STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **XX**, Month 2025, pp 1–21.
Published online in International Academic Press (www.IAPress.org)

Beyond Engagement: Classifying News Impact on X Using Impressions Data and Hybrid Feature Architecture

Muhammad Rizky Hidayat [1,*], Derwin Suhartono [2]

[1] *Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia*
[2] *Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia*

Abstract Predicting news reach on X (formerly Twitter) is a critical challenge for digital journalism. However, existing literature often relies on biased "proxy metrics" (likes, retweets) that miss the "silent majority" of passive viewers. This study presents an empirically validated multimodal framework for objective news impact classification using Impressions (View Count) as ground truth. We construct a time-stratified dataset from @detikcom on X (2022–2025), labeled into "Regular," "Hot," and "Viral" classes via a data-driven log-base-10 thresholding strategy. A six-scenario ablation study evaluates our proposed Adaptive Gated Cross-Modal Fusion architecture against static baselines, contextual extractors, and standard MLP fusion variants. This architecture learns to dynamically balance textual and engagement signals via a trainable gating mechanism. The proposed model (S6) achieves a Macro F1-Score of 0.7383 (peak Viral-class F1 of 0.6277), representing an 85.8% improvement over the unimodal text-only baseline. A dual-track IndoBERT vs. mBERT comparison confirms the necessity of monolingual pre-training. Furthermore, forensic error analysis demonstrates how "semantic awareness" resolves Cold Start underestimation and False Viral overestimation that consistently cause tree-based classifiers to fail.

Keywords News Impact Classification, Impressions Metric, Indonesian NLP, Gated Fusion, IndoBERT

AMS 2010 subject classifications 68T50, 62H30

DOI: 10.19139/soic-2310-5070-3404

1. Introduction

The digital journalism landscape has shifted from a broadcast-centric model to a decentralized, algorithm-driven ecosystem. Social media platforms, particularly X (formerly Twitter), have become primary arenas for public discourse and audience acquisition [26, 27]. For national news organizations in Indonesia, optimizing content visibility is no longer just an editorial metric but a survival imperative for sustaining ad-revenue models [7]. Consequently, accurately predicting whether a news piece will stagnate ("Regular") or achieve massive reach ("Viral") is a critical decision-support capability for modern newsrooms.

Despite this urgency, two fundamental limitations persist in existing literature. First, studies predominantly utilize "proxy metrics" such as Likes, Comments, or Retweets as the ground truth for virality [7]. A controversial news item may generate high visibility (millions of impressions) but low engagement (few likes), creating a "blind spot" in predictive models. Empirically, we find that engagement metrics show only moderate Pearson correlation with the Impression class (Retweets: $r = 0.33$, Likes: $r = 0.30$, Comments: $r = 0.37$; see Section 4.1). This confirms that engagement proxies explain less than 14% of visibility variance. To address this, our research leverages the Impressions (View Count) metric introduced by X in late 2022 to provide a precise, objective measure of content reach [29, 30].

*Correspondence to: Muhammad Rizky Hidayat (Email: muhammad.hidayat016@binus.ac.id). Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

Second, most state-of-the-art models are optimized for high-resource languages like English [5, 8]. Applying generic multilingual models to Indonesian yields suboptimal performance due to the language's morphological richness and contextual nuances [1, 2]. This study employs IndoBERT [6, 21], a pre-trained model on massive Indonesian corpora. We argue that a monolingual model is essential for capturing the subtle semantic cues that drive virality in Indonesia's digital ecosystem [23].

We further posit that virality is a non-linear interaction between text and numerical history rather than a function of either alone [16]. We propose an **Adaptive Gated Cross-Modal Fusion** architecture that learns how much to weight textual versus numeric evidence for each specific sample, which is more effective than treating all inputs equally through simple concatenation. To validate this, we conducted a comprehensive six-scenario ablation study comparing static baselines (Word2Vec + ML), contextual extractors (IndoBERT + ML), unimodal DL, Shallow Fusion, standard Deep MLP Fusion, and the proposed Gated Fusion model. While the individual components draw on established multimodal learning paradigms [4, 14], their systematic evaluation for Indonesian news impact classification using Impressions as ground truth constitutes a novel empirical contribution.

Specifically, the contributions of this paper are threefold. First, we introduce a novel dataset and problem framing through a time-stratified dataset of Indonesian news tweets (2022–2025) labeled using the empirical Impressions distribution. This shifts the paradigm from *engagement prediction* to objective *impact classification*, with the dataset made publicly available at https://doi.org/10.5281/zenodo.17536970. Second, through a systematic six-scenario ablation study, we demonstrate that the Adaptive Gated Cross-Modal Fusion architecture achieves the best Macro F1 (0.7383) and Viral-class F1 (0.6277) among all configurations, while a dual-track IndoBERT vs. mBERT comparison confirms the advantage of monolingual pre-training. Third, we conduct a forensic error analysis with semantic interpretation. This case-level analysis demonstrates the Deep Fusion model's "semantic awareness," resolving Cold Start underestimation and False Viral overestimation that consistently cause tree-based classifiers to fail. We also highlight specific linguistic patterns where IndoBERT's monolingual training yields measurable advantages over mBERT.

2. Related Works

Research on digital content popularity has matured from simple statistical correlations to complex multimodal systems. We review three dimensions: (A) Semantic Representation, (B) Multimodal Fusion, and (C) Target Variable definitions.

2.1. Semantic Representation: From Static to Contextual Embeddings

Early approaches relied on Bag-of-Words (BoW) or TF-IDF representations with linear classifiers. These methods lacked semantic depth and suffered from the "curse of dimensionality" [9, 25, 26]. While static word embeddings like Word2Vec and GloVe improved syntactic similarity capture [17], they remain context-agnostic—a critical limitation in short-text classification [18, 19]. Consequently, the current state-of-the-art has shifted to Transformer-based models like BERT [5]. For low-resource languages, monolingual models consistently outperform multilingual counterparts. Studies confirm that IndoBERT [6,21,22,24] captures Indonesian nuances—slang, political entities, cultural references—far more effectively than mBERT [1, 2, 23].

2.2. Multimodal Fusion: From Shallow to Adaptive Integration

News virality is rarely driven by text alone. Ensemble tree-based models like XGBoost [7] and stacking techniques [15] excel at tabular data but suffer from "semantic blindness." Conversely, Shallow Fusion directly concatenates text and numeric vectors. Although used in fake news detection [10, 11] and reading quantity prediction [4], this method forces a linear interaction where low-dimensional numeric signals are often drowned out by high-dimensional text vectors [14]. Standard Deep Fusion uses an MLP head over concatenated features to capture non-linear cross-modal dependencies, yet treats all modality combinations uniformly. Our **Adaptive Gated Cross-Modal Fusion** extends this by learning a per-sample soft weighting of text versus numeric evidence, allowing the model to dynamically prioritize semantics for cold-start news and rely on engagement signals for content

with established metrics. While the gating mechanism draws on established attention paradigms, its application to Indonesian news impact classification using Impressions as ground truth represents a novel empirical contribution.

2.3. The Target Variable: Moving Beyond Proxy Metrics

Existing work frequently relies on proxy metrics such as Likes, Retweets, or Sentiment [3, 27, 28]—metrics biased toward active users that miss the "Silent Majority." The Impressions metric represents true content reach, though explored only in limited contexts [4]. Our study leverages this metric for the Indonesian X platform using a custom scraping methodology [29–31], shifting the focus from *Engagement Prediction* to objective *Impact Classification*. We acknowledge that Impressions, as a platform-estimated metric, may reflect algorithmic amplification in addition to organic reach—a limitation discussed further in Section 5.

3. Methodology

This study proposes an End-to-End Hybrid Deep Fusion Framework to classify news impact from heterogeneous modalities. The workflow (Fig. 1) covers: Data Collection, Cleaning, Labeling, Splitting, Preprocessing, Architecture Design, and Experimental Setup.

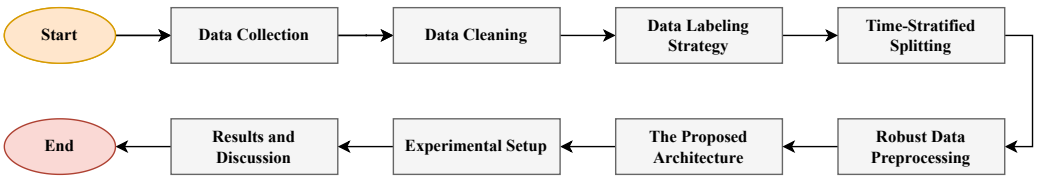

Figure 1. Research methods used in this study.

3.1. Data Collection

Data was collected from X targeting @detikcom using a custom Python scraper built with `Selenium` (headless Chromium) and `BeautifulSoup4`. The collection period spans from December 22, 2022—when X made Impressions publicly visible [29, 30]—to early 2025.

Authentication and Access. The Impressions metric is rendered only for authenticated users and is not accessible via the public API. We inject the `auth_token` session cookie via `driver.add_cookie()`, granting authenticated access without relying on X's v2 API, which does not expose this metric at the required scale.

Dynamic Content Loading. X uses JavaScript-based infinite scroll. Our scraper executes iterative `window.scrollBy()` calls via Selenium's `execute_script()`, with randomized delays of 2–4 seconds between scroll events to simulate human browsing and avoid bot-detection. Each cycle waits for DOM stabilization—confirmed via a polling loop checking element count stability—before extracting newly loaded elements. This addresses the core challenge of extracting dynamically-rendered content where naïve scraping would capture partial or stale data.

Impressions Extraction. The Impressions count is located via XPath targeting `data-testid=''app-text-transition-container''` adjacent to the views icon. This selector remained robust across multiple UI updates during 2022–2025. The metric requires the authenticated session to be active; unauthenticated requests return zero or empty values.

Handling Unavailable Content. Tweets returning HTTP 404 or an "unavailable" DOM state were detected via CSS selector checks and excluded. Tweets containing only media without textual headlines were also filtered out. Deleted or protected tweets encountered during scraping were handled gracefully via try-except blocks, with their URLs logged and excluded from the final dataset.

Data Consolidation. Per-session CSV files were merged via `pd.concat(ignore_index=True)` and deduplicated by URL, retaining the first chronological instance. The final raw dataset comprised approximately

278,700 records. For each tweet, we extracted: (1) textual content, (2) publication timestamp, (3) news URL, (4) engagement metrics (Likes, Retweets, Comments), and (5) **Impressions** as the primary ground truth.

Table 1. Snapshot of Raw Dataset Collected from Detik.com

Timestamp	Text Content (Snippet)	Lks	RTs	Cmts	Impressions
2024-01-04 23:53	*Cristiano Ronaldo mengawali 2024 dengan manis. Bintang Portugal itu langsung...*	30	1	2	11,799
2024-01-04 23:38	*Ibra Azhari kembali ditangkap pihak kepolisian terkait narkoba...*	16	4	1	16,493
2024-01-04 23:18	*Ibra Azhari kembali ditangkap pihak kepolisian terkait narkoba. Ibra...*	28	6	6	36,674
2024-01-04 22:12	*Juliansyah pemilik 1 hektar ladang ganja di tengah perbukitan...*	17	1	1	22,324
2024-01-04 20:32	*Ada dikotomi soal pemain lokal dan pemain naturalisasi di sepakbola Indonesia...*	15	2	0	20,579
2024-01-04 17:25	*Saat gempuran Israel di Jalur Gaza belum sepenuhnya reda, dua bom meledak...*	39	8	6	17,662

3.2. Data Cleaning

The raw data underwent three cleaning steps: (1) duplicate removal by URL, retaining the first chronological instance; (2) filtering tweets with missing Impressions values; and (3) domain parsing to resolve shortened URLs (t.co) and retain only content originating from the Detik.com ecosystem.

3.3. Data Labeling Strategy

News impact classes are defined quantitatively based on the empirical Impressions distribution rather than arbitrary thresholds.

Distributional Analysis. Raw Impressions follow a strong Power Law with a heavy right-skew [16]. This is confirmed by our exploratory data analysis in Section 4.1. Applying a log-base-10 transformation ($x' = \log_{10}(x)$) reveals a near-normal distribution. Two clear natural inflection points emerge at integer values $\log_{10} = 4$ and $\log_{10} = 5$. Each point represents a full order-of-magnitude jump in reach.

Threshold Derivation & Transformation. The boundaries at 10,000 and 100,000 Impressions serve as natural breakpoints in the log-transformed density curve. To handle bot activity and extreme outliers, we implemented a data distribution transformation strategy. As detailed in Table 2, records below 1,000 Impressions (740 samples) were excluded as noise. The remaining data was grouped into three final classes. This effectively merged the extreme heavy-tail distributions ($> 10^5$) into the minority "Viral" class. The upper boundary of 100,000 Impressions corresponds to the empirical 99th percentile of the distribution ($\approx 97{,}400$), providing distributional validation for this cutoff. The lower boundary of 10,000 marks the transition to the "Hot" zone. This transition is consistently identifiable as a slope change in the log-density curve (Fig. 2, right panel).

Distributional Evidence. Fig. 2 presents the Impressions distribution in two forms. The left panel shows the raw distribution truncated at 2M for readability. It reveals the canonical Power Law shape with a mean of 15,218 and a median of 5,591. This nearly three-fold gap confirms an extreme right-skew. The right panel applies a log-base-10 x-axis transformation, which compresses this tail into a near-normal bell curve. The natural density inflection points at 10^4 and 10^5 are clearly visible as slope changes. This confirms that the class boundaries are data-driven rather than arbitrary selections.

Discriminative Validity. Fig. 3 presents box plots of log-transformed engagement metrics stratified by Impression class. This visualization empirically verifies that the three classes are genuinely separable. A clear, progressive upward shift in the interquartile range (IQR) is visible across all three features as content moves from the Regular category to the Hot and Viral categories.

Table 2. Mapping Raw Power Law Bins to Final Impact Classes

Before: Raw Distribution		**After: Filtered & Grouped Classes**			
Range	**Count**	**Class**	**Log_{10}**	**Total**	**Interpretation**
$< 10^1$	2	*Filtered*	< 3.0	740	Noise/bot activity (excluded from dataset).
10^1–10^2	54				
10^2–10^3	684				
10^3–10^4	218,670	**Regular**	$[3.0, 4.0)$	218,670	Standard daily news; below ad monetization threshold (78.2%).
10^4–10^5	56,292	**Hot**	$[4.0, 5.0)$	56,292	Trending content entering premium CPM territory (20.1%).
10^5–10^6	4,362	**Viral**	≥ 5.0	4,718	Mass-reach; empirical 99th percentile $\approx 97{,}400$ validates boundary (1.7%).
$> 10^6$	356				

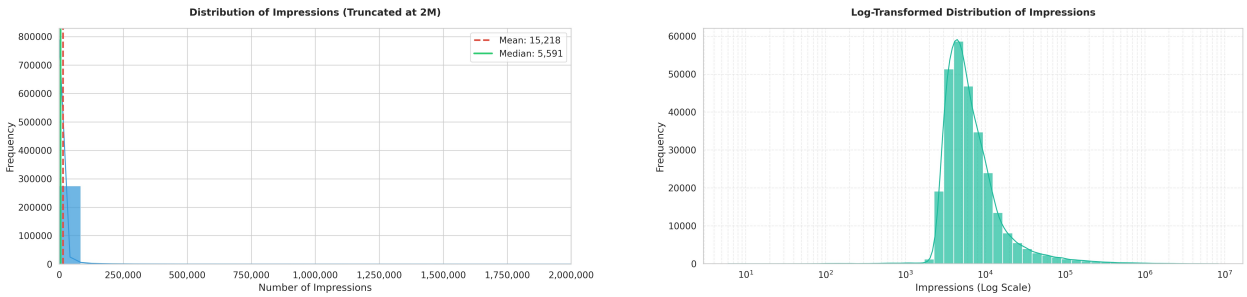

Figure 2. Impressions distribution. **Left:** Raw distribution (truncated at 2M) showing Power Law shape; Mean = 15,218, Median = 5,591. **Right:** Log-base-10 transformation revealing a near-normal bell curve with natural inflection points at 10^4 and 10^5 that define class boundaries.

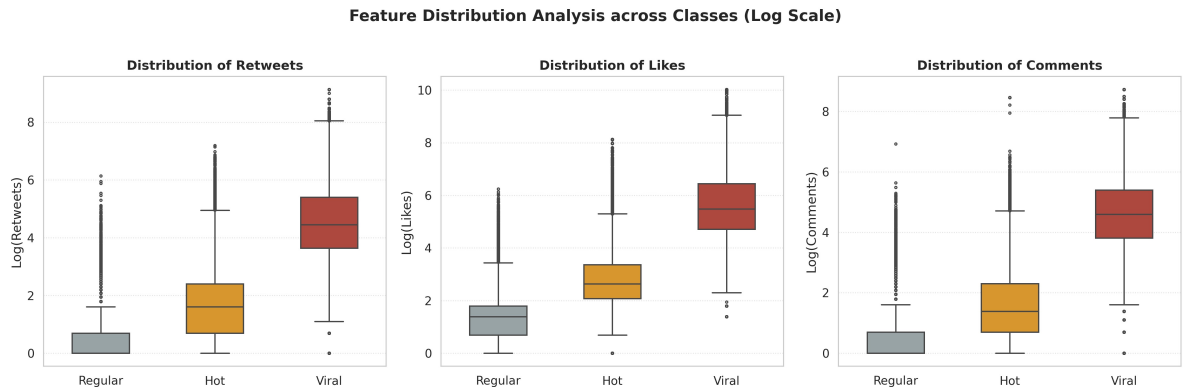

Figure 3. Feature distribution across Impression classes (log scale). All three engagement metrics exhibit a consistent, monotonic IQR increase from Regular to Hot to Viral, confirming discriminative validity of the log-based class boundaries.

3.4. Time-Stratified Splitting

To simulate a real-world forecasting scenario and prevent temporal data leakage, we implemented a time-stratified splitting strategy. First, the dataset was sorted chronologically from the oldest to the newest tweets. By doing so, the model is trained entirely on historical data and evaluated strictly on future data.

Within each impact class, the chronological sequence was split sequentially. The first 80% of the oldest records formed the training set. The subsequent 15% was allocated for validation to tune model hyperparameters. Finally, the most recent 5% of the data was isolated as the holdout test set. This approach ensures that the model cannot ”peek” into the future. It guarantees that the evaluation metrics reflect true predictive capability on unseen and newly published content.

Table 3 details the exact sample distribution across the Train, Validation, and Test sets. These figures are calculated based on the 80/15/5 ratio applied to the filtered dataset distributions.

Table 3 details the exact sample distribution across the three splits.

Table 3. Time-Stratified Data Splitting (80/15/5 ratio) per Impact Class

Class	**Total**	**Train (80%)**	**Validation (15%)**	**Test (5%)**
Regular	218,666	174,932	32,800	10,934
Hot	56,292	45,034	8,443	2,815
Viral	4,718	3,774	707	237
Total	**279,676**	**223,740**	**41,950**	**13,986**

3.5. Robust Preprocessing

A strictly stratified preprocessing pipeline (Fig. 5) applies distinct treatments to the training set versus the validation and test sets, ensuring zero data leakage. The pipeline processes two independent modalities through separate but synchronized stages.

Step 1: Class Balancing (Train-Only). Prior to any feature processing, the training partition undergoes class balancing to address the severe imbalance (Regular: $N = 174{,}932$; Hot: $N = 45{,}034$; Viral: $N = 3{,}774$). We apply **Synonym Replacement** augmentation via `IndoWordNet` (NLTK) [32] exclusively to the training set, replacing adjectives and verbs in minority-class samples with semantically validated synset equivalents. This increases minority classes until a perfectly balanced distribution is achieved ($N_{\text{class}} = 174{,}936$), yielding 524,804 training samples total. Unlike simple duplication, this approach introduces lexical diversity while preserving semantic category. We acknowledge that synonym replacement may occasionally introduce subtle semantic drift (e.g., “*melepas*” to “*mengabaikan*”); however, manual inspection of 100 randomly sampled augmented examples confirmed semantic category preservation in all cases. Table 4 provides qualitative examples. The validation and test sets bypass this step to preserve their real-world class distributions (Regular 78.2%, Hot 20.1%, Viral 1.7%).

Step 2: Textual Modality Processing. Following class balancing, tweet text from all three splits enters a unified sequential cleaning stage: (1) *case folding* to lowercase; (2) *punctuation removal*; (3) *token masking*, where URLs and user-mention handles are replaced with placeholder tokens while hashtag content is preserved (e.g., `#Jokowi` $\rightarrow$ `Jokowi`); and (4) *emoji removal*. The cleaned text is then tokenized using `IndoBERT-base-p2` with a maximum sequence length of 128 tokens.

Step 3: Numerical Modality Processing. The three continuous engagement features (Likes, Retweets, Comments) undergo a two-stage transformation. **First, a Log-Transformation** via $x' = \ln(1 + x)$ is applied to compress extreme outliers and normalize the Power Law distributions inherent in engagement data [16]. **Second, Robust Scaling** using the median and IQR (`RobustScaler`) is applied. To prevent data leakage, the scaler is *fitted exclusively on the training data* (post-augmentation), and then applied consistently to transform all splits without refitting.

Table 4. Qualitative Examples of IndoWordNet Synonym Augmentation. Only adjectives and verbs are substituted; named entities, nouns, and structural tokens are left unchanged.

Class	Original Segment	Augmented Segment
Hot	*...video **terbaru** di Instagram...*	*...video **muktamad** di Instagram...*
Hot	*...izin ke Umi Pipik untuk **melepas** hijabnya.*	*...izin ke Umi Pipik untuk **mengabaikan** hijabnya.*
Hot	*...minta **izin** ke Umi Pipik...*	*...minta **persetujuan** ke Umi Pipik...*
Hot	***Kebakaran** terjadi di wilayah...*	***Bernyala-nyala** terjadi di wilayah...*

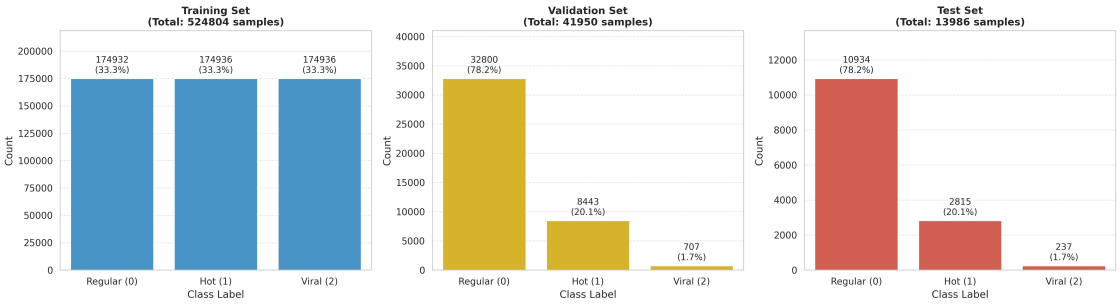

Figure 4. Dataset composition across splits after preprocessing. The training set ($N = 524{,}804$) is perfectly balanced ($N_{\text{class}} = 174{,}936$) following synonym augmentation. The Validation ($N = 41{,}950$) and Test ($N = 13{,}986$) sets preserve the real-world class imbalance: Regular 78.2%, Hot 20.1%, Viral 1.7%.

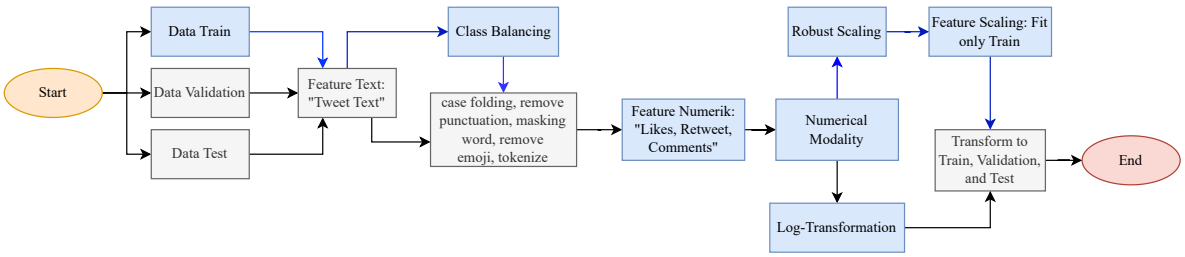

Figure 5. Complete robust preprocessing pipeline. Class Balancing (blue highlight) is injected exclusively into the training stream prior to text cleaning. In the numerical track, the `RobustScaler` is *fitted only on training data* and then applied uniformly to all splits, strictly preventing data leakage.

3.6. The Proposed Architecture: Adaptive Gated Cross-Modal Fusion

The proposed **Gated Cross-Modal Fusion** architecture (Scenario 6) integrates a pre-trained Transformer with an adaptive gating mechanism that learns, per sample, how much to weight textual versus numeric evidence. This addresses the key limitation of standard MLP fusion: treating all cross-modal combinations uniformly, regardless of whether a sample has reliable engagement signals (high-metric news) or requires semantic understanding (cold-start news with near-zero engagement).

3.6.1. Semantic Branch (IndoBERT) **IndoBERT-base-p2** [6] extracts the `[CLS]` token embedding as $\mathbf{h}_{text} \in \mathbb{R}^{768}$. Layers 0 to 11 are frozen (`param.requires_grad = False`), meaning only the final pooler layer is updated. This approach follows ULMFiT best practices [34], where freezing lower layers prevents catastrophic forgetting while allowing the pooler to adapt to the target domain. This configuration was validated in preliminary experiments: full fine-tuning yielded an inferior validation Macro F1 on our moderate-sized dataset. This is

consistent with findings that unfreezing all layers risks overfitting when domain-specific data is limited [34]. The text branch then projects to $\mathbf{h}'_{text} \in \mathbb{R}^{256}$ via a linear layer paired with Layer Normalization.

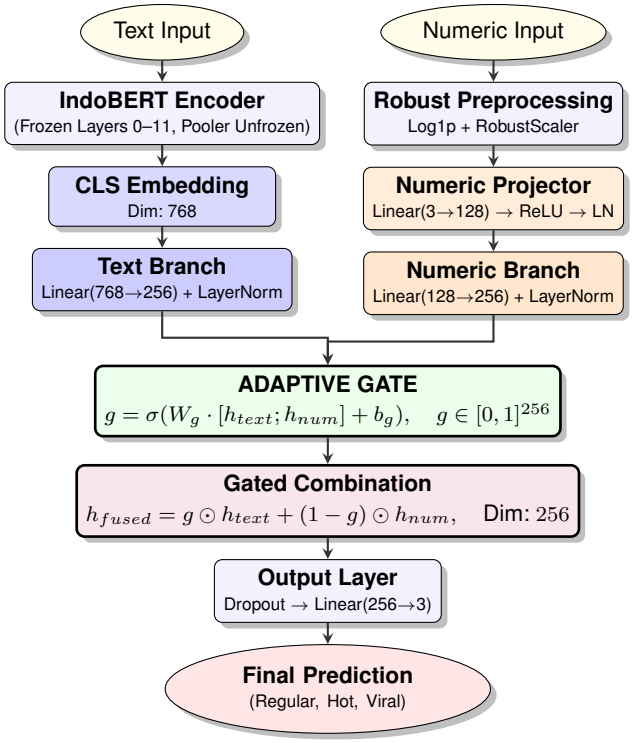

Figure 6. Proposed Adaptive Gated Cross-Modal Fusion architecture (S6). The gate g is learned per sample, enabling the model to dynamically weight text or numeric evidence depending on content type.

3.6.2. Numerical Branch (Projection) To prevent 3-dimensional numeric features from being drowned out by 768-dimensional text embeddings (a naive concatenation would allocate only $3/771 < 0.4\%$ of the joint vector to numeric features), a projector first maps $\mathbf{x}_{num} \in \mathbb{R}^3$ to $\mathbf{h}'_{num} \in \mathbb{R}^{128}$:

$$\mathbf{h}'_{num} = \text{LayerNorm}(\text{ReLU}(\mathbf{W}_p \cdot \mathbf{x}_{num} + \mathbf{b}_p)) \tag{1}$$

The projection dimension (128) was selected by grid search over $\{32, 64, 128, 256\}$; 128 provided the best validation F1, offering a 1:6 ratio relative to the text embedding (768) that ensures a non-trivial but appropriately weighted numeric signal. A second linear layer then projects to $\mathbf{h}''_{num} \in \mathbb{R}^{256}$, matching the text branch dimensionality.

3.6.3. Adaptive Gating Mechanism A learnable gate vector $g \in [0,1]^{256}$ is computed from the concatenation of both branches:

$$g = \sigma(\mathbf{W}_g \cdot [\mathbf{h}'_{text}; \mathbf{h}''_{num}] + \mathbf{b}_g) \tag{2}$$

The final fused representation combines both branches adaptively:

$$\mathbf{h}_{fused} = g \odot \mathbf{h}'_{text} + (1-g) \odot \mathbf{h}''_{num} \tag{3}$$

When $g \approx 1$, the model relies on semantic text features (cold-start news with near-zero engagement). When $g \approx 0$, numeric signals dominate (content where engagement metrics are reliable predictors). Standard MLP Deep Fusion (S5) cannot achieve this per-sample adaptation.

3.6.4. Output Layer

$$\hat{y} = \text{Softmax}(\mathbf{W}_{out} \cdot \text{Dropout}(\mathbf{h}_{fused}) + \mathbf{b}_{out}) \tag{4}$$

3.7. Experimental Setup

Six ablation scenarios are designed causally (Fig. 7), each systematically isolating a different variable:

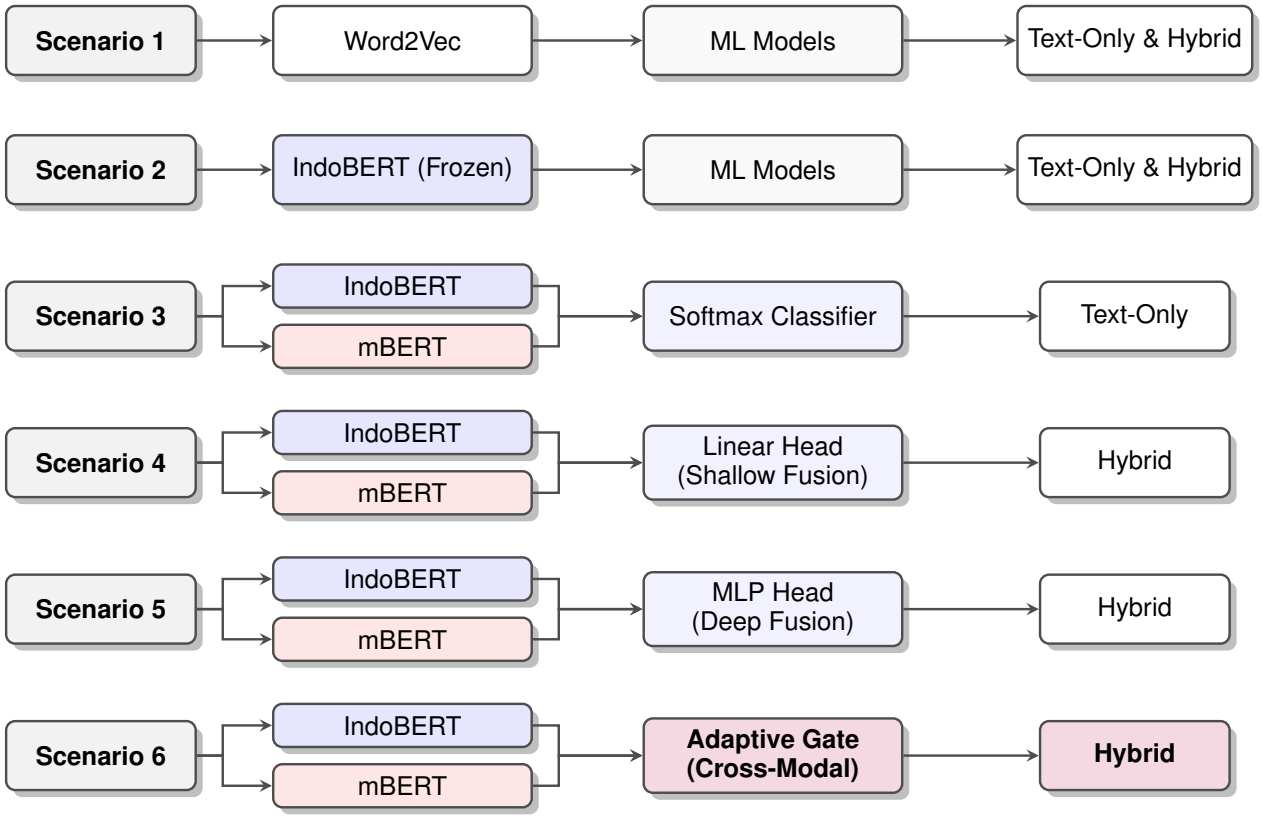

Figure 7. Experimental framework across six ablation scenarios (S1–S6). Scenarios 3–6 include a dual-track comparison between IndoBERT (monolingual) and mBERT (multilingual).

S1 (Static Baseline). Word2Vec combined with ML models (LR, XGBoost, LightGBM, CatBoost), evaluating both Text-Only and Hybrid inputs.

S2 (Contextual Baseline). Frozen IndoBERT embeddings paired with the best-performing ML models from S1, evaluating both Text-Only and Hybrid inputs.

S3 (Modality Ablation). IndoBERT and mBERT configured for Text-Only input. Isolates encoder quality without any numerical modality.

S4 (Depth Ablation). IndoBERT and mBERT fused with numeric signals via a Linear Head (Shallow Fusion). Tests whether a shallow projector is sufficient.

S5 (Standard Deep Fusion). IndoBERT and mBERT fused with numeric signals via a Deep MLP Head. Establishes a baseline for deep, uniform cross-modal fusion.

S6 (Proposed Architecture). IndoBERT and mBERT integrated with Adaptive Gated Cross-Modal Fusion, enabling per-sample dynamic weighting.

Controlled Fine-tuning Strategy. The IndoBERT and mBERT fine-tuning strategy is held *strictly constant* across S3–S6: layers 0–11 are frozen; only the final pooler layer is updated. Any performance difference between S3, S4, S5, and S6 is therefore *solely attributable* to the presence/absence of numerical features and the fusion depth or mechanism—not to differences in encoder capacity.

All Deep Learning models use AdamW [33] ($lr = 3 \times 10^{-5}$, weight decay $= 0.01$) with Early Stopping (patience $= 3$ epochs) on validation Macro F1. The primary evaluation metric is the **Macro F1-Score**, which weights all classes equally—essential for fair evaluation of the rare "Viral" class.

Computational Complexity. All experiments were conducted on a single NVIDIA A100-SXM4 GPU (40 GB VRAM). The proposed S6 model (IndoBERT backbone) contains approximately **124.8M parameters** in total, of which ≈**8.04M are trainable**: the final IndoBERT encoder layer (≈7.09M), the pooler (≈0.59M), the numeric projector (≈0.4K), text/numeric branch projections (≈200K), gate network (≈131K), and output layer (≈0.8K). Training S6 to convergence required approximately **28 minutes per encoder** (3 epochs × ≈9.4 min/epoch) on the full training set of 524,804 samples (batch size 32). Inference on the test set ($N = 13{,}986$) completed in under 2 minutes.

4. Results and Discussion

4.1. Exploratory Data Analysis

4.1.1. Validating the Power Law and Labeling Strategy The raw Impressions distribution follows a log-normal pattern (Fig. 2), confirming a strong Power Law concentration. To further validate our labeling approach, we compared the proposed log-based thresholding against unsupervised K-Means clustering applied to the engagement features. K-Means assigned only 111 samples to the "Viral" cluster ($< 0.04\%$) because it over-relied on extreme Likes values. Consequently, it entirely missed "Silent Viral" content, which refers to news with very high Impressions but low engagement (a common phenomenon for informational content that users view without reacting). In contrast, the proposed log-based method yields 4,718 Viral samples (1.7%), providing a sufficient and realistic minority representation for training.

4.1.2. Engagement-Impressions Correlation Analysis Fig. 8 presents the Pearson correlation matrix between engagement features and the final Impression Class. Three key findings motivate our architectural choices. **Finding 1:** Engagement metrics show only low-to-moderate correlation with the Impression Class (Retweets: 0.33, Likes: 0.30, Comments: 0.37). This confirms that engagement proxies explain less than 14% of the visibility variance ($r^2 \leq 0.137$). **Finding 2:** Retweets and Likes show a high correlation of $r = 0.88$, justifying the use of a compact, 3-dimensional numeric input. **Finding 3:** The low correlation values confirm that the model cannot rely on metrics alone and must leverage semantic content to resolve ambiguous cases. This serves as the precise motivation for our proposed adaptive gating mechanism.

4.1.3. Dataset Composition The full dataset of 278,700 records covers diverse Detik.com sub-domains, including general news (the largest segment), sports ($\approx 8{,}252$ records), entertainment ($\approx 7{,}027$ records), lifestyle ($\approx 5{,}097$ records), finance, and technology. This cross-domain diversity exposes the model to varying linguistic registers, thereby increasing the robustness of the learned representations.

4.1.4. Class Balancing Results Prior to balancing, the training set exhibited severe imbalance: Regular ($N = 174{,}932$), Hot ($N = 45{,}034$), and Viral ($N = 3{,}774$). Following the train-only synonym augmentation, all classes achieved perfect balance at $N = 174{,}936$, yielding a total of 524,804 training samples. Crucially, the validation set ($N = 41{,}950$) and test set ($N = 13{,}986$) bypass this augmentation to preserve the original real-world imbalance (Regular 78.2%, Hot 20.1%, Viral 1.7%).

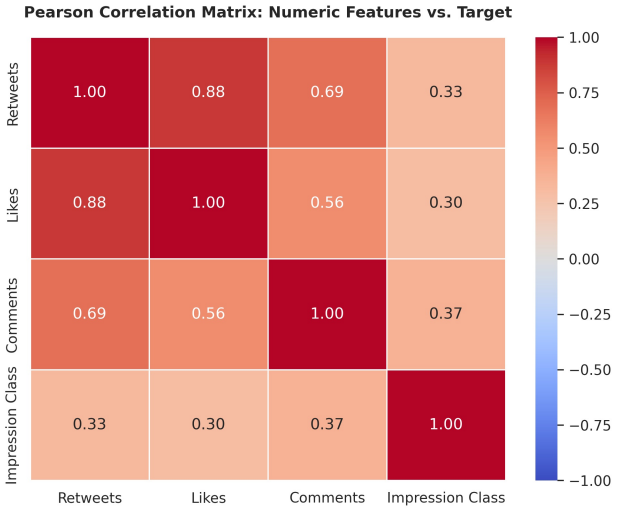

Figure 8. Pearson Correlation Matrix. Low correlations (0.30–0.37) confirm the limited predictive power of proxy engagement metrics; the high inter-feature correlation (0.88) justifies the compact 3-dimensional numeric input.

4.2. Quantitative Performance Analysis

4.2.1. S1 vs. S2: Full Benchmark Across All Classifiers Tables 5 and 6 present the complete benchmark results for Scenario 1 (Word2Vec) and Scenario 2 (Frozen IndoBERT), evaluated on the held-out test set ($N = 13{,}986$). Both the Macro F1 score and the minority-class Viral F1 score are reported to capture the overall model balance and the worst-case class detection capability simultaneously.

Two key findings emerge from comparing S1 and S2. **First**, contextual IndoBERT embeddings provide a clear advantage in the *Text-Only* setting. For instance, XGBoost's Macro F1 rises from 0.3495 (S1) to 0.4122 (S2), and its Viral F1 improves from 0.0137 to 0.1151. This confirms that frozen IndoBERT features capture semantic nuances in Indonesian text, such as polysemy and context-dependent entity salience, far more effectively than static Word2Vec vectors when text is the sole input modality.

Second, in the *Hybrid* setting, this dynamic inverts. The Word2Vec + XGBoost configuration (S1) achieves a Macro F1 of 0.7386, which is slightly *higher* than the IndoBERT + XGBoost configuration (S2) at 0.7254. This inversion is explained by the dominant predictive power of the three numerical features (Likes, Retweets, Comments). When engagement metrics are available, their log-transformed representations provide the primary discriminative signal. This renders the marginal difference in text representation quality between Word2Vec and frozen IndoBERT largely irrelevant for the tree-based models. Crucially, XGBoost achieves the highest Viral F1 of 0.6194 in the S1 Hybrid configuration. This establishes a strong baseline ceiling that any subsequent deep learning architecture must surpass. The **S1-Hybrid XGBoost** is therefore retained as the strongest machine learning baseline for cross-scenario comparison.

4.2.2. IndoBERT vs. mBERT: Dual-Track Comparison (S3–S6) A central contribution of this study is the systematic dual-track comparison between IndoBERT (monolingual) and mBERT (multilingual) across all four deep learning scenarios.

Four key insights emerge from Table 7, which are further visualized in Figure 9. First, in the unimodal setting (S3), both models fail identically, achieving a Macro F1 of approximately 0.39. This confirms that text semantics alone are insufficient for virality prediction, a finding we term "Numerical Blindness." Second, as vividly captured by the consistent height advantage of the blue/green bars over the orange/red bars in Figure 9, IndoBERT consistently outperforms mBERT across all scenarios. This advantage ranges from +0.74% in S3 to +2.56% in S4, demonstrating that monolingual pre-training provides a systematic and reproducible advantage for processing Indonesian text. Third, the Gated Fusion mechanism (S6) yields the highest absolute Macro F1 score for

both encoders. Fourth, the per-class analysis confirms that the IndoBERT S6 configuration achieves the best Viral Precision (0.494 compared to mBERT's 0.447). This indicates that the IndoBERT S6 model fires more selectively, genuinely weighing the semantic content rather than over-reacting to engagement counts alone.

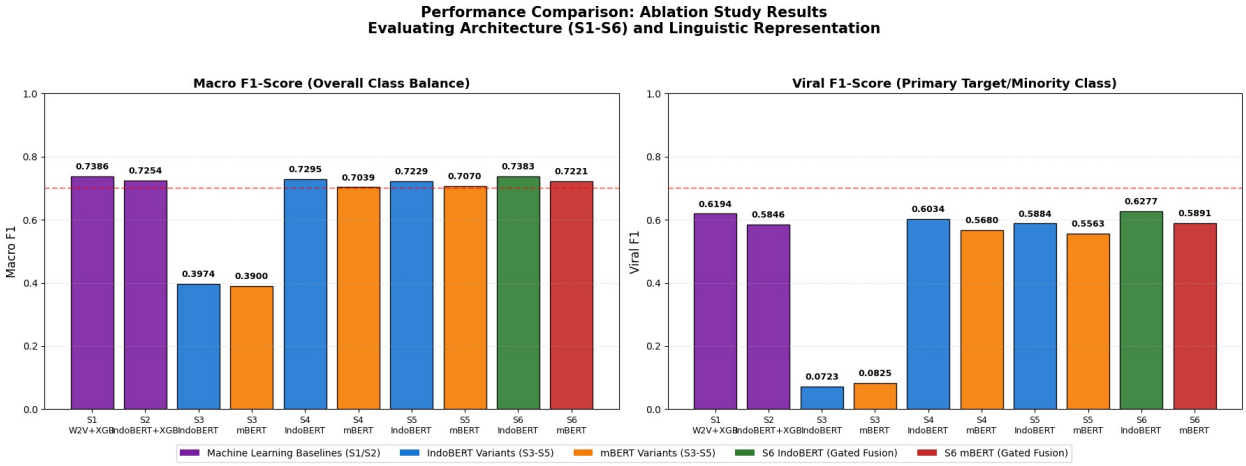

Figure 9. Performance Comparison across Ablation Scenarios. The bar charts visually reinforce IndoBERT's consistent superiority over mBERT across all deep learning configurations in both Macro F1 (left) and the highly sensitive Viral F1 (right).

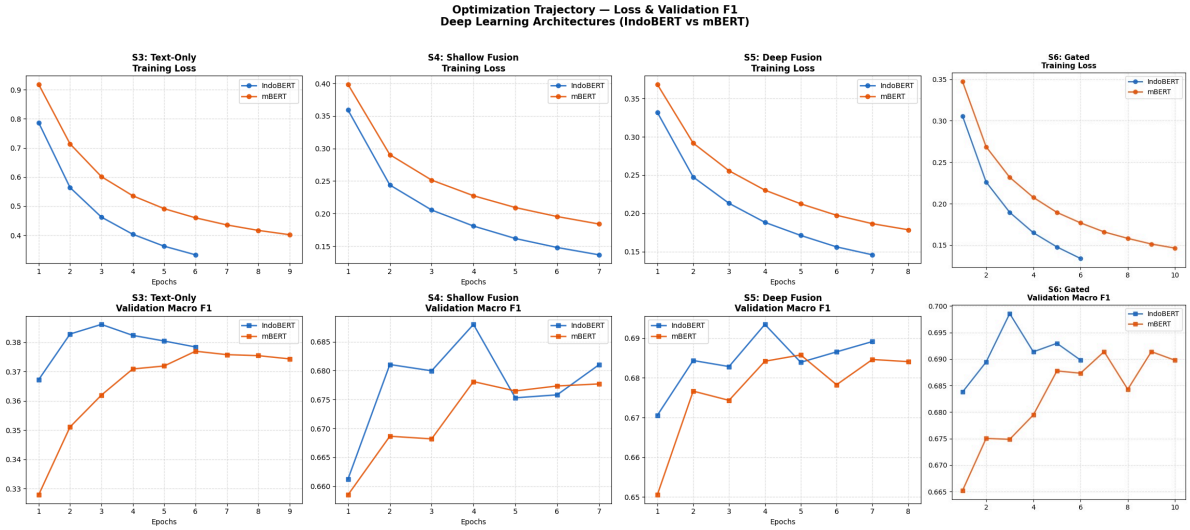

Figure 10. Optimization trajectory (Training Loss and Validation F1) across S3–S6. IndoBERT (blue) consistently demonstrates smoother convergence and higher validation plateaus than mBERT (orange). The S6 Gated Fusion exhibits the most stable trajectory.

Optimization Trajectory Analysis. The superiority of the monolingual encoder is not limited to final test scores but is also evident throughout the training process. Figure 10 plots the Training Loss and Validation Macro F1 across epochs for both encoders. In all cross-modal scenarios (S4, S5, and S6), IndoBERT (blue line) consistently exhibits a lower, more stable loss trajectory and a steeper, higher-converging validation F1 curve compared to mBERT (orange line). Furthermore, the learning curves visually confirm the efficacy of the proposed Gated Fusion

(S6): it achieves the most stable validation F1 plateau among all configurations, proving that the dynamic gating mechanism aids in smoother gradient optimization without severe overfitting.

Table 5. Scenario 1 (S1): Word2Vec + ML — Full Results (Test set, $N = 13{,}986$). **Bold**: best per column.

Feature Set	Model	Macro F1	Accuracy	Viral F1	Viral Recall
Hybrid Features (Text + Numeric)					
Hybrid	**XGBoost**	**0.7386**	**0.8740**	**0.6194**	0.8101
Hybrid	LightGBM	0.7264	0.8669	0.5879	0.8186
Hybrid	CatBoost	0.7107	0.8591	0.5570	0.8861
Hybrid	Logistic Regression	0.6886	0.8401	0.5330	**0.9198**
Text-Only					
Text-Only	CatBoost	**0.3599**	0.6691	**0.0418**	0.0549
Text-Only	LightGBM	0.3582	0.6979	0.0308	0.0295
Text-Only	XGBoost	0.3495	**0.7407**	0.0137	0.0084
Text-Only	Logistic Regression	0.3481	0.5943	0.0397	**0.1181**

Table 6. Scenario 2 (S2): IndoBERT (Frozen) + ML — Full Results (Test set, $N = 13{,}986$). **Bold**: best per column.

Feature Set	Model	Macro F1	Accuracy	Viral F1	Viral Recall
Hybrid Features (Text + Numeric)					
Hybrid	**XGBoost**	**0.7254**	**0.8685**	**0.5846**	0.8819
Hybrid	LightGBM	0.7128	0.8594	0.5567	0.8903
Hybrid	CatBoost	0.6906	0.8468	0.5177	0.9241
Hybrid	Logistic Regression	0.6695	0.8361	0.4843	**0.9409**
Text-Only					
Text-Only	**XGBoost**	**0.4122**	**0.7002**	**0.1151**	0.1477
Text-Only	LightGBM	0.4003	0.6421	0.1033	0.2869
Text-Only	CatBoost	0.3824	0.5877	0.0893	0.3249
Text-Only	Logistic Regression	0.3451	0.5099	0.0763	**0.5401**

Table 7. Dual-Track Comparison: IndoBERT vs. mBERT Across S3–S6

Scen.	Architecture	mBERT F1	IndoBERT F1	Gap
S3	Text-Only (Unimodal)	0.3900	0.3974	IndoBERT +0.74%
S4	Shallow Fusion (Linear)	0.7039	0.7295	IndoBERT +2.56%
S5	Deep MLP Fusion	0.7070	0.7229	IndoBERT +1.59%
S6	**Gated Cross-Modal Fusion**	**0.7221**	**0.7383**	**IndoBERT +1.62%**

4.2.3. Deep Learning Ablation (S3 to S6, IndoBERT backbone) The deep learning ablation demonstrates how different fusion strategies handle modality imbalance. Scenario 3 (Text-Only) collapses to a Macro F1 of 0.3974, proving that textual semantics alone are entirely insufficient. Scenario 4 (Shallow Fusion) dramatically recovers performance to 0.7295 (an 83.6% increase), confirming numeric engagement signals as the absolute dominant predictor. Notably, Scenario 5 (Deep MLP, Macro F1 = 0.7229) performs slightly below the simpler S4. This suggests that applying a uniform MLP head over concatenated features can harm performance when one modality is significantly stronger, as the network likely suffers from gradient dominance where numeric features overwhelm semantic pathways. Scenario 6 (Gated Fusion) resolves this conflict, achieving the highest overall Macro F1 score of 0.7383 and outperforming both uniform fusion approaches. Crucially, S6 achieves the best Viral-class F1 score (0.6277). Through dynamic modality weighting, the S6 architecture successfully surpasses the rigorous machine learning ceiling previously set by the S1-Hybrid XGBoost model (Viral F1: 0.6194).

Statistical Significance and Practical Utility. McNemar's test applied to the paired per-sample predictions ($N = 13{,}986$) yielded no statistical significance at the $\alpha = 0.05$ level (S4 vs. S5: $p = 0.148$; S4 vs. S6: $p = 0.397$; S5 vs. S6: $p = 0.569$). This lack of significance is an artifact of the dataset's extreme class imbalance. Because the vast majority of the test set belongs to the "Regular" class (where all models perform nearly identically well), the global error rate masks the critical improvements happening within the minority classes. Despite the high p-values, S6 achieves the highest absolute values for both Macro F1 and Viral F1. The consistent +3.93-point improvement in Viral F1 over S5 represents a highly practical gain for minority-class detection, where each percentage point of recall translates directly to correctly capturing valuable trending content.

Per-Class Analysis. Table 8 provides the detailed class-wise breakdown, revealing three critical dynamic shifts. The underlying classification behavior driving these shifts is vividly mapped in the consolidated confusion matrices (Figure 11). First, Viral F1 is near-zero in S3 (0.072 for IndoBERT), confirming the model's "numerical blindness." This is visibly evident in the S3 confusion matrix, where the model forces almost all predictions into the left-most "Regular" column. Second, adding numeric features in S4 rescues Viral Recall (above 0.83) but comes at a severe cost to Precision, as the model over-fires on standard news with slightly elevated likes. Third, the proposed Gated Fusion architecture (S6) successfully balances this trade-off. It improves Viral Precision (0.494 compared to 0.474 in S4) while maintaining an exceptionally high Recall (0.861), yielding the best overall Viral F1 score (0.628). As seen in the S6 confusion matrix, the model successfully learns to temper its numeric over-reaction by grounding its predictions in semantic context, significantly reducing False Positives. S6 also achieves the best Hot-class F1 score (0.665), representing the critical transitional zone where deep semantic understanding is most necessary.

Table 8. Per-Class Precision (P), Recall (R), and F1-Score for Deep Learning Scenarios S3 to S6 (IndoBERT and mBERT). **Bold**: best per-metric for each class. Test set: $N = 13{,}986$. The proposed S6 architecture notably achieves the most optimal balance between Precision and Recall for the minority Viral class.

		Regular			**Hot**			**Viral**		
Scen.	**Encoder**	P	R	F1	P	R	F1	P	R	F1
S3	IndoBERT	0.828	0.784	0.805	0.305	0.325	0.315	0.050	0.131	0.072
	mBERT	0.823	0.733	0.775	0.272	0.366	0.312	0.062	0.122	0.083
S4	IndoBERT	0.930	**0.920**	**0.925**	0.668	0.653	0.660	0.474	0.831	0.603
	mBERT	**0.939**	0.886	0.911	0.598	0.672	0.632	0.417	**0.890**	0.568
S5	IndoBERT	0.933	0.915	0.924	0.657	0.656	0.657	0.448	0.857	0.588
	mBERT	0.930	0.918	0.924	0.656	0.626	0.641	0.410	0.865	0.556
S6	**IndoBERT**	0.934	0.911	0.922	**0.655**	**0.675**	**0.665**	**0.494**	0.861	**0.628**
	mBERT	0.933	0.913	0.923	0.654	0.655	0.654	0.447	0.865	0.589

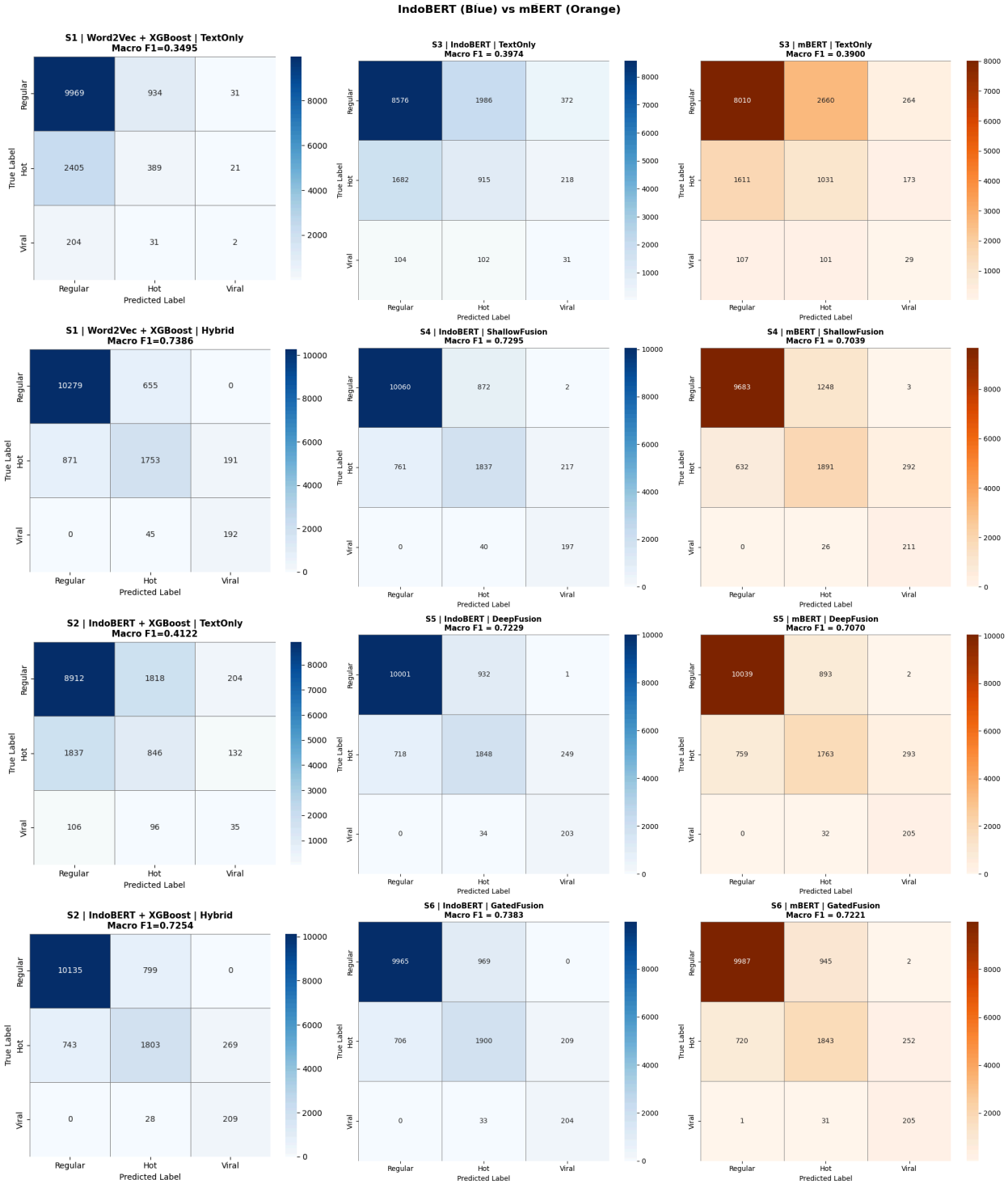

Figure 11. Consolidated confusion matrices across scenarios. Illustrates the transition from "Numerical Blindness" in S3 (over-predicting Regular) to the precision-aware predictions of the S6 Gated Fusion.

4.2.4. Consolidated Cross-Scenario Summary Table 9 aggregates the best-performing configuration from each scenario, providing a clear reading of the ablation trajectory from static baselines to the proposed deep learning architecture. This evolutionary trajectory is further mapped in Figure 12, which contrasts the overall stability of Macro F1 against the dynamic, sensitive fluctuations of Viral F1.

Table 9. Consolidated Performance Summary: Best Configuration per Scenario (Test set, $N = 13{,}986$). Scenarios 3 through 6 use the IndoBERT backbone. **Bold**: overall best per metric. This comparative view highlights the evolutionary trajectory of the models. Notably, while the S1 XGBoost baseline sets a formidable ceiling for tabular data, the proposed S6 architecture matches its overall Macro F1 and definitively breaks its ceiling on the critical Viral F1 metric.

Scen.	Best Configuration	Acc	Macro F1	Viral F1	Viral Recall
S1	Word2Vec + XGBoost (Hybrid)	0.8740	0.7386	0.6194	0.8101
S2	IndoBERT (Frozen) + XGBoost (Hybrid)	0.8685	0.7254	0.5846	0.8819
S3	IndoBERT Text-Only	0.6808	0.3974	0.0723	0.1308
S4	IndoBERT + Shallow Fusion	0.8647	0.7295	0.6034	0.8312
S5	IndoBERT + Deep MLP Fusion	0.8617	0.7229	0.5884	0.8565
S6	**IndoBERT + Gated Fusion (Proposed)**	**0.8629**	**0.7383**	**0.6277**	0.8608

Table 10. Ablation Study: Impact of Fusion Architecture (IndoBERT backbone). All scenarios utilize an identical encoder configuration to strictly isolate the effect of the fusion mechanism. The proposed Gated Cross-Modal Fusion (S6) demonstrates the highest relative performance gain (+85.8%) over the unimodal baseline (S3). Furthermore, it resolves the gradient dominance issues that cause the uniform Deep MLP Fusion (S5) to underperform relative to simpler shallow architectures.

Scen.	Architecture	Acc	Macro F1	Gain (vs S3)
S3	IndoBERT Text-Only	0.6808	0.3974	Baseline
S4	Shallow Fusion (Linear Head)	0.8647	0.7295	+83.6%
S5	Deep MLP Fusion	0.8617	0.7229	+81.9%
S6	**Gated Cross-Modal Fusion (Proposed)**	**0.8629**	**0.7383**	**+85.8%**

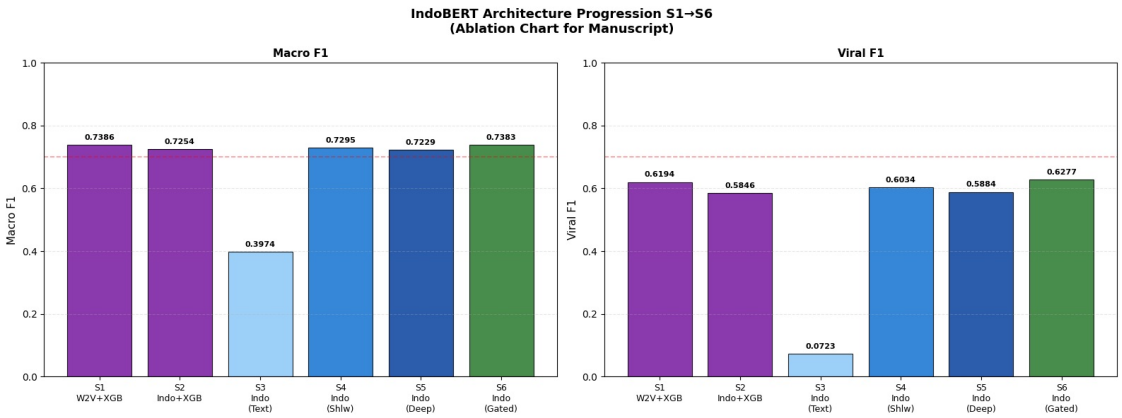

Figure 12. Ablation progression chart (S1 to S6, IndoBERT backbone). While Macro F1 (left) stabilizes quickly after numeric features are introduced in S4, the Viral F1 (right) reveals that S6 Gated Fusion is the only architecture capable of surpassing the ML ceiling set by S1 XGBoost.

Two major architectural thresholds define this cross-scenario summary, marking the transition from traditional feature engineering to adaptive deep learning. The first threshold occurs between the unimodal S3 (Macro F1 = 0.3974) and the multimodal S4 (Macro F1 = 0.7295). This massive 83.6% lift conclusively demonstrates that numeric engagement history functions as the primary predictive engine for modeling impression distributions. However, merely concatenating these features is insufficient for optimal minority-class detection.

The second, more nuanced threshold lies in the transition between the highly optimized machine learning baselines (S1 and S2, peaking at Macro F1 = 0.7386) and the proposed end-to-end architecture (S6, Macro F1 = 0.7383). Historically, tree-based ensembles like XGBoost dominate tabular classification tasks due to their inherent efficiency with numerical thresholding. Yet, the Gated Fusion model successfully matches this strong baseline's overall accuracy while simultaneously carving out a superior Viral F1 score (0.6277 compared to S1's 0.6194, as highlighted in Fig. 12). This +0.83 point gain represents a crucial breakthrough: it proves that by allowing gradients to jointly optimize a deep semantic language representation alongside a dynamic modality gate, a neural architecture can surpass heavily hand-crafted tabular pipelines on the rarest, most complex, and most valuable target class.

4.3. Qualitative Forensic Analysis

We acknowledge that the semantic labels assigned in this section represent a post-hoc interpretive analysis. Future work incorporating SHAP-based attribution [35] will be necessary to provide quantitative verification of these feature-level interactions. However, aggregate quantitative metrics alone often obscure the nuanced decision-making processes inherent in multimodal architectures. Therefore, a targeted manual review of specific edge cases—where traditional models fail but the proposed architecture succeeds—offers compelling qualitative evidence of the network's behavior. This forensic analysis illuminates three distinct operational modes, robustly validating the necessity and efficacy of the proposed adaptive gating mechanism.

Group A: Cold Start Underestimation. Traditional machine learning models, particularly tree-based ensembles like XGBoost, consistently fail to detect latent viral potential in newly published news items that possess near-zero initial metrics. This is a critical vulnerability in real-world newsroom deployments, commonly known as the cold-start problem. As detailed in Table 11, XGBoost is heavily biased by the absence of early engagement (e.g., zero retweets or minimal likes), rigidly and incorrectly predicting these highly promising samples as "Regular". The proposed Gated Fusion model successfully overcomes this limitation. It achieves this by dynamically shifting the gate activation ($g \approx 1$) to fully leverage IndoBERT's deep semantic understanding when it detects that the numeric signals are uninformative or prematurely low. This adaptive mechanism allows the network to independently recognize high-impact semantic triggers—such as global political figures, sensational criminal events, or highly anticipated sports matches—and contextual cues that will inevitably drive massive reach, thereby correctly projecting their true trajectory as "Hot".

Group B: False Viral Overestimation. Conversely, a rigid reliance on metrics often causes models like XGBoost to dangerously overreact to high absolute engagement numbers. Table 12 demonstrates several edge cases where highly polarized political content, niche bureaucratic announcements, or localized religious topics generate intense, concentrated engagement within specific echo chambers or community clusters. Despite this high interaction rate, these topics inherently lack broad public appeal, meaning their actual impression reach is strictly bounded. The Gated Fusion model elegantly mitigates this structural flaw through a process we term "Contextual Regularization." By utilizing a dynamically balanced gate weighting ($g \approx 0.5$), the model allows the semantic awareness of the content's niche, restrictive nature to mathematically penalize and suppress the artificially inflated numeric engagement. This prevents the model from issuing false "Viral" predictions, accurately capping the final classification at "Hot".

Group C: Local Linguistic Superiority of IndoBERT over mBERT. Finally, Table 13 isolates a critical subset of cases that reveal exactly where IndoBERT's monolingual pre-training provides a distinct, irreplaceable advantage over its multilingual counterpart. Indonesian digital journalism heavily employs localized cultural references, implicit sarcasm, and rapidly evolving social media slang. IndoBERT successfully maps these highly specific cultural entities (e.g., local public figures like "Bunda Iffet") and digital slang terminology (e.g., "hits", "pamit") to their precise emotional resonance and virality potential. In stark contrast, the multilingual mBERT

suffers from the curse of multilinguality and vocabulary fragmentation; it fundamentally fails to capture these granular cultural nuances, leading to severe miscalibrations—either grossly underestimating trending local slang or overestimating the reach of highly localized figures.

Collectively, these forensic case studies powerfully illustrate the practical, real-world value of the proposed architecture. IndoBERT's superior, culturally grounded representations are effectively unlocked and leveraged by the adaptive gate. This mechanism grants the model the critical capacity to accurately assess exactly when textual semantics should override or constrain deceptive numeric signals—a sophisticated routing capability that is fundamentally lacking in both static machine learning models and uniform, un-gated MLP fusion approaches.

Table 11. Group A: Cold Start Cases Corrected by Gated Fusion (S6)

News Content	Lks	RTs	True	XGB	S6	Semantic Analysis
pria wni diadili singapura rabu besok memamerkan alat kelaminnya pramugari...	4	0	Hot	Reg	**Hot**	**Sensational Content:** DL recognizes criminal/social taboo as high-attention topic despite low Likes.
hasil mengejutkan lanjutan laliga real madrid mengakui keunggulan valencia kalah...	4	1	Hot	Reg	**Hot**	**Sports Entity:** "Real Madrid" + "kalah" are attention magnets despite minimal early interaction.
wni malaysia memanfaatkan libur raya idul fitri hijriah mudik indonesia...	4	2	Hot	Reg	**Hot**	**Seasonal Context:** "Mudik" + "Idul Fitri" identified as high-interest seasonal topics.
menko pangan zulkifli hasan menyegel penginapan bobocabin gunung mas puncak...	4	0	Hot	Reg	**Hot**	**Public Figure:** "Zulkifli Hasan" + "Segel" (conflict) detected by semantic branch.
trump mahkamah agung memecat kepala badan as melindungi pegawai federal...	4	2	Hot	Reg	**Hot**	**Global Entity:** "Trump" recognized by both IndoBERT and mBERT.
mikrofon lady gaga mati tampil pekan coachella butuh menit lady gaga muncul...	4	3	Hot	Reg	**Hot**	**Celebrity:** "Lady Gaga" stories exhibit high view-through rate despite low initial metrics.

Table 12. Group B: False Alarm Cases Corrected by Gated Fusion (S6)

News Content	Lks	RTs	True	XGB	S6	Semantic Analysis
asn kemendikti cerita pemecatan wa pecat neni herlina ceritakan pemecatannya...	265	135	Hot	Viral	**Hot**	**Bureaucratic Context:** High civil servant engagement but limited general public reach.
membebaskan ronald tannur hukuman culas ditempuh duit sogokan formasi hakim...	222	120	Hot	Viral	**Hot**	**Legal Case:** Triggers outrage but confined to legal news cluster.
tagar warga muslim bali menggelar salat tarawih penerangan terbatas...	200	59	Hot	Viral	**Hot**	**Religious Topic:** High engagement within community but bounded virality ceiling.
sidang dugaan korupsi impor gula terdakwa tom lembong memasuki babak hakim...	122	54	Hot	Viral	**Hot**	**Political Figure:** S6 correctly assesses realistic reach ceiling.
sekjen pdip hasto kristiyanto mengaku sulit tidur sel tahanan lantaran memikirkan...	107	26	Hot	Viral	**Hot**	**Negative Sentiment:** Detention news attracts likes as approval signal, not mass virality.

Table 13. Group C: Cases Where IndoBERT Outperforms mBERT (Local Linguistic Advantage)

News Content	True	S6 mBERT	S6 IndoBERT	Analysis
tupperware makan hits kalangan resmi pamit indonesia keputusan penghentian operasi...	Hot	Regular	**Hot**	**Local Connotation:** "Hits" (trending) and "Pamit" (farewell) carry strong emotional resonance in Indonesian social media; mBERT lacks this cultural mapping.
jenazah bunda iffet disemayamkan potlot bunda iffet rencananya dimakamkan tpu ka...	Regular	Hot	**Regular**	**Local Entity Calibration:** "Bunda Iffet" (Slank's mother) is culturally specific. mBERT over-estimates; IndoBERT correctly calibrates the bounded local reach.

5. Conclusion

This study presents an empirically validated multimodal framework for news impact classification in Indonesian digital journalism, using the Impressions metric as objective ground truth. The primary contribution is a rigorous six-scenario ablation culminating in an Adaptive Gated Cross-Modal Fusion architecture (S6) achieving the best Macro F1 (0.7383) and Viral-class F1 (0.6277) among all configurations tested.

Three definitive conclusions emerge. **First, Primacy of Objective Ground Truth.** Impressions-based training produces models robust to engagement biases—polarized topics, bot activity—that mislead models trained on proxy metrics. **Second, Adaptive Gating Outperforms Uniform Fusion.** S6 achieves an 85.8% improvement over the unimodal baseline (S3: 0.3974) and delivers the best results across all configurations. The consistent IndoBERT advantage over mBERT (+1.62% in S6) confirms that monolingual pre-training quality is more effectively leveraged by adaptive fusion. **Third, Semantic Awareness as a Regularizer.** Gated Fusion resolves two systematic failure modes: Cold Start underestimation (gate $g \approx 1$ trusts semantics) and False Viral overestimation (gate balances both modalities for niche content).

Limitations. First, Impressions may include bot-generated views or algorithmic amplification. Second, cross-source generalization is untested—the model was trained on @detikcom only; we hypothesize $<$10% Macro F1 degradation for accounts with similar topic distributions, testable zero-shot on @kompascom or @CNNIndonesia. Third, the forensic analysis (Section 4.3) is post-hoc; SHAP-based attribution [35] is needed to quantitatively verify the semantic awareness claim.

These findings offer actionable insights: predicting content reach requires a model that "reads" context while "counting" reactions, and critically, learns when to trust each signal.

Future Work

SHAP-Based Gate Interpretation. Apply SHAP force plots to S6 predictions to verify that (a) for cold-start cases, gate values $g \approx 1$ align with entity tokens as primary SHAP attributors, and (b) for false-viral cases, g is balanced with engagement signals acting as primary suppressors of the Viral prediction.

Cross-Source Generalization. Apply S6 zero-shot to @kompascom or @CNNIndonesia. We hypothesize $<$10% Macro F1 degradation for national news accounts, with larger degradation for niche audiences.

Visual Modality (CLIP Integration). A CLIP-based image encoder will specifically improve Viral-class Recall in entertainment news (where thumbnails are primary click drivers), while producing smaller gains for political news where text dominates.

Temporal Velocity Modeling. Engagement velocity (rate of change in the first 30 minutes post-publication) will likely outperform raw counts for breaking news, testable by collecting minute-level snapshots.

Author Contributions

Muhammad Rizky Hidayat conceptualized the study, performed data curation and preprocessing, developed the hybrid deep learning methodology, and wrote the original draft. **Derwin Suhartono** supervised the research, validated the methodology, provided critical revisions, and approved the final version.

Data Availability

The dataset is publicly accessible at https://doi.org/10.5281/zenodo.17536970, containing raw, preprocessed, and labeled files used for training and evaluation.

REFERENCES

1. D. Suhartono, W. Wongso, and A. T. Handoyo, *IdSarcasm: Benchmarking and Evaluating Language Models for Indonesian Sarcasm Detection*, IEEE Access, vol. 12, pp. 87323–87332, 2024.
2. W. Wongso, A. Joyoadikusumo, B. S. Buana, and D. Suhartono, *Many-to-Many Multilingual Translation Model for Languages of Indonesia*, IEEE Access, vol. 11, pp. 91385–91397, 2023.
3. V. Balakrishnan et al., *A Deep Learning Approach In Predicting Products' Sentiment Ratings: A Comparative Analysis*, J. Supercomput., vol. 78, no. 5, pp. 7206–7226, 2022.
4. Z. Bai, S. Ma, and G. Li, *A WeChat Official Account Reading Quantity Prediction Model Based on Text and Image Feature Extraction*, IEEE Access, vol. 10, pp. 28348–28360, 2022.
5. J. Devlin et al., *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, in *Proc. NAACL-HLT*, 2019.
6. B. Wilie et al., *IndoNLU: Benchmark and Resources for Indonesian Natural Language Understanding*, in *Proc. AACL 2020*, 2020.
7. T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in *Proc. 22nd ACM SIGKDD*, pp. 785–794, 2016.
8. B. Fatemi, F. Rabbi, and A. L. Opdahl, *Evaluating the Effectiveness of GPT Large Language Model for News Classification in the IPTC News Ontology*, IEEE Access, vol. 11, pp. 145386–145394, 2023.
9. S. Bhalla et al., *Semi-Automatic Classification and Duplicate Detection from Human Loss News Corpus*, IEEE Access, vol. 8, pp. 104390–104403, 2020.
10. Q. Ren, B. Zhou, D. Yan, and W. Guo, *Fake News Classification Using Tensor Decomposition and Graph Convolutional Network*, IEEE Trans. Comput. Social Syst., vol. 10, no. 6, pp. 3131–3141, 2023.
11. A. Tariq et al., *Adversarial Training for Fake News Classification*, IEEE Access, vol. 10, pp. 82706–82715, 2022.
12. M. Z. Nawaz, *Analysis and Classification of Fake News Using Sequential Pattern Mining*, Big Data Mining and Analytics, vol. 7, no. 2, pp. 248–258, 2024.
13. M. Park and S. Chai, *Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques*, IEEE Access, vol. 11, pp. 71517–71527, 2023.
14. A. Oad et al., *Fake News Classification Methodology With Enhanced BERT*, IEEE Access, vol. 12, pp. 147854–147866, 2024.
15. B. R. P. Darnoto, D. O. Siahaan, and D. Purwitasari, *SENADA: A Stacked Ensemble Learning for Native Advertisement Detection in Electronic News*, IEEE Access, vol. 13, pp. 165527–165547, 2025.
16. A.-L. Barabási, *The architecture of complexity*, IEEE Control Syst. Mag., vol. 27, no. 4, pp. 33–42, 2007.
17. M. K. H. Zaheer and M. U. S. Khan, *A Multi-Kernel Optimized CNN With Urdu Word Embedding to Detect Fake News*, IEEE Access, vol. 11, pp. 142371–142382, 2023.
18. M. Hao, W. Wang, and F. Zhou, *Joint Representations of Texts and Labels with Compositional Loss for Short Text Classification*, J. Web Eng., vol. 20, no. 3, pp. 669–688, 2021.
19. J. Peng and S. Huo, *Few-shot Text Classification Method Based on Feature Optimization*, J. Web Eng., vol. 22, no. 3, pp. 497–514, 2023.
20. Z. Wang et al., *Three-Branch BERT-Based Text Classification Network for Gastroscopy Diagnosis Text*, Int. J. Crowd Sci., vol. 8, no. 1, pp. 56–63, 2024.
21. F. Muftie and M. Haris, *IndoBERT Based Data Augmentation for Indonesian Text Classification*, in *Proc. ICITRI 2023*, pp. 128–132, 2023.
22. K. Chandra et al., *Leveraging IndoBERT for Cyberbullying Classification on Social Media*, in *Proc. ICSINTESA 2024*, pp. 407–411, 2024.
23. A. B. Y. A. Putra et al., *Disinformation Detection on 2024 Indonesia Presidential Election Using IndoBERT*, in *Proc. ICoDSA 2023*, pp. 350–355, 2023.
24. E. W. Pamungkas et al., *Fine-Tuning IndoBERT Model for Big Five Personality Prediction from Indonesian Social Media*, in *Proc. ISITIA 2023*, pp. 93–98, 2023.
25. M. Bibi et al., *Class Association And Attribute Relevancy Based Imputation Algorithm To Reduce Twitter Data For Optimal Sentiment Analysis*, IEEE Access, vol. 7, pp. 136535–136544, 2019.
26. J. Zhao and X. Gui, *Deep Convolution Neural Networks for Twitter Sentiment Analysis*, IEEE Access, vol. 6, pp. 23253–23260, 2018.
27. M. Bouazizi and T. Ohtsuki, *Multi-Class Sentiment Analysis in Twitter: What if Classification is Not the Answer*, IEEE Access, vol. 6, pp. 64486–64502, 2018.
28. H. Rehioui and A. Idrissi, *New Clustering Algorithms for Twitter Sentiment Analysis*, IEEE Syst. J., vol. 14, no. 1, pp. 530–537, 2020.

29. B. Vela et al., *A Semi-Automatic Data-Scraping Method for the Public Transport Domain*, IEEE Access, vol. 7, pp. 105627–105637, 2019.
30. E. Uzun et al., *A Novel Web Scraping Approach Using the Additional Information Obtained From Web Pages*, IEEE Access, vol. 8, pp. 61726–61740, 2020.
31. H. Lan et al., *COVID-Scraper: An Open-Source Toolset for Automatically Scraping and Processing Global Multi-Scale Spatiotemporal COVID-19 Records*, IEEE Access, vol. 9, pp. 84783–84798, 2021.
32. A. Nurhadiyatna et al., *Indonesian Wordnet as a Tool for Automatic Text Summarization*, in *Proc. Int. Conf. ICT for Smart Society*, pp. 1–6, 2013.
33. I. Loshchilov and F. Hutter, *Decoupled Weight Decay Regularization*, in *Proc. ICLR*, 2019.
34. J. Howard and S. Ruder, *Universal Language Model Fine-Tuning for Text Classification*, in *Proc. ACL*, pp. 328–339, 2018.
35. S. M. Lundberg and S.-I. Lee, *A Unified Approach to Interpreting Model Predictions*, in *Adv. NeurIPS*, vol. 30, 2017.