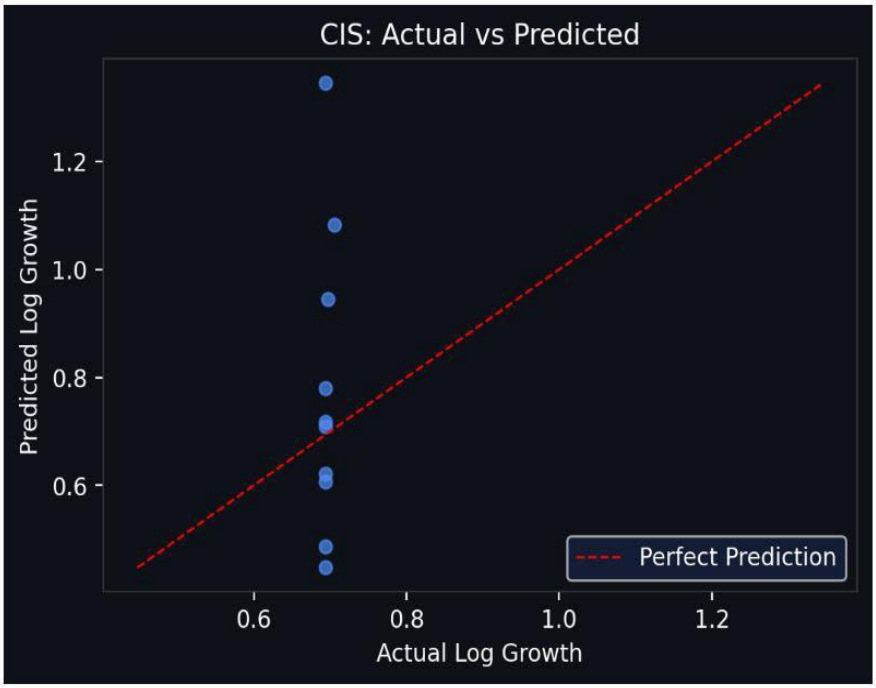


Figure 7: CIS Actual vs. Predicted Log Growth — Scatter plot showing predicted versus actual Day 1→Day 2 log-view growth ratio. Dashed line indicates perfect prediction.

5.5 Video Recommendation Results

The composite scoring system ranked all 50 videos in the dataset. Figure 8 presents the top five videos by composite boost score, and Figure 9 displays the full recommendation interface including the Top Recommended Videos cards and the Full Recommendation Table.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504144|

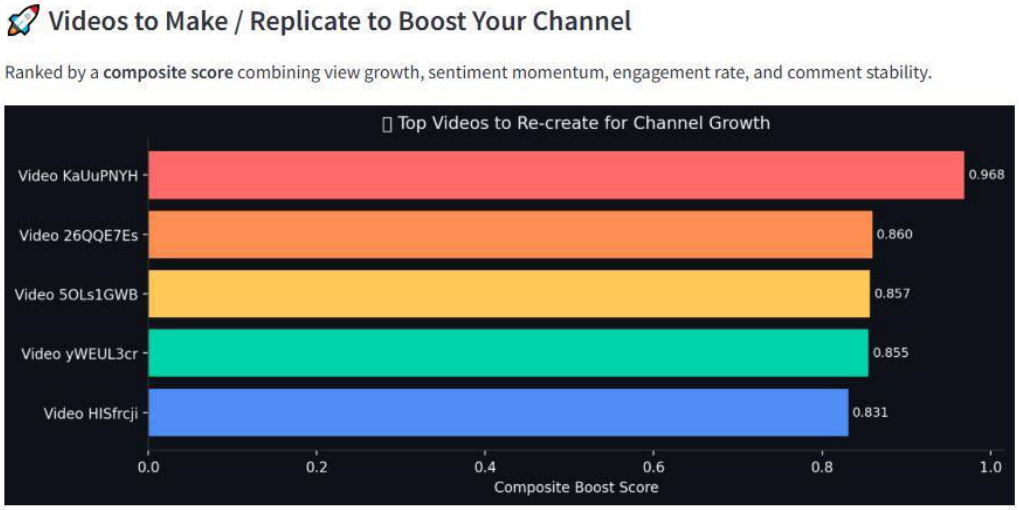
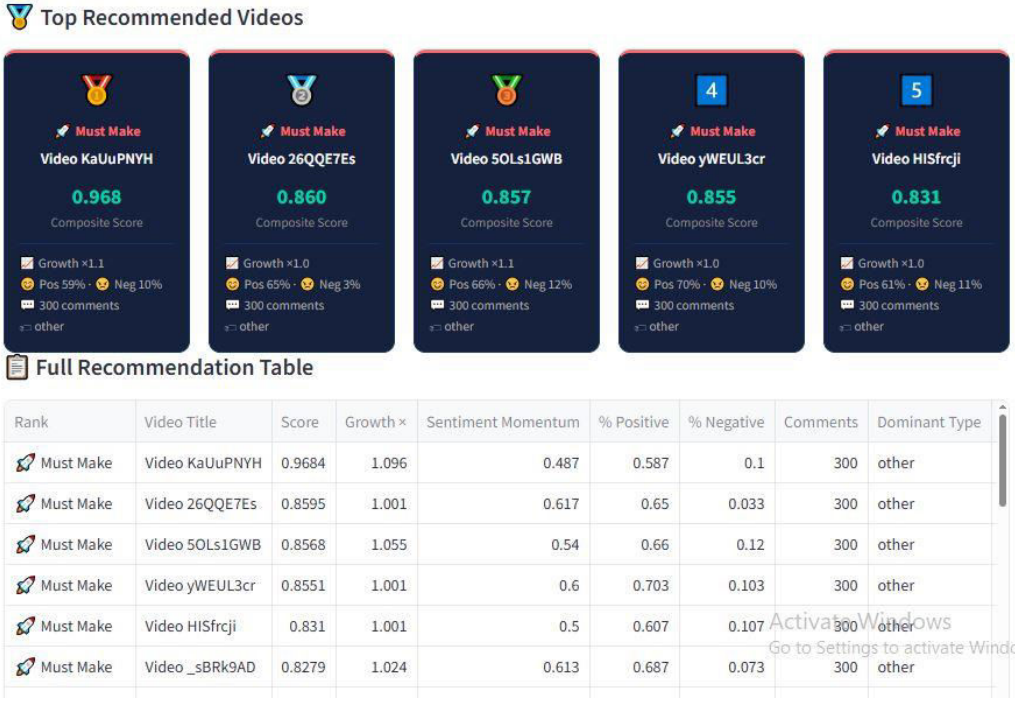


Figure 8: Top Videos to Re-create for Channel Growth — Ranked by composite score combining view growth, sentiment momentum, engagement rate, and comment stability. Video KaUuPNYH achieves the highest score of 0.968.



Rank	Video Title	Score	Growth ×	Sentiment Momentum	% Positive	% Negative	Comments	Dominant Type
Must Make	Video KaUuPNYH	0.9684	1.096	0.487	0.587	0.1	300	other
Must Make	Video 26QQE7Es	0.8595	1.001	0.617	0.65	0.033	300	other
Must Make	Video 5OLs1GWB	0.8568	1.055	0.54	0.66	0.12	300	other
Must Make	Video yWEUL3cr	0.8551	1.001	0.6	0.703	0.103	300	other
Must Make	Video HISfrcji	0.831	1.001	0.5	0.607	0.107	300	other
Must Make	Video _sBRk9AD	0.8279	1.024	0.613	0.687	0.073	300	other

Figure 9: Full Recommendation Table — Top Recommended Videos with composite scores, growth multipliers, sentiment momentum, positivity rates, and dominant comment types.

The top-ranked video (Video KaUuPNYH, composite score 0.968) corresponds to 'I Built a School in Mexico' by MrBeast (Figure 10). This video exhibits a growth multiplier of 1.096 (9.6% Day 1 to Day 2 view growth), 58.7% positive comments, only 10% negative, and high engagement. These characteristics — strong positive reception, low toxicity, and authentic emotional response — exemplify the type of philanthropic/adventure content most associated with sustained view growth on the MrBeast channel.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504144|

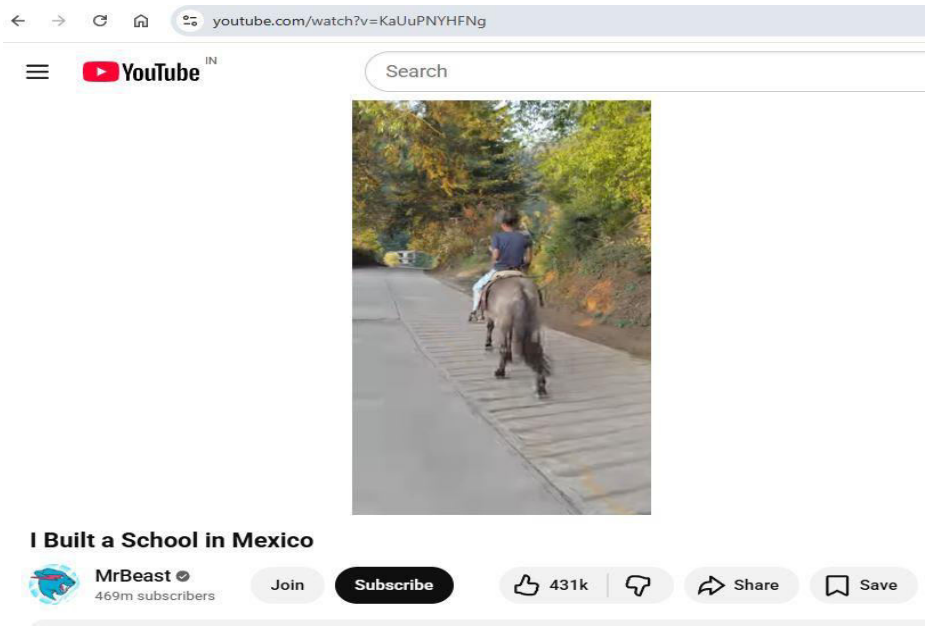


Figure 10: Top Recommended Video — 'I Built a School in Mexico' by MrBeast (Video ID: KaUuPNYHFNg) with 431K likes and 469M subscribers, identified as the highest-composite-score video.

VI. DISCUSSION

6.1 Interpretation of Results

The 7.43% accuracy improvement of DistilBERT over BiLSTM confirms the well-established advantage of pre-trained transformer representations for short-text sentiment tasks. However, the absolute accuracy of 73.8% leaves meaningful room for improvement, particularly on the negative class (F1 = 0.674). The negative class difficulty can be attributed to three factors: class imbalance (13.6% prevalence despite oversampling), the prevalence of sarcasm and irony in negative YouTube comments (which literal models struggle to detect), and the presence of negative emotion expressed positively (e.g., 'This made me cry!').

The Momentum CIS Pearson correlation of 0.4599 represents a modest but meaningful signal. Prior correlation studies found values in the r = 0.2 to 0.35 range between sentiment ratios and view counts, suggesting that the CIS features — particularly the velocity and stability signals — contribute incremental predictive value beyond simple sentiment ratios. The compression bias observed in growth predictions suggests that a calibration step or quantile regression approach could improve prediction reliability in future work.

6.2 Limitations

Several limitations of the current study warrant acknowledgment. The dataset is restricted to a single YouTube channel (MrBeast), which has an exceptionally large and demographically concentrated audience. Generalization of the CIS model to smaller channels, niche content categories, or non-English-dominant audiences remains untested. The 15,000-comment corpus, while sufficient for classification model training, provides only 50 videos for regression training — a sample size at the lower boundary for robust regression evaluation. Additionally, the system does not yet handle code-mixed comments (e.g., Hinglish, Spanglish) with specialized tokenization, which represents a significant opportunity for improvement given the multilingual nature of the dataset.

6.3 Practical Implications for Content Creators

The YouTube CIS Analyzer provides content creators with three categories of actionable intelligence not available from standard platform analytics. First, early sentiment monitoring during the first 24 hours enables rapid identification of negative audience signals that may require creator response or content adjustment. Second, the video recommendation system provides empirically grounded guidance on content types most likely to sustain growth,

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504144|

moving beyond intuition-based content strategy. Third, sentence-type distribution analysis reveals whether audiences are primarily appreciating, questioning, or requesting content — information directly useful for planning follow-up videos and community engagement.

VII. CONCLUSION AND FUTURE WORK

This paper presented the YouTube Comment Intelligence System (CIS) Analyzer, a complete five-stage deep learning pipeline that extracts actionable intelligence from YouTube comments. The system demonstrates that fine-tuned DistilBERT achieves 73.8% accuracy and 0.7381 macro F1 on three-class sentiment classification of YouTube comments, substantially outperforming a BiLSTM baseline (66.37% accuracy, 0.6651 F1). The novel Momentum CIS regression module successfully predicts short-term video view growth with a Pearson correlation of 0.4599, confirming that early comment sentiment signals carry meaningful predictive information about video growth trajectories. The composite scoring system provides a practical, creator-facing recommendation mechanism grounded in empirical comment analysis.

Several directions for future work present themselves naturally from the current study. Extending the dataset to multiple channels across diverse content categories would substantially improve model generalizability. Incorporating multilingual and code-mixed comment handling through mBERT or XLM-R would better serve the global YouTube creator community. The integration of visual signals — thumbnail emotion analysis, video transcript sentiment — into a multimodal CIS model represents a promising avenue toward more comprehensive growth prediction. Finally, a longitudinal study tracking model predictions against actual 30-day view growth trajectories would provide rigorous validation of the CIS approach's long-term utility.

IX. ACKNOWLEDGMENT

The authors acknowledge the YouTube Data API v3 for enabling programmatic data collection. This research was conducted as part of academic coursework requirements and does not represent a commercial application. All data was collected in compliance with YouTube's Terms of Service and used solely for academic research purposes.

REFERENCES

1. Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.
2. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019, 4171–4186.
3. Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the 8th International AAAI Conference on Weblogs and Social Media.
4. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
5. Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: A review. Artificial Intelligence Review, 53(6), 4335–4385.
6. Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. Electronics, 9(3), 483.
7. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.
8. Smetanin, S., & Komarov, M. (2021). Sentiment analysis of product reviews in Russian using convolutional neural networks. Proceedings of the 2021 IEEE Conference on Artificial Intelligence.
9. Poria, S., Cambria, E., Hazarika, D., & Vij, P. (2016). A deeper look into sarcastic tweets using deep convolutional neural networks. Proceedings of COLING 2016.
10. Roh, Y., Heo, G., & Whang, S. E. (2019). A survey on data collection for machine learning: A big data – AI integrated perspective. IEEE Transactions on Knowledge and Data Engineering, 33(4), 1328–1347.
11. Gao, W., Peng, M., Wang, H., Zhang, Y., Xie, Q., & Tian, G. (2019). Incorporating word embeddings into topic modeling of short text. Knowledge and Information Systems, 61(2), 1123–1145.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504144|

12. Naseem, U., Razzak, I., Khushi, M., Eklund, P. W., & Kim, J. (2021). COVIDSenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis. IEEE Transactions on Computational Social Systems, 8(4), 898–908.
13. Zhu, X., & Zhang, S. (2022). Video popularity prediction by sentiment propagation via implicit networks. Proceedings of ACM SIGKDD 2022.
14. Ruder, S. (2017). An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098.
15. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of ACM KDD 2016, 785–794.
16. Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533.
17. YouTube Data API v3. (2024). YouTube Data API overview. Google Developers. https://developers.google.com/youtube/v3
18. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv preprint arXiv:1606.06565.
19. Wolf, T., Debut, L., Sanh, V., et al. (2020). Transformers: State-of-the-art natural language processing. Proceedings of EMNLP 2020 System Demonstrations, 38–45.
20. Scikit-learn developers. (2024). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
21. Streamlit Inc. (2024). Streamlit — A faster way to build and share data apps. https://streamlit.io

BIOGRAPHY

Kamalkishor Sureshkumar Sunwasiya is a postgraduate candidate in the M.Sc. Data Science programme (Semester IV) at S.I.E.S.College of Arts, Science & Commerce (Autonomous), Mumbai, India. His research interests encompass supervised machine learning, predictive analytics for marketing intelligence, and full-stack data science application development.This paper represents the principal output of his research project undertaken during the academic year 2025–26.

Dr. Abuzar Ansari is Professor and Head of the Department of Data Science at S.I.E.S. College of Arts, Science & Commerce (Autonomous), Mumbai. His research domain spans applied machine learning, educational data mining, and natural language processing. He serves as Principal Investigator for multiple funded data science research projects within the institution.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details