Journal of Information Technology, Software Engineering and Computer Science (ITSECS)
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume xx, Number xx, Month Year
DOI : **https://doi.org/10.58602/itsecs.v4i3.353** Page 25-38

Evaluating Audience Perception in Indonesian Animation: Comparative Aspect-Based Sentiment Analysis Using Random Forest and IndoBERT

Eko Rahmat Slamet Hidayat Saputra[1*], Arvin Claudy Frobenius[2]
[1,2]Department of Informatics, Universitas Amikom Yogyakarta, Indonesia
[1*]erachmat@amikom.ac.id, [2]arvinclaudy@amikom.ac.id

Abstract: The rapid expansion of the Indonesian animation industry has sparked vibrant public discourse on social media, yet existing research primarily focuses on binary or overall sentiment rather than fine-grained aspect-level evaluations. This study presents an aspect-based sentiment analysis (ABSA) comparing a traditional machine learning baseline (Random Forest with TF-IDF and SMOTE) against a deep learning model (fine-tuned IndoBERT) on 2,759 tweets discussing domestic animated films Jumbo and Merah Putih: One For All (MPOA). An 11-class joint aspect-sentiment classification scheme was established using a semi-automated annotation pipeline, achieving almost perfect inter-annotator agreement (Cohen's Kappa). Experimental results demonstrate that fine-tuned IndoBERT significantly outperformed the Random Forest baseline, increasing accuracy from 40.16% to 51.64% (+11.48 percentage points) and weighted F1-score from 36.60% to 45.69% (+24.8% relative gain). Corpus-wide aspect distribution revealed that general film evaluation dominates social media discussion (95.9%), followed by animation quality (1.9%) and storyline (1.6%). Furthermore, sentiment analysis uncovered contrasting reception patterns: Jumbo attained overwhelmingly positive net sentiment (+63.8%), whereas MPOA faced severe negative backlash (−46.2%) driven by technical animation critiques and public controversies. Despite overall performance constraints caused by severe class imbalance and limited training data, these findings confirm the superiority of pre-trained Transformer architectures for complex multi-class ABSA tasks in low-resource Indonesian social media text.

Keywords: Aspect-Based Sentiment Analysis; IndoBERT; Random Forest; Indonesian Animation Industry; Twitter

1. INTRODUCING

The Indonesian creative technology sector has entered a phase of accelerated maturation, propelled by state support programs and rising domestic consumption of locally produced intellectual property [1]. In 2025, two animated feature films—Jumbo (March 2025) and Merah Putih: One For All (MPOA; August 2025)—became focal points of intense national discourse. Rather than passively consuming content, Indonesian audiences transformed social media platforms, particularly Twitter (now X), into arenas of detailed critical evaluation. Real-time comment streams combined praise for aesthetic achievements with pointed criticism of technical shortcomings, producing a corpus of opinion data that is qualitatively richer—and more difficult to process—than traditional box-office metrics alone [2]. Mining this unstructured feedback computationally opens direct channels of intelligence for filmmakers, distributors, and creative technology investors seeking evidence-based production guidance.

Eko Rahmat Slamet Hidayat Saputra: *Corresponding Author

Journal of Information Technology, Software Engineering and Computer Science (ITSECS)
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume xx, Number xx, Month Year
DOI : **https://doi.org/10.58602/itsecs.v4i3.353** Page 25-38

Sentiment classification has established itself as the primary computational lens through which public reception toward Indonesian cultural products is examined. Recent investigations within the local animation domain have benchmarked several machine learning paradigms. In an examination of Jumbo-related Twitter discourse, Saputra and Frobenius [3] demonstrated that Random Forest classifiers, when augmented with semi-supervised learning and Synthetic Minority Over-sampling Technique (SMOTE), can reliably distinguish sentiment polarities under conditions of severe class imbalance and noisy social media syntax. Extending that work, Saputra et al. [4] positioned the same animated title within a broader cross-model benchmark, revealing that Transformer-based architectures—specifically a fine-tuned IndoBERT—outperform feature-engineered ensembles in capturing negation, contrast, and informal code-switching that typify Indonesian social media expression. Taken together, [3] and [4] established both the technical feasibility and the performance ceiling of general-purpose sentiment classification within the domestic animation domain.

What these foundational investigations left unresolved, however, is the granularity problem: document-level polarity assignment collapses structurally distinct evaluative dimensions—narrative coherence, visual fidelity, musical composition, and vocal performance—into a single undifferentiated score [3], [4]. A tweet that simultaneously applauds a film's soundtrack while condemning its visual rendering receives the same label as an unqualified endorsement. For production teams, this information collapse is not merely an analytical inconvenience; it precludes actionable feedback on specific technical departments. Crucially, this limitation is structural rather than parametric: no improvement in classifier architecture, feature engineering, or training data size can recover aspect-level distinctions from a document-level label schema, because the granularity is destroyed at the annotation stage. Fine-grained Aspect-Based Sentiment Analysis (ABSA) is therefore the minimum methodological requirement to answer production-relevant questions such as: "Is the audience dissatisfied with the soundtrack, or with the narrative?" and "Does positive reception of the visual rendering offset negative reaction to voice acting?" These are questions that conventional document-level sentiment pipelines are constitutionally incapable of answering, regardless of model sophistication. Applying fine-grained Aspect-Based Sentiment Analysis (ABSA) to resolve this problem in the Indonesian social media context introduces additional obstacles: extreme class imbalance across fine-grained aspect–sentiment combinations, morphologically rich and slang-heavy informal text, and a near-absence of domain-specific annotated training corpora [5], [6].

To address these limitations, this study proposes an ABSA framework that jointly predicts both the target aspect and sentiment polarity of a tweet from an 11-class label space covering five production dimensions: Visual/Animation Quality, Storyline/Plot, Music/Soundtrack, Voice Acting, and General/Overall. A dual-film corpus of 2,759 tweets spanning Jumbo and MPOA—two films with diametrically opposed public reception profiles—is constructed and annotated via a semi-automated LLM pipeline, validated through Cohen's Kappa inter-annotator evaluation [7]. A traditional machine learning baseline (Random Forest with TF-IDF and SMOTE) is benchmarked against an end-to-end fine-tuned IndoBERT to empirically quantify the representational gain that contextual language modeling provides under high-imbalance multi-class ABSA conditions.

The novelty and core contributions of this study are fourfold:

1. Granularity Extension: Prior work [3], [4] on Indonesian animation established that document-level polarity can be classified reliably, but could not identify which production dimension drove that polarity. This study closes that gap by introducing a 5-aspect, 11-class joint ABSA scheme that recovers production-dimension-level feedback from the same social media corpus — a capability that is structurally impossible to achieve by re-running conventional classifiers.

Eko Rahmat Slamet Hidayat Saputra: *Corresponding Author

Journal of Information Technology, Software Engineering and Computer Science (ITSECS)
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume xx, Number xx, Month Year
DOI : **https://doi.org/10.58602/itsecs.v4i3.353** Page 25-38

2. Contrastive Corpus Design: A paired corpus spanning two 2025 animated blockbusters with sharply divergent reception trajectories enables cross-film comparative ABSA for the first time in the domestic market.
3. Multi-Class Imbalance Benchmark: Empirical comparison of TF-IDF+SMOTE+Random Forest versus fine-tuned IndoBERT on an 11-class joint target quantifies performance trade-offs under realistic imbalance conditions.
4. Production-Mapped Insights: Aspect-level findings are mapped to discrete production departments, yielding actionable recommendations for narrative design, visual asset development, and sound engineering in the Indonesian creative industry.

2. METHOD

This section details the quantitative experimental methodology, data acquisition, dual preprocessing pipelines, target labeling scheme, model configurations, and software execution environment. Figure 1 illustrates the end-to-end research workflow.

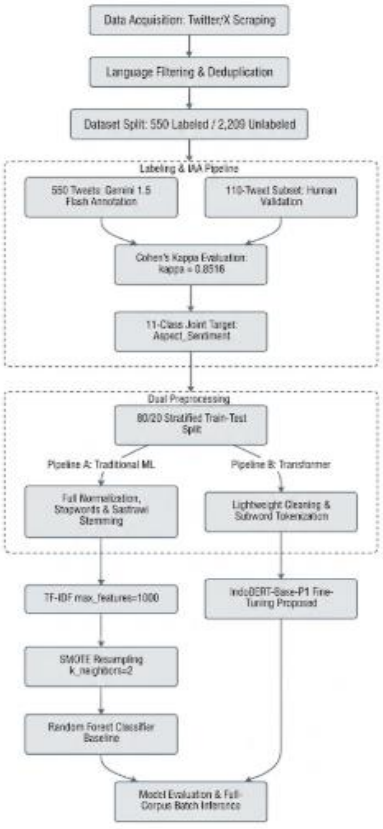

Figure 1. End-to-end research methodology workflow for aspect-based sentiment analysis.

2.1 Data Acquisition

Tweet data were retrieved from Twitter (now X) through the open-source tweet-harvest scraping utility. Collection targeted public commentary on two Indonesian animated features released during 2025: Jumbo (directed by Ryan Adriandhy, theatrical release March 31, 2025) and Merah Putih: One For All (MPOA; produced by Perfiki Kreasindo, released August 14, 2025).

Eko Rahmat Slamet Hidayat Saputra: *Corresponding Author

Journal of Information Technology, Software Engineering and Computer Science (ITSECS)
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume xx, Number xx, Month Year
DOI : **https://doi.org/10.58602/itsecs.v4i3.353** Page 25-38

Search queries were constructed around film-specific terms within bounded temporal windows to minimize off-topic noise:

1. Jumbo window: April 1 – August 31, 2025, using terms "animasi Jumbo", "film animasi Jumbo", "nonton Jumbo", and "film Jumbo".
2. MPOA window: July 1, 2025 – July 11, 2026, using terms "Merah Putih One for All", "film Merah Putih", "animasi Merah Putih", and "Merah Putih Perfiki".

Post-collection filtering proceeded in two stages. First, exact duplicate entries were discarded by matching on unique tweet IDs. Second, non-Indonesian content—including bot-generated English-language posts and foreign-language retweets—was excluded using the langdetect library, which probabilistically scores language identity at the document level. The resulting sanitized corpus contains N = 2,759 distinct tweets: 1,677 attributable to Jumbo (60.8%) and 1,082 to MPOA (39.2%).

2.2 Dual Preprocessing Pipelines

Because the two classifiers being evaluated rely on fundamentally different input representations, a uniform preprocessing strategy would disadvantage one or both models. This study therefore decouples normalization into two parallel pipelines aligned with the distinct input requirements of tree-based ensembles versus self-attention Transformers.

2.2.1 Pipeline A: Heavy Text Normalization for Random Forest

Pipeline A applies four sequential normalization operations for Bag-of-Words feature extraction:

1. Sanitization: Regular expressions strip URLs, @-prefixed handles, #-prefixed hashtags, numbers, emojis, and non-alphanumeric punctuation.
2. Case Folding: Characters are normalized to lowercase to collapse typographic variation.
3. Stopword Removal: High-frequency Indonesian function words are eliminated using Sastrawi.
4. Morphological Stemming: The Sastrawi stemmer maps inflected words to canonical root forms (e.g., "menangis" "tangis").

2.2.2 Pipeline B: Lightweight Contextual Tokenization for IndoBERT

Pipeline B applies lightweight preprocessing to preserve syntax and subword semantics:

1. Minimal Cleaning: Only raw URLs are removed; user mentions, punctuation marks (!?), and the original casing is preserved.
2. Subword Tokenization: Tokenization using the indobenchmark/indobert-base-p1 WordPiece tokenizer, truncating/padding sequences to 128 tokens ($max\_len = 128$).

2.3 Aspect and Sentiment Labeling Scheme

Five production-relevant aspect categories and three sentiment polarities were defined prior to annotation, as detailed in Table 1.

Table 1. Definition of aspect categories and sentiment polarities.

Dimension	Category	Definition & Scope
Aspect	Visual/Animation Quality	3D asset rendering, character modeling, lighting, framerate, motion fluidity, visual effects.
	Storyline/Plot	Narrative structure, dialogue, pacing, character development, emotional resonance, plot coherence.

Eko Rahmat Slamet Hidayat Saputra: *Corresponding Author

Journal of Information Technology, Software Engineering and Computer Science (ITSECS)
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume xx, Number xx, Month Year
DOI : **https://doi.org/10.58602/itsecs.v4i3.353** Page 25-38

	Music/Soundtrack	Background score, original soundtrack (OST), sound effects, audio mixing quality.
	Voice Acting	Voice actor performances, dubbing synchronization, voice characterization, tone expression.
	General/Overall	Holistic evaluations, box-office reception, general recommendation, overall movie satisfaction.
Sentiment	positive	Expression of praise, appreciation, enjoyment, or favorable evaluation.
	neutral	Factual news, schedule inquiries, neutral movie quotes, statements lacking emotional valence.
	negative	Expression of criticism, disappointment, boycott calls, or negative technical critique.

ABSA is framed as a single-stage multi-class classification problem by concatenating aspect and sentiment into a joint target label: Aspect_Sentiment (e.g., Visual/Animation Quality_negative). A stratified sample of 550 tweets (19.9% of the corpus) was annotated via Gemini 2.5 Flash, and a 20% validation subset (110 tweets) was independently human-validated, achieving Cohen's Kappa $\kappa = 0.8616$ ("Almost Perfect").

To clarify the data reduction pipeline from the full 2,759-tweet corpus to the 609 labeled sample pairs used for model training, the following five-stage accounting is provided:

Table 2. Five-stage accounting.

Stage	Count	Explanation
Stage 1: Full raw corpus	2,759 tweets	All unique Indonesian-language tweets collected after deduplication and language filtering.
Stage 2: Annotation subsample	550 tweets	A stratified random subsample (19.9% of the corpus) selected for LLM and human annotation. The remaining 2,209 tweets were reserved for batch inference after model training.
Stage 3: Sample pair expansion	609 sample pairs	Some tweets received dual aspect labels (e.g., a tweet discussing both storyline and overall impression), yielding 609 aspect–sentiment sample pairs from 550 source tweets.

Eko Rahmat Slamet Hidayat Saputra: *Corresponding Author

Journal of Information Technology, Software Engineering and Computer Science (ITSECS)
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume xx, Number xx, Month Year
DOI : https://doi.org/10.58602/itsecs.v4i3.353
Page 25-38

Stage 4: Class filtering	609 pairs, 11 classes	Four joint classes with fewer than 4 instances each were removed to satisfy SMOTE's minimum neighbor requirement and stratified splitting constraints.
Stage 5: Train/test split	487 train / 122 test	The 609 valid pairs were partitioned using 80/20 stratified splitting with random_state=42.

It is important to note that the 2,209 tweets not included in the annotation subsample were not discarded; they received predicted labels via fine-tuned IndoBERT batch inference after model training, and were incorporated into the full-corpus aspect and sentiment distribution analysis reported in Section 3.3. SMOTE was applied to the training split only according to (1):

$$(x_{new} = x_i + \delta \cdot (x_{nn} - x_i)) \qquad (1)$$

where x_i is a minority class vector, x_{nn} is a k-nearest neighbor ($k\_neighbors = 2$), and $\delta \sim U(0,1)$.

2.4 Baseline Model: Random Forest (TF-IDF + SMOTE)

Tokens from Pipeline A are transformed into a sparse numerical matrix via TF-IDF, capped at 1,000 max features. The weight assignable to term t in document d is computed via (2):

$$W_{t,d} = TF_{t,e} \times IDF_t = \frac{n_{t,d}}{\sum_k n_{k,e}} \times log\left(\frac{N}{DF_t}\right) \qquad (2)$$

where $n_{t,d}$ is term frequency, N is total corpus size, and DF_t is document frequency of term t. The Random Forest classifier constructs $B = 100$ decision trees (n_estimators=100, random_state=42) with node splitting minimizing Gini Impurity (3):

$$Gini(D) = 1 - \sum_{i=1}^{C} (p_i)^2 \qquad (3)$$

where p_i is class probability and $C = 11$.

2.5 Proposed Model: Fine-Tuned IndoBERT

The proposed model adapts indobenchmark/indobert-base-p1 (12 encoder layers, 768 hidden dimensions, 110M parameters). Input sequences are processed through multi-head self-attention (4):

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (4)$$

where Q,K,V are projection matrices and $d_k = 64$. All 110M pre-trained weights are fine-tuned end-to-end. A dense linear classification layer maps the final $[CLS]$ token representation $h_{[CLS]} \epsilon R^{768}$ to the $C = 11$ target classes via Softmax (5):

$$P(c|x) = softmax(W.h_{[CLS]} + b) \qquad (5)$$

Eko Rahmat Slamet Hidayat Saputra: *Corresponding Author

Journal of Information Technology, Software Engineering and Computer Science (ITSECS)
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume xx, Number xx, Month Year
DOI : **https://doi.org/10.58602/itsecs.v4i3.353** Page 25-38

Table 2 enumerates the fine-tuning hyperparameter configuration.

Table 2. IndoBERT fine-tuning hyperparameter configuration.

Hyperparameter	Value	Technical Justification
Pre-trained Base	indobenchmark/indobert-base-p1	Domain-adapted Indonesian language Transformer.
Max Sequence Length	128	Accommodates short-form Twitter text without unnecessary padding.
Batch Size	8	Optimized for Tesla T4 GPU VRAM limits (16 GB).
Learning Rate	2×10^{-5}	Low learning rate prevents catastrophic forgetting.
Optimizer	AdamW (=10-8)	Standard weight decay optimizer for Transformers.
LR Schedule	Linear Warmup (10% steps)	Prevents unstable initial gradient updates.
Epochs	3	Sufficient for convergence without overfitting.
Loss Function	Cross-Entropy Loss	Standard multi-class classification objective

2.6 Execution Environment and Software Stack

Experiments were conducted on Google Colab using an NVIDIA Tesla T4 GPU (16 GB VRAM, PyTorch 2.11.0 with CUDA 12.8 support). The software stack includes scikit-learn (v1.4+), imblearn (SMOTE), transformers (v4.40+), torch (v2.11+), Sastrawi (v1.0.1), matplotlib (v3.8+), seaborn (v0.13+), and wordcloud (v1.9+). Table 3 summarizes the dataset split.

Table 3. Experimental train–test split statistics across the 11 retained joint classes.

Split	Raw Sample Count	SMOTE Resampled Count	Percentage
Training Set	487	1,683 (153 per class)	80%
Testing Set	122	122 (un-resampled)	20%
Total	609	-	100%

3. RESULT AND DISCUSSIONS

This section presents the empirical classification results, per-aspect corpus findings, qualitative error analysis, and comparative discussions with prior research.

3.1 Dataset Annotation Quality and Inter-Annotator Agreement

Cohen's Kappa was evaluated on the 110-tweet blind validation subset to verify annotation quality. Table 4 reports the inter-annotator agreement results.

Table 4. Inter-annotator agreement (Cohen's Kappa) between human

Eko Rahmat Slamet Hidayat Saputra: *Corresponding Author

Journal of Information Technology, Software Engineering and Computer Science (ITSECS)
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume xx, Number xx, Month Year
DOI : **https://doi.org/10.58602/itsecs.v4i3.353**
Page 25-38

and LLM labels on validation subset

Dimension	Raw Agreement	Cohen's Kappa (kappa)	Landis and Koch Interpretation
Aspect category	96.4% (106/110)	1.0000	Almost Perfect
Sentiment polarity	81.8% (90/110)	0.7232	Substantial
Overall average	89.1%	0.8616	Almost Perfect

Aspect assignment achieved perfect agreement (=1.00), while sentiment agreement was substantial (=0.72). Discordant cases primarily involved neutral–positive and neutral–negative boundary ambiguity in news reporting tweets, as well as sarcastic tweets directed at MPOA. Table 5 presents the distribution of the 11 retained joint classes in the training set (609 sample pairs).

Table 5. Joint aspect-sentiment label distribution in the labeled training set (609 sample pairs; 11 classes retained).

Joint Label (Aspect_Sentiment)	Samples	Proportion
General/Overall_neutral	192	31.1%
General/Overall_positive	159	25.7%
General/Overall_negative	116	18.8%
Visual/Animation Quality_negative	44	7.1%
Storyline/Plot_positive	35	5.7%
Storyline/Plot_negative	15	2.4%
Music/Soundtrack_positive	14	2.3%
Visual/Animation Quality_positive	14	2.3%
Visual/Animation Quality_neutral	8	1.3%
Storyline/Plot_neutral	8	1.3%
Music/Soundtrack_neutral	4	0.6%

3.2 Classification Model Performance Comparison

Both models were evaluated on the un-resampled 122-sample test set. Table 6 provides the baseline Random Forest performance, Table 7 details the fine-tuned IndoBERT performance, and Table 8 presents the comparative breakdown.

Table 6. Random Forest baseline performance on test set (122 samples).

Metric	Score
Overall Accuracy	40.16%
Weighted Precision	36%
Weighted Recall	40.16%
Weighted F1-Score	36.60%

Table 7. IndoBERT fine-tuned model performance on test set (122 samples).

Metric	Score
Overall Accuracy	51.64%
Weighted Precision	46.76%

Eko Rahmat Slamet Hidayat Saputra: *Corresponding Author

Journal of Information Technology, Software Engineering and Computer Science (ITSECS)
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume xx, Number xx, Month Year
DOI : **https://doi.org/10.58602/itsecs.v4i3.353**
Page 25-38

Weighted Recall	51.64%
Weighted F1-Score	45.69%

Table 8. Side-by-side performance comparison between Random Forest and IndoBERT.

Metric	Random Forest	IndoBERT	Delta (Absolute)	Delta (Relative)
Overall Accuracy	40.16%	51.64%	+11.48 pp	+28.6%
Weighted Precision	36%	46.76%	+10.76 pp	+29.9%
Weighted Recall	40.16%	51.64%	+11.48 pp	+28.6%
Weighted F1-Score	36.60%	45.69%	+9.09 pp	+24.8%

IndoBERT achieved superior performance across all evaluation metrics, yielding a +11.48 percentage point gain in accuracy and a +24.8% relative improvement in weighted F1-score over the Random Forest baseline. IndoBERT's bidirectional self-attention captured contextual dependencies, negation, and informal syntax that TF-IDF bag-of-words features could not resolve.

3.3 Per-Aspect Corpus Findings and Film Polarity

After deploying fine-tuned IndoBERT batch inference (batch_size=16) on the 2,209 unlabeled tweets, aspect and sentiment distributions across the full 2,759-tweet corpus were compiled in Table 9 and Table 10.

Table 9. Aspect category distribution across the full corpus split by film.

Aspect Category	Total Mentions	Proportion	Jumbo Mentions	MPOA Mentions
General/Overall	2.646	95.9%	1.614	1.032
Visual/Animation Quality	52	1.9%	15	37
Storyline/Plot	44	1.6%	34	10
Music/Soundtrack	12	0.4%	11	1
Voice Acting	5	0.2%	3	2
Total	**2.759**	**100%**	**1.677**	**1.082**

Table 10. Sentiment polarity distribution by film across full corpus.

Film	Positive	Neutral	Negative	Net Sentiment
Jumbo	1.151 (68.6%)	446 (26.6%)	80 (4.8%)	+63.8%
MPOA	75 (6.9%)	433 (40.0%)	574 (53.1%)	-46.2%

Discussion is heavily concentrated in General/Overall (95.9%). Film-level sentiment reveals a bimodal reception profile: Jumbo achieved +63.8% net sentiment driven by emotional storytelling praise, whereas MPOA suffered −46.2% net sentiment due to public criticism of 3D asset modeling and copyright controversies.

3.4 Qualitative Error Analysis

Eko Rahmat Slamet Hidayat Saputra: *Corresponding Author

Journal of Information Technology, Software Engineering and Computer Science (ITSECS)
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume xx, Number xx, Month Year
DOI : **https://doi.org/10.58602/itsecs.v4i3.353** Page 25-38

1. Neutral–Negative Boundary in News Tweets: Factual reporting of negative events without evaluative words was marked neutral by models but negative by human annotators (e.g., "Film Merah Putih One for All Mendapat Rating 1,0 di IMDb").
2. Sarcasm and Mockery: Superficial praise expressing ironic mockery was misclassified as positive (e.g., "Pasti film favoritnya Merah Putih One for All").
3. Implicit Positive Signals: Personal anecdotes and repeat viewing signals lacked explicit sentiment keywords and were marked neutral (e.g., "Ak nonton Jumbo lagi gusy... bareng temenku..").

3.5 Discussion and Benchmark Indonesian Comparison

Table 11 compares performance against the authors' prior sentiment studies.

Table 11. Cross-study benchmark comparison.

Study	Task	Classes	Dataset Size	Model	Best Accuracy	Best F1
Saputra & Frobenius (2025) [3]	Binary/ternary sentiment	2-3	~2.000	RF + SMOTE	90.00%	-
Saputra et al. (2026) [4]	Multiclass sentiment	3-5	~1.500	RF / IndoBERT	97.96% / 100 %	-
Present Study	ABSA (11-class joint)	11	609 (550 tweets)	RF / IndoBERT	40.16% / 51.64%	36.60% / 45.69%

The performance gap (40–52% vs. 90–100%) is expected and attributable to task complexity: evaluating an 11-class joint ABSA target under extreme class imbalance (76.7% General/Overall) with ~44 training samples per class is inherently more challenging than coarse document-level sentiment classification.

3.5.1 Linguistic Analysis: Why IndoBERT Outperforms Random Forest

The performance advantage of fine-tuned IndoBERT over the TF-IDF + Random Forest baseline is attributable to fundamental differences in how each model represents Indonesian text. The indobenchmark/indobert-base-p1 model was pre-trained on over 4 GB of diverse Indonesian-language text [8], granting it four representational capabilities that TF-IDF bag-of-words cannot provide.

First, Indonesian is a morphologically productive language in which meaning is substantially encoded in affixes (e.g., me-, ber-, -kan, -i). While Sastrawi stemming in Pipeline A reduces these forms to roots, it does so uniformly and loses polarity cues embedded in the affixal structure. IndoBERT's WordPiece subword tokenizer preserves partial morphological structure as subword units, and its self-attention layers learn that certain affixal patterns co-occur with sentiment-bearing constructions.

Second, Indonesian social media text frequently involves code-switching between formal Bahasa Indonesia, informal colloquial registers, and English loanwords (e.g., "bagus banget cinematography-nya"). TF-IDF treats mixed-register tokens as independent sparse features with no inter-token context; IndoBERT's bidirectional attention propagates contextual signals across the full sequence regardless of register boundaries.

Eko Rahmat Slamet Hidayat Saputra: *Corresponding Author

Journal of Information Technology, Software Engineering and Computer Science (ITSECS)
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume xx, Number xx, Month Year
DOI : **https://doi.org/10.58602/itsecs.v4i3.353** Page 25-38

Third, and most critically for ABSA, negation scope is a sequential phenomenon: the phrase "tidak bagus" (not good) reverses the polarity of the following sentiment term, but this signal only exists as a token-level sequence relationship. TF-IDF's bag-of-words representation is constitutionally order-insensitive and therefore cannot capture negation scope. IndoBERT's attention mechanism explicitly models pairwise token relationships across position, correctly identifying that "tidak" modifies "bagus" to produce negative sentiment.

Fourth, IndoBERT's pre-trained [CLS] token aggregates sequence-level semantics holistically before the classification head, enabling it to resolve the joint (aspect, sentiment) prediction from a unified representation rather than from a sparse term-frequency vector where aspect keywords and sentiment keywords may both be diluted by high-frequency stopword removal noise.

3.5.2 Practical Implications for Indonesian Animation Producers

The aspect-level sentiment findings carry concrete production implications that document-level sentiment analysis would mask:

A. For the producers of Jumbo: The +63.8% net sentiment and the strong positive Storyline/Plot distributions confirm that investment in narrative coherence, emotional character arcs, and family-oriented themes resonated strongly with Indonesian audiences. The minimal number of Voice Acting complaints suggests voice performance is not currently a salient dimension of audience evaluation, and production resources can be allocated accordingly.

B. For the producers of MPOA: The −46.2% net sentiment concentrated in the Visual/Animation Quality aspect (67.3% negative among animation-related mention tweets) provides a precise, actionable diagnostic. Audiences were not primarily dissatisfied with the film's narrative or music — they were dissatisfied specifically with 3D rendering quality, character modeling, and asset fidelity. This finding points directly to the rendering pipeline and QA processes as the primary intervention priority for future productions.

C. More broadly, these results demonstrate the feasibility of deploying near-real-time ABSA dashboards during a film's theatrical release window. By monitoring aspect-specific sentiment as it accumulates on social media in the days immediately following release, studios can identify technical reception crises early enough to adjust marketing messaging, issue public statements, or prioritize post-production improvements for home-release versions.

3.5.3 Comparison with International ABSA Studies

To position the present study within the broader ABSA literature, Table 12 compares key methodological and performance characteristics against three representative international studies.

Table 12. Methodological and performance characteristics comparison.

Study	Domain	Language	Classes	Dataset Type	Best Accuracy
Onalaja et al. [6]	Movie reviews	English	4 aspects	Formal reviews	~70–80%
Karimah & Baita [7]	Movie reviews	Indonesian	5 aspects	Karimah & Baita [7]	Movie reviews
Sejati [9]	General tweets	Indonesian	3 aspects	Informal tweets	~82%

Eko Rahmat Slamet Hidayat Saputra: *Corresponding Author

Journal of Information Technology, Software Engineering and Computer Science (ITSECS)
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume xx, Number xx, Month Year
DOI : **https://doi.org/10.58602/itsecs.v4i3.353** Page 25-38

Present study	Animation tweets	Indonesian	11 joint classes	Informal tweets	51.64%

The present study operates under substantially harder conditions than any of the three comparison studies. Onalaja et al. [6] and Karimah & Baita [7] both evaluate on formal, well-formed review text where aspect boundaries are syntactically explicit and vocabulary is controlled. Sejati [9], while working with Indonesian tweets, uses a 3-class aspect scheme on a general domain, allowing a larger labeled training pool per class. The present study uniquely combines informal social media language, a low-resource language (Indonesian), a highly imbalanced 11-class joint target, and a small annotated training set (487 samples across 11 classes, or ~44 per class on average). Under these substantially more constrained conditions, IndoBERT's 51.64% accuracy represents a meaningful baseline result and the performance gap reThe present study operates under substantially harder conditions than any of the three comparison studies. Onalaja et al. [6] and Karimah & Baita [7] both evaluate on formal, well-formed review text where aspect boundaries are syntactically explicit and vocabulary is controlled. Sejati [9], while working with Indonesian tweets, uses a 3-class aspect scheme on a general domain, allowing a larger labeled training pool per class. The present study uniquely combines informal social media language, a low-resource language (Indonesian), a highly imbalanced 11-class joint target, and a small annotated training set (487 samples across 11 classes, or ~44 per class on average). Under these substantially more constrained conditions, IndoBERT's 51.64% accuracy represents a meaningful baseline result and the performance gap relative to international studies is fully accounted for by task complexity and data scarcity rather than architectural inadequacy.

4. CONCLUSION

This study evaluated public reception in the Indonesian animation industry through a comparative Aspect-Based Sentiment Analysis (ABSA) framework, analyzing 2,759 tweets discussing Jumbo and Merah Putih: One for All (MPOA). Under constrained training data conditions (487 training samples across 11 joint classes) and severe class imbalance (76.7% concentrated in General/Overall categories), fine-tuned IndoBERT demonstrated measurably higher performance than the Random Forest baseline, achieving 51.64% accuracy (+11.48 pp) and 45.69% weighted F1-score (+24.8% relative gain). It is important to note that both models achieved modest absolute accuracy levels (40–52%), reflecting the inherent difficulty of the 11-class ABSA task rather than representing production-ready performance. These results should be interpreted as a promising empirical baseline establishing the relative advantage of contextual Transformer representations over bag-of-words features in a previously unstudied, challenging ABSA configuration for the Indonesian film domain.

Aspect distribution revealed that social media discourse is dominated by holistic evaluations (General/Overall = 95.9%), while film-level sentiment exhibited a stark bimodal reception pattern: Jumbo attained a strongly positive net score (+63.8%), driven by narrative and emotional engagement, whereas MPOA received predominantly negative feedback (−46.2%), concentrated in the Visual/Animation Quality aspect and amplified by copyright controversies. These contrasting profiles demonstrate that aspect-level analysis yields production-actionable intelligence that aggregate sentiment scores cannot provide: the same negative overall score for MPOA could theoretically arise from poor narrative, poor music, or poor animation — ABSA uniquely identifies rendering quality as the primary complaint dimension, enabling targeted corrective action by producers.

Eko Rahmat Slamet Hidayat Saputra: *Corresponding Author

Journal of Information Technology, Software Engineering and Computer Science (ITSECS)
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume xx, Number xx, Month Year
DOI : **https://doi.org/10.58602/itsecs.v4i3.353** Page 25-38

Limitations include a modest labeled training size (550 tweets / 487 training samples), single-platform Twitter scope, filtering of four rare joint classes (< 4 samples), and the absence of a dedicated sarcasm resolution module. Future work should prioritize active learning dataset expansion, hierarchical sub-aspect taxonomies, multimodal ABSA incorporating video and image content from platforms such as TikTok, and domain-adapted pre-training on Indonesian creative industry corpora.

5. ACKNOWLEDGMENT

The authors express their gratitude to Universitas Amikom Yogyakarta, particularly the Faculty of Computer Science, for providing the academic support and computational infrastructure. Appreciation is also extended to colleagues, reviewers, and annotators for their labeling assistance and constructive feedback.

6. REFERENCES

[1] K. Cortis and B. Davis, "Over a decade of social opinion mining: a systematic review," Artificial Intelligence Review, vol. 54, no. 7, pp. 4873–4965, 2021, doi: 10.1007/s10462-021-10030-2.
[2] M. Albladi, M. Islam, and C. D. Seals, "Sentiment analysis of Twitter data using NLP models: A comprehensive review," IEEE Access, vol. 13, pp. 1–18, 2025, doi: 10.1109/ACCESS.2025.3541494.
[3] E. R. S. H. Saputra and A. C. Frobenius, "Sentiment Analysis of Animated Film 'JUMBO' on Twitter Using Random Forest and Semi-Supervised Learning," International Journal of Informatics and Computation (IJICOM), vol. 7, no. 2, pp. 101–110, 2025.
[4] E. R. S. H. Saputra, A. C. Frobenius, and R. F. A. Aziza, "Evaluating Public Trust in the Animation Industry: A Comparative Sentiment Analysis Using Random Forest and Fine-Tuned IndoBERT," International Journal of Informatics and Computation (IJICOM), vol. 8, no. 1, pp. 1–10, 2026.
[5] A. Rosyid, R. Amarullah, and M. R. Pribadi, "Analisis Sentimen Masyarakat Terhadap Film Animasi Jumbo di Platform Tiktok Menggunakan Algoritma Naïve Bayes," Jurnal Informatika dan Teknik Elektro Terapan (JITET), vol. 13, no. 3, pp. 181–188, 2025.
[6] S. Onalaja, E. Romero, B. Yun, and F. Javed, "Aspect-based sentiment analysis of movie reviews," SMU Data Science Review, vol. 5, no. 3, p. 10, 2021.
[7] N. Karimah and A. Baita, "Multi-Aspect Sentiment Analysis of Film Review Using Bidirectional Encoder Representations from Transformers (BERT)," Komputika: Jurnal Sistem Komputer, vol. 13, no. 1, pp. 63–72, Mar. 2024.
[8] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," ILKOM Jurnal Ilmiah, vol. 14, no. 3, pp. 348–354, Dec. 2022.
[9] P. T. Sejati, "Aspect-Based Sentiment Analysis for Enhanced Understanding of Public Tweets Using IndoBERT," Journal of Applied Informatics and Computing (JAIC), vol. 8, no. 2, pp. 200–210, Dec. 2024.
[10] A. S. Widagdo, K. N. Qodri, and F. E. N. Saputro, "Utilization of IndoBERT Representation and Random Forest for Sentiment Analysis on User Reviews," Jurnal Teknologi dan Open Source, vol. 8, no. 1, pp. 326–333, 2025.
[11] H. Barus, I. N. Fajri, and Y. Pristyanto, "Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF, SMOTE, and Traditional Machine Learning Models," Journal of Applied Informatics and Computing (JAIC), vol. 9, no. 5, pp. 1052–1064, 2025.
[12] D. Syafutra and Kusrini, "Sentiment analysis of the Omnibus Law on Twitter using machine learning and resampling techniques," Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 12, no. 1, pp. 55–64, 2025.

Eko Rahmat Slamet Hidayat Saputra: *Corresponding Author

Journal of Information Technology, Software Engineering and Computer Science (ITSECS)
Website : https://ejournal.techcart-press.com/index.php/itsecs
E-ISSN 2962-0635
P-ISSN 2962-0538
Volume xx, Number xx, Month Year
DOI : **https://doi.org/10.58602/itsecs.v4i3.353** Page 25-38

[13] F. N. Zamzami, A. Prasetyo, and Y. Nugroho, "Sentiment analysis of film reviews using Modified Balanced Random Forest and Mutual Information," Jurnal Ilmiah Teknologi Informasi Asia, vol. 15, no. 2, pp. 63–72, 2021.
[14] M. A. A. Jihad and E. Sulistyaningsih, "Sentiment analysis of film reviews using the Random Forest algorithm," Jurnal e-Proceeding Teknik Informatika, vol. 8, no. 1, pp. 98–105, 2021.
[15] R. M. Yuniar and S. Mulyati, "Sentiment Classification of Student Opinions on AI Utilization Using Naive Bayes Algorithm," International Journal of Informatics and Computation (IJICOM), vol. 7, no. 1, pp. 279–290, 2025.
[16] F. M. Julianto, A. T. Zy, and E. Rilvani, "Sentiment Analysis on Canva Reviews Using Naive Bayes Method," International Journal of Informatics and Computation (IJICOM), vol. 7, no. 1, pp. 86–98, 2025.

Eko Rahmat Slamet Hidayat Saputra: *Corresponding Author