Jurnal ELTIKOM :
Jurnal Teknik Elektro, Teknologi Informasi dan Komputer

Vol. 10, No. 1, June 2026, page. 118-130
ISSN 2598-3245 (Print), ISSN 2598-3288 (Online)
DOI: http://doi.org/10.31961/eltikom.v10i1.2132
Available online at http://eltikom.poliban.ac.id

LSTM-BASED CAUSAL ATTRIBUTION MODELING OF THE 2025 SUMATRA FLASH FLOOD DISCOURSE ON YOUTUBE

Kunti Najma Jalia*, Adi Suwondo, Hidayatus Sibyan
Department of Informatics Engineering, Quranic Science University, Wonosobo, Indonesia
e-mail: najmajalia@mhs.unsiq.ac.id, adisuwondo@unsiq.ac.id, hsibyan@fastikom-unsiq.ac.id

Received: 19 December 2025 – Revised: 25 March 2026 – Accepted: 21 April 2026

ABSTRACT

Existing disaster sentiment analysis mainly focuses on emotional polarity classification, while often overlooking the causal reasoning that shapes public discourse on responsibility for disaster outcomes. This study proposes and assesses a Long Short-Term Memory (LSTM)-based causal attribution classification framework to examine YouTube comments related to the 2025 Sumatra flash flood. It compares LSTM performance with Support Vector Machine (SVM) and Naïve Bayes baselines. A total of 17,503 publicly available comments were collected through the YouTube Data API v3 and processed into a final dataset of 12,299 comments. The comments were classified into two causal categories, human factor and nature/prayer factor, using lexicon-based scoring validated by three independent annotators (Cohen's κ = 0.81). The experimental results show that LSTM achieves 98.17% accuracy with strong stability (±0.25% standard deviation) under stratified five-fold cross-validation, substantially outperforming SVM (82.83%) and Naïve Bayes (75.04%). These findings indicate that sequence-based architectures can capture the contextual dependencies in causal attribution discourse, offering a replicable framework for disaster risk communication monitoring systems.

Keywords*: causal attribution, disaster discourse, LSTM, social media analysis, YouTube comments.*

I. INTRODUCTION

FLASH floods and landslides in Sumatra at the end of 2025 were recorded as the largest hydrometeorological disaster in the past decade, with catastrophic impacts exceeding the government's worst-case scenario projections [1]. Official statistics from the National Disaster Management Agency (BNPB), updated on December 13, 2025, reported 1,006 fatalities, 217 missing persons, and 654,652 displaced residents following extreme weather anomalies induced by Tropical Cyclone Senyar [2]. Daily rainfall intensity exceeded 300 mm, equivalent to a normal monthly accumulation, and was worsened by upstream watershed degradation caused by massive deforestation. As a result, critical infrastructure across three affected provinces was paralyzed [3]. The scale of this impact positions the event not only as a regional physical crisis but also as a social phenomenon that drew national and international attention and triggered extensive digital discourse on disaster causality and accountability.

Social media proliferation has fundamentally changed how the public expresses views on natural catastrophes, with YouTube becoming one of the dominant platforms for disaster-related content consumption and discussion in Indonesia [4]. Unlike microblogging platforms such as Twitter/X, which impose character limits, YouTube comment sections allow more extended and reasoned viewpoints, making users' causal reasoning more explicit [5], [6]. Previous studies show that comments on YouTube news content contain a higher proportion of structured argumentation [7], making the platform a rich data source for analyzing public understanding of disaster causation. Digital interactions surrounding the 2025 Sumatra flash flood captured a polarization of public perceptions between theological narratives grounded in fate and anthropogenic narratives emphasizing human negligence and governance failures [8], [9].

Computational approaches using machine learning have become increasingly prominent in disaster research for analyzing public reactions and supplementing traditional monitoring infrastructure. A comprehensive meta-analysis of 139 publications shows the prevalence of traditional classifiers, especially SVM and Naïve Bayes, due to their processing speed and model transparency [10]. In forest fire analysis, SVM models enhanced with the Synthetic Minority Oversampling Technique (SMOTE) have reached 90.31% accuracy on datasets with clearly expressed linguistic features [11]. Probabilistic methods such as Naïve Bayes commonly show accuracy plateaus between 76% and 79%, mainly because their feature independence assumptions limit adaptability to the diverse and heterogeneous nature of social media text [12]. The dependence of conventional algorithms on domain-specific linguistic characteristics indicates limitations in cross-disaster generalizability.

Processing Indonesian disaster narratives presents complex linguistic challenges, including sarcasm, irony, multilingual code-switching, and rapidly evolving informal vocabulary [13], [14]. Studies on sarcastic flood-related texts report that hybrid models achieve a maximum accuracy of 77%, while standard models remain limited to approximately 76% on imbalanced datasets [15]. These limitations reflect the inability of bag-of-words approaches to capture sentence-level context and semantic relationships across words in long, implicitly nuanced texts [16]. The use of static sentiment lexicons further limits model adaptability to the dynamic nature of informal social media language [17]. This performance gap becomes critical when the analytical objective shifts from sentiment polarity to causal attribution.

Such limitations provide a rationale for examining deep learning–based models, especially LSTM networks, because they are effective in learning long-term relationships in sequential representations [18], [19]. LSTM gate structures support context-sensitive information selection and retention [20], making the architecture useful for capturing implicit criticism and policy-related irony in disaster discourse [21]. Recent studies show that LSTM implementations for Indonesian-language extreme-weather sentiment analysis achieve an average validation accuracy of 94.33%, reinforcing the advantage of sequential approaches over manual feature-engineering methods [22].

Despite the growing body of research on disaster-related social media analysis, existing computational studies predominantly operationalize public responses through emotional polarity classification, categorizing discourse as positive, negative, or neutral without examining the underlying causal reasoning structures. Although attribution theory has been widely applied in crisis communication research to explain how individuals assign responsibility for disasters [23], its use in computational text classification remains underdeveloped, particularly for low-resource languages such as Indonesian. Prior computational studies on blame and responsibility attribution in disaster discourse have relied mainly on keyword-based approaches or topic modeling rather than supervised sequence classification [24], limiting their capacity to capture the implicit and context-dependent nature of causal reasoning in informal social media text. Furthermore, even recent studies examining social media discourse related to the 2025 Sumatra disaster have focused only on emotional sentiment and trauma-related expression rather than causal attribution patterns [25]. This gap is substantive because causal attribution directly informs policy response, institutional accountability, and risk communication strategy in ways that sentiment polarity alone cannot capture.

Based on these gaps, this study addresses three research questions: (1) What is the distribution of causal attribution categories in YouTube discourse regarding the 2025 Sumatra flash flood? (2) How does LSTM performance compare with SVM and Naïve Bayes in classifying causal attribution? (3) What classification patterns emerge from error analysis across models?

The novelty of this study is mainly analytical and domain-specific rather than architectural. From an application perspective, it reframes disaster discourse analysis from sentiment polarity to causal attribution, treating public discourse as reasoned judgment rather than emotional reaction, an analytical orientation that has not been systematically applied to Indonesian disaster communication on YouTube. From a methodological perspective, it introduces a validated lexicon-based labeling scheme purpose-built for Indonesian disaster discourse, distinguishes anthropogenic from theological or naturalistic attributions, and empirically evaluates the suitability of sequential architectures for capturing implicit causal reasoning patterns within this domain.

This research presents three primary contributions: a theoretical contribution by reframing disaster sentiment analysis as causal reasoning classification; a methodological contribution through the development of a validated causal annotation framework with verified inter-annotator reliability, applicable

TABLE 1
YOUTUBE VIDEO SOURCES USED FOR DATA COLLECTION

No	Video ID	Video Tittle	Channel
1	7tj0KRKvYto	EKSKLUSIF! Banjir Sumatra: Melihat Perusakan Hutan Lebih Dekat	Liputan6
2	R4CJ2JaX5zA	SUMBAR BERDUKA 11/12/2025: Sumatera Barat Kembali Banjir Besar & Longsor	Bencana Populer
3	OTmJI8MyFDE	FULL! Greenpeace Soal Bencana Sumatra: Tak Ada Ekonomi Tumbuh di Bumi yang Mati	KompasTV
4	Pf44-AlpIW8	Dandhy Laksono Bongkar Dalang Kerusakan Alam di Sumatra	dr. Richard Lee
5	85pfSIuHuk0	Kegagapan Pemerintah dan Penyebab Banjir dan Longsor di Sumatera	Tempodotco
6	W7ys2FOR3Ao	Presiden Prabowo Tinjau Lokasi Bencana di Sumatra dan Aceh	Metro TV
7	Lo_D1qaG60I	Prabowo Tiba di Bireuen Aceh, Temui Korban Banjir hingga Tinjau Perbaikan Jembatan	KompasTV
8	vK4daD9T9cY	Apa yang Terjadi dalam Banjir-Longsor di Aceh, Sumut, dan Sumbar?	Kompas.com
9	hiP76ZBZnw4	Gubernur Aceh Bicara: Status Bencana, Kepala Daerah yang Menyerah dan Bantuan Asing	Najwa Shihab
`10	BopPg9R5Y1k	[FULL] Melacak Biang Kerok Bencana Sumatra	tvOneNews

["yg": "yang", "ga": "tidak", "gak": "tidak", "nggak": "tidak", "tdk": "tidak", "dgn": "dengan", "dr": "dari", "bgt": "banget", "krn": "karena", "krna": "karena", "utk": "untuk", "tlg": "tolong", "sdh": "sudah", "udh": "sudah", "aja": "saja", "bkn": "bukan", "kalo": "kalau", "kl": "kalau", "tuh": "itu", "ni": "ini", "jgn": "jangan", "sampe": "sampai", "knp": "kenapa", "gmn": "bagaimana", "bbrp": "beberapa", "trus": "terus", "skrg": "sekarang", "tau": "tahu", "sumut": "sumatera utara", "sumbar": "sumatera barat", "aceh": "aceh", "konoha": "indonesia", "wakanda": "indonesia", "rezim": "pemerintah", "olga": "oligarki", "oligarki": "oligarki"]
Figure 1. Slang Word

to Indonesian disaster discourse; and a practical contribution by providing empirical evidence that supports the use of LSTM architectures in disaster-related public opinion monitoring systems.

This study applies an LSTM-based method to categorize public causal perceptions related to the 2025 Sumatra flash flood and evaluates its performance against baseline classifiers, including SVM and Naïve Bayes. The dataset was collected from YouTube comments on disaster-related news videos and labeled using a semi-automated lexicon-based causal attribution approach validated through manual annotation by three independent annotators. This study contributes theoretically by systematically mapping public causal perceptions, methodologically by developing a replicable causal labeling framework, and practically by recommending an optimal model architecture for disaster-related public opinion monitoring systems. The findings are expected to support the formulation of more responsive disaster mitigation communication strategies aligned with Sustainable Development Goals (SDGs) 11 and 13.

II. RESEARCH METHOD

This study employs a quantitative, computational-experiment-based approach to classify public causal attributions regarding the 2025 Sumatra flash flood in YouTube comments, focusing on the distinction between human-factor and nature/prayer-factor attributions. The methodological pipeline is structured sequentially, covering data collection, preprocessing, labeling, manual annotation validation, feature extraction, modeling, training, and validation to ensure experimental reproducibility and methodological consistency.

A. Data Collection

Comment data were acquired through the YouTube Data API v3 using the commentThreads endpoint from ten news and public discussion videos related to the 2025 Sumatra flash flood. The video curation criteria included channel reputation, topical alignment, and engagement metrics. Comments were collected between November 26 and December 17, 2025, to capture the dynamics of causal attribution from the initial disaster phase through the post-disaster period. All collected data consisted only of publicly available comments, and no personally identifiable user information was stored or analyzed in this study. The raw dataset comprised 17,503 comments, indicating the scale of the machine learning dataset, and was stored in CSV format to ensure traceability and research reproducibility. To ensure full experimental traceability, the ten YouTube video sources used for comment extraction are documented in Table 1.

B. Data Preprocessing

Data preprocessing was conducted to prepare YouTube comment texts for causal attribution classification modeling. Given the colloquial, unstructured, and linguistically heterogeneous characteristics of social media text, systematic preprocessing procedures were applied to reduce linguistic noise while preserving semantically and causally relevant information within the disaster context [26].

1) Cleaning and Non-Essential Character Removal

Initial cleaning procedures eliminated extraneous elements such as punctuation marks, numeric values, hyperlinks, user mentions, hashtag symbols, and non-ASCII encodings that do not contribute to

causal attribution analysis [27]. This process standardized the text and facilitated subsequent steps, including normalization, stemming, and lexical labeling. In addition, empty and duplicate comments were removed based on similarity in the cleaned text to prevent distributional bias caused by the dominance of identical opinions.

2) Case Folding

Text normalization through case folding transformed all characters to lowercase. This step aimed to reduce token redundancy caused by capitalization differences that carry no semantic meaning. As a result, identical words were not represented as distinct entities solely because of letter-case variation, thereby maintaining a more controlled feature space [28].

3) Informal Language Normalization

Informal language normalization was performed to standardize non-standard word variants commonly found in YouTube comments into their formal Indonesian equivalents, reducing feature-space expansion and semantic distortion [29]. This process used a contextually constructed slang dictionary tailored to disaster and political discourse, covering common abbreviations, informal spelling variants, and symbolic terms representing regions, actors, and institutions. The slang dictionary used in this normalization process is shown in Figure 1. Each token was replaced with its standardized form before further processing, ensuring consistency in word representation and reducing feature fragmentation. Normalization was applied before stemming so that morphological reduction operated on standardized word forms.

4) Tokenization

Token segmentation used whitespace delimitation, partitioning text at space boundaries so that each word was represented as a single token. This approach aligns with normalization and stemming procedures and provides the basis for numerical representation in causal labeling and feature extraction [30].

5) Stopword Removal

Functional word filtering removed high-frequency terms with limited causal semantic value, such as conjunctions and grammatical particles. An Indonesian stopword list was applied selectively to avoid removing words that carried important information about disasters, social actors, or institutions. This selective strategy maintained a balance between reducing linguistic noise and preserving semantic meaning [31], [32].

6) Stemming

Morphological reduction through Indonesian stemming transformed words to their base forms to reduce morphological variation and simplify textual representation. The resulting preprocessed text, which had been cleaned, normalized, tokenized, and stemmed, was then used for the causal labeling and modeling stages. Although sequence-oriented models such as LSTM can capture complex morphological patterns, stemming and stopword removal were retained to ensure methodological consistency across classifiers and to reduce lexical sparsity in highly informal and diverse Indonesian online texts. In informal Indonesian social media text, where morphological variation caused by typos, abbreviations, and non-standard spellings is pervasive, stemming mainly functions as a normalization mechanism that reduces out-of-vocabulary tokens rather than causing semantic loss. This approach is consistent with prior studies showing accuracy improvements when stemming is applied to non-formal Indonesian text classification tasks [33]. This strategy was adopted to stabilize token distributions without removing core causal meanings relevant to disaster discourse.

C. Labeling

Labeling in this study adopts a disaster-causal-attribution approach that emphasizes causal reasoning rather than emotional polarity. The labeling scheme consists of two causal classes: human factor, representing anthropogenic narratives related to human activities, structural negligence, and institutional policy failures; and nature/prayer factor, representing theological and naturalistic narratives that frame disasters as external events, including destiny, natural phenomena, and religious expressions of empathy.

1) Lexicon-Based Labeling with Causal Attribution Scoring

Label assignment was conducted automatically through a lexicon-driven scoring method using a purpose-built keyword ontology tailored to disaster sociological frameworks. Each comment was represented as a set of tokens generated through preprocessing [34]. Two lexicon sets were constructed to represent anthropogenic and theological attribution indicators. Category determination relied on aggregate keyword frequency computation within each comment. The anthropogenic attribution score

$$S_{antro}(C) = \sum_{\omega \in C} count(\omega \in L_{antro}) \quad (1)$$

$$S_{teo}(C) = \sum_{\omega \in C} count(\omega \in L_{teo}) \quad (2)$$

$$Y(C) = \begin{cases} \text{Human factor}, & \text{if } S_{antro} \geq S_{teo} \text{ and } S_{antro} > 0 \\ \text{Nature/prayer factor}, & \text{if } S_{teo} > S_{antro} \\ \text{Neutral}, & \text{otherwise} \end{cases} \quad (3)$$

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (4)$$

$S_{antro}\ (C)$ and theological attribution score $S_{teo}\ (C)$, computed using Equations (1) and (2), quantify the extent to which a comment attributes the disaster to human-related causes or to nature- and prayer-based narratives, respectively. Where ω represents a token in comment C, and L_{antro} and L_{teo} represent the keyword sets for each causal category. Final label assignment Y(C) followed the decision function defined in Equation (3).

When both scores were equal, the human factor label was assigned as the priority. Equation (3) formalizes the causal attribution decision rule by prioritizing anthropogenic attribution as actionable information for disaster risk mitigation, while theological attributions are treated as non-actionable normative responses. This approach is relevant because fate-based attributions may weaken social resilience [35] and are often used in elite discourse to deflect responsibility for environmental governance and public policy failures [36].

The anthropogenic keyword set includes terms related to governance failure, institutional negligence, deforestation, and policy accountability, such as *korupsi*, *lalai*, *deforestasi*, *penebangan*, *pemerintah*, and *kebijakan*. The theological keyword set includes religious expressions, fatalistic phrasing, divine will attribution, and prayer-related vocabulary, such as *musibah*, *takdir*, *doa*, *cobaan*, *bencana alam*, and *Allah berkehendak*. Although lexicon-based scoring enables large-scale causal labeling, it has inherent limitations, including bias toward dominant lexical cues and the risk of circularity, where machine learning models may replicate keyword associations rather than capture implicit causal reasoning [37]. Therefore, this study does not assume that model performance reflects deep causal understanding. Instead, it evaluates the extent to which sequential models can generalize causal attribution patterns beyond explicit keyword matching through manual annotation validation and cross-model comparison.

2) Manual Annotation Validation

Human annotation verification was conducted to ensure the dependability of automated labeling and its alignment with human interpretation, given the potential for linguistic ambiguity and implicit context in social media comments [38]. The quality assessment used a random sample comprising 10% of the complete dataset (n = 1,230 comments) to maintain representativeness. The selected samples were independently evaluated by three human coders who had no access to the automated labels under a blind annotation scheme to minimize confirmation bias. Cohen's Kappa coefficient κ, expressed in (4), was used to quantify inter-annotator agreement. Equation (4) measures agreement beyond chance and was used in this study to validate the reliability of the causal labeling scheme before model training. Where P_o represents observed agreement and P_e represents agreement expected by chance. The interpretation of κ values followed the COOPA standard, summarized in Table 2. This standard was used to assess the suitability of the causal labeling scheme before modeling [39]. Table 2 presents the COOPA standard interpretation ranges, where values above 0.81 indicate almost perfect agreement and are suitable for reliable classification tasks. The κ value of 0.81 obtained in this study confirms that the labeling scheme has sufficient reliability for subsequent modeling.

D. Feature Extraction

The preprocessed textual data were converted into numerical feature representations tailored to the requirements of each classification model. The Naïve Bayes and SVM models used TF-IDF representations with unigram and bigram schemes to capture relevant lexical contributions while suppressing common terms, with feature dimensionality constrained for computational stability [40]. LSTM used sequential representations generated from tokenized and length-standardized text, in which words were

TABLE 2
INTERPRETATION OF COHEN'S KAPPA VALUES (COOPA STANDARD)

Kappa Value	Degree
< 0.20	Poor
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantive
0.81 – 1.00	Almost Perfect

TABLE 3
LSTM ARCHITECTURE CONFIGURATION

Component	Specification
Implementation Framework	TensorFlow / Keras (Python 3.12)
Optimizer	Adam (lr = 0.001)
Loss Function	Categorical Crossentropy
Batch Size	64
Max Epochs (train-test split)	15
Max Epochs (cross-validation)	5
Early Stopping	val_accuracy, patience = 3
Train-Test Split Ratio	80:20 (random_seed = 42)
Input	(batch_size, 100)
Vocabulary Size	5.000
Sequence Length	100
Embedding Dimension	100
Dropout Regularization	0,2 (spatial, dropout, recurrent)
LSTM Units	100
Output Activation	Softmax
Number of Classes	2

mapped to numerical indices and then projected through an embedding layer [41], [42]. This method supports the learning of semantic associations and temporal relationships without manual feature engineering. Pre-existing word vectors were excluded to prevent domain incompatibility and allow the model to learn semantic representations specific to the disaster discourse context directly from the data.

E. Model Architecture

LSTM serves as the principal architecture for causal classification within Indonesian social media discourse in this study, while conventional machine learning models serve as baselines. The emphasis is placed on LSTM's ability to capture sequential dependencies and semantic context [43]. The LSTM model processes text as fixed-length token sequences of 100 tokens obtained through tokenization and padding. Inputs are represented as matrices with the shape (batch_size, 100), with the vocabulary size limited to 5,000 words, and are projected through an end-to-end learned embedding layer without pre-trained embeddings. Regularization is applied through spatial dropout, dropout, and recurrent dropout to reduce overfitting. LSTM units can capture dependencies across local and extended temporal contexts, while a Softmax activation function is used in the output layer to generate probabilistic predictions for the binary classes. This architecture balances representational capacity and training stability for Indonesian social media data. Detailed architectural configurations are presented in Table 3. As shown in Table 3, the architecture uses moderate regularization with a 0.2 dropout rate and a constrained vocabulary size of 5,000 words to prevent overfitting while maintaining sufficient representational capacity for informal Indonesian text.

F. Model Training and Validation

Training and validation procedures evaluated the model's generalization capacity in causal attribution classification using a controlled and uniform evaluation protocol. Category proportions were continuously monitored during the experiments and did not show extreme imbalance. Therefore, class weighting was not applied to preserve the natural distribution of public discourse [44]. Initial data partitioning followed an 80:20 training-testing ratio with a fixed random seed, followed by stratified five-fold cross-validation to maintain class proportions across splits.

TF-IDF–based feature vectors were used to train the Naïve Bayes and SVM models, with no class weighting or data resampling applied. Sequential token inputs were used to train the LSTM model, also without class weighting. The maximum number of epochs was set at 15 for the train and test split and five for cross-validation. Validation-triggered early stopping was applied to prevent overfitting. Classi-

TABLE 4
MODEL METRICS COMPARISON

Model	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	0.7504	0.7680	0.7504	0.7460
SVM	0.8283	0.8372	0.8283	0.8261
LSTM	0.9817	0.9817	0.9817	0.9817

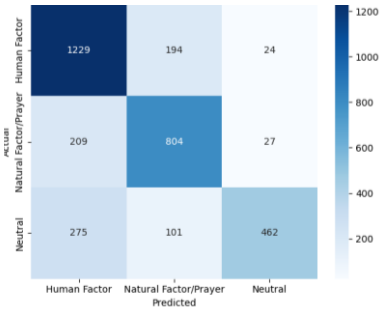

Figure 2. Confusion Matrix Naïve Bayes

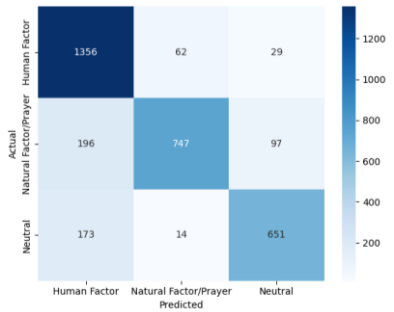

Figure 3. Confusion Matrix SVM

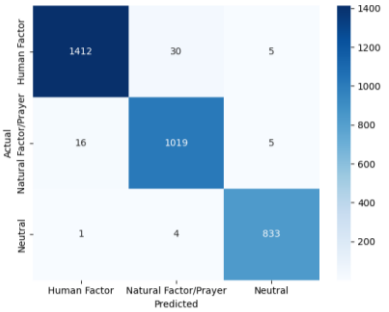

Figure 4. Confusion Matrix LSTM

fication effectiveness was assessed using accuracy, precision, recall, and F1-score indicators. This assessment was supplemented by misclassification analysis through confusion matrices and discriminative capacity evaluation using ROC-AUC as the benchmark metric.

III. RESULTS AND DISCUSSION

A. Result

This research began with 17,503 unprocessed comments harvested from YouTube, representing online public discourse related to the disaster event. After basic cleaning procedures to remove duplicate, empty, and non-informative comments, 16,623 comments were retained for the causal labeling process. Automated labeling initially produced three categories: human factor, nature/prayer factor, and neutral. A total of 4,324 neutral comments were excluded from the dataset to maintain the rigor of the binary classification formulation and avoid conceptual ambiguity during the modeling stage. The final dataset used in all experiments comprised 12,299 comments, including 7,183 comments labeled as human factor and 5,116 comments labeled as nature/prayer factor. This dataset formed the basis for all subsequent modeling and evaluation processes without qualitative interpretation of discourse meaning.

Automated label quality was verified through human annotation procedures, yielding $\kappa = 0.81$, which indicates strong inter-annotator agreement. This result confirms that the labels have sufficient interpretative stability, allowing model evaluation outcomes to be attributed to algorithmic performance rather than annotation uncertainty.

Based on the finalized dataset, modeling and training were conducted using three classification approaches: Naïve Bayes, SVM, and LSTM. All models used the same train and test split configuration to ensure a consistent basis for quantitative performance comparison.

As shown in Table 4, model performance was assessed using standard classification metrics under a consistent train-test split. The experimental results indicate accuracies of 75.04% for Naïve Bayes, 82.83% for SVM, and 98.17% for the LSTM model, with LSTM showing the strongest overall performance. The consistent superiority of LSTM across all evaluation metrics reflects a progressive improvement from Naïve Bayes to SVM, followed by a substantial gain with LSTM. This pattern suggests that the observed performance advantage is not solely attributable to increased model complexity but also to the suitability of sequential representations for capturing the linguistic characteristics of disaster-related comments.

Classification error patterns are illustrated through confusion matrices in Figures 2-4. In these matrices, diagonal cells represent correct classifications, namely true positives and true negatives, while off-diagonal cells indicate misclassification errors. Higher diagonal values relative to off-diagonal values indicate better classification performance. Naïve Bayes exhibits a relatively high misclassification rate for the nature/prayer factor class, reflecting the limitations of the feature independence assumption in capturing implicit causal narratives. SVM reduces some of these misclassifications but still shows bias toward the dominant class. Compared with the other models, LSTM shows a well-balanced distribution of classification errors, with minimal false positives and false negatives for each causal class.

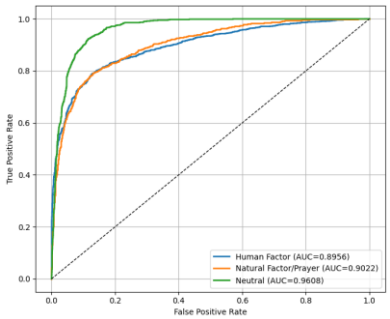

Figure 5. ROC-AUC Naïve Bayes

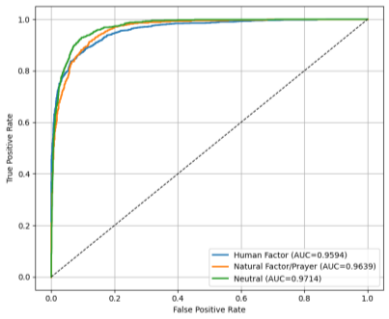

Figure 6. ROC-AUC SVM

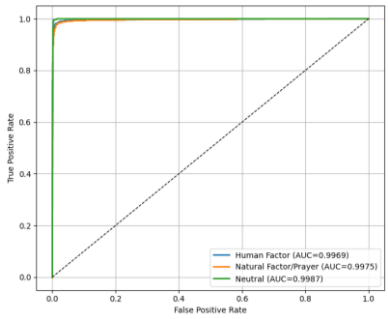

Figure 7. ROC-AUC LSTM

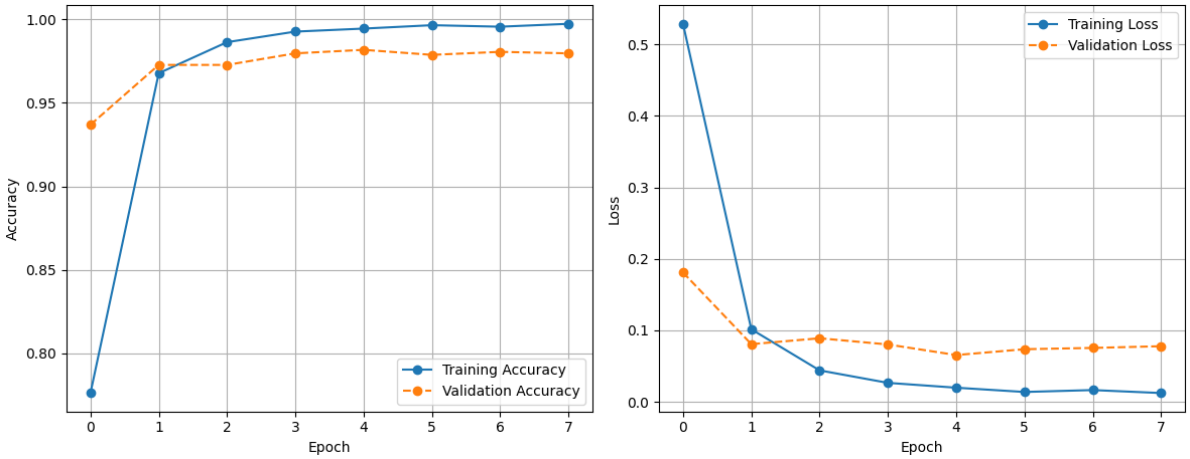

Figure 8. Accuracy and Loss Evolution of the LSTM Model During Training and Validation

Model discrimination capability was further evaluated using ROC–AUC curves, as shown in Figures 5-7. In ROC curves, the y-axis represents the true positive rate and the x-axis represents the false positive rate. A curve closer to the top-left corner indicates superior discrimination capability, with the AUC value quantifying the area under each curve. The ROC analysis shows that the LSTM model maintains near-perfect discrimination performance, as indicated by an AUC score of 0.99. SVM obtained an AUC of 0.96, while Naïve Bayes reached approximately 0.90, reinforcing the limitations of simple probabilistic models in distinguishing causal classes with high lexical overlap.

Figure 8 presents the evolution of the LNN model's training and validation performance across multiple epochs. Accuracy and loss values are plotted against epoch indices, revealing smooth convergence patterns. The absence of divergence between the curves indicates effective generalization during model training. The curves also show rapid convergence in the early epochs and stable validation performance without significant divergence, suggesting the absence of overfitting and strong generalization capability.

Model stability was assessed using stratified 5-fold cross-validation, with per-fold accuracy results summarized in Table 5. Table 5 presents the per-fold accuracy values across five stratified cross-validation iterations, demonstrating the consistency of model performance across different data partitions. LSTM achieved an average accuracy of 97.76% with a very low standard deviation (±0.25%), indicating substantially higher stability than SVM (82.93% ± 0.80%) and Naïve Bayes (74.85% ± 1.34%). This stability is further visualized in Figure 9, which shows the narrowest accuracy distribution for LSTM, indicating consistent performance across different training subsets. These results strengthen the argument that LSTM's superiority does not depend on a particular data split but instead reflects robust generalization across the analyzed discourse structure.

B. Discussion

These findings indicate that causal attribution in disaster-related public discourse is not merely a lexical phenomenon, but a form of structured social reasoning articulated through narrative sequences that encode responsibility, justification, and moral evaluation. Instead of expressing isolated opinions, users

TABLE 5
COMPARATIVE EVALUATION OF MODEL PERFORMANCE BASED ON 5-FOLD CROSS-VALIDATION

Validation Fold	Naïve Bayes	SVM	LSTM
1	0.7293	0.8232	0.9789
2	0.7663	0.8400	0.9780
3	0.7504	0.8277	0.9771
4	0.7518	0.8348	0.9735
5	0.7446	0.8210	0.9801

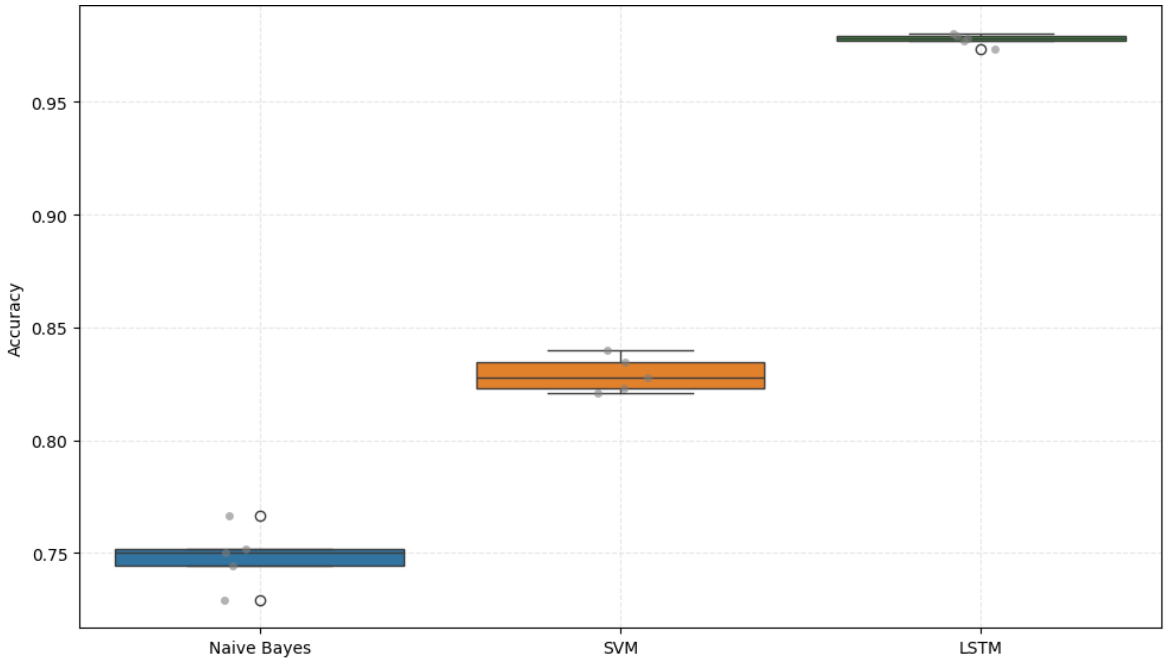

Figure 9. Model Stability Comparison - 5-Fold Cross Validation

construct implicit causal chains that link actors, actions, and consequences across sentences. This observation underscores that disaster discourse in digital environments operates as a narrative process in which meaning emerges from contextual continuity rather than from token-level sentiment signals alone.

The 98.17% accuracy achieved by LSTM in this study surpasses the performance reported in comparable disaster text classification research. Maharani et al. [11] reported 90.31% accuracy using SVM with SMOTE optimization on forest fire sentiment data, while Khamidah et al. [15] achieved only 77% accuracy on sarcastic flood-related texts using deep learning approaches. The stronger performance in the present study can be attributed to two factors: the causal attribution framing reduces classification ambiguity compared with sentiment polarity, and the LSTM architecture effectively captures sequential dependencies that encode implicit reasoning in Indonesian social media discourse. Abdurahmonov et al. [22] achieved 94.33% validation accuracy in extreme-weather sentiment analysis using LSTM, which aligns with the performance trajectory observed in this study. However, the 3.84 percentage point improvement in the present study suggests that causal attribution classification may present a more learnable task structure than sentiment polarity when discourse contains explicit responsibility markers.

The unusually high classification accuracy requires further contextual justification. Three factors collectively explain this performance level. First, the task is binary rather than multi-class, which structurally reduces classification complexity compared with multi-label sentiment polarity tasks. Second, the two target categories, anthropogenic attribution and theological or naturalistic attribution, are linguistically distinct in Indonesian disaster discourse. Governance critique vocabulary and religious empathy vocabulary have limited lexical overlap, which reduces decision boundary ambiguity for the model. Third, the consistency of LSTM performance across all five stratified cross-validation folds (range: 97.35%–98.01%, $\sigma = \pm 0.25\%$) confirms that the result is not an artifact of a favorable data partition. Collectively, these factors indicate that the observed accuracy reflects a genuinely learnable task structure rather than data leakage or labeling circularity, while acknowledging that performance on out-of-domain or temporally distant disaster datasets may differ. The Cohen's kappa value of 0.81 obtained in this study indicates almost perfect agreement according to the Landis and Koch interpretation scale,

exceeding the typical threshold of 0.60 considered acceptable for text annotation tasks. This reliability metric is comparable to or exceeds values reported in prior causal extraction research [38], supporting the validity of the labeling framework for replication in future studies.

The interpretation is corroborated by the LSTM model's sustained performance advantages throughout the experimental analysis. Its advantage should not be understood only as a technical outcome of sequential modeling, but also as empirical evidence that public causal reasoning is context-dependent and temporally ordered. Prior studies in disaster-related sentiment analysis have largely emphasized polarity detection and lexical salience. However, the present findings demonstrate that such approaches are insufficient to capture how responsibility and blame are discursively constructed. In contrast, LSTM can capture how causal attribution unfolds progressively within user-generated narratives, positioning sequential modeling as a more appropriate analytical approach for this form of public reasoning.

From a theoretical perspective, this study reframes disaster-related sentiment analysis as a problem of causal reasoning rather than emotional polarity. By shifting the analytical focus from affective evaluation to causal attribution, the findings extend computational approaches toward a discourse-oriented understanding of public sense-making. This reframing positions causal attribution not as an auxiliary annotation layer, but as a core analytical construct for interpreting disaster-related communication. Such a theoretical shift is particularly relevant in disaster studies, where public discourse often functions less as emotional expression and more as a site of moral judgment and political negotiation.

The attribution categories employed in this study should be interpreted strictly as analytical labels rather than normative classifications. Narratives categorized under the nature/prayer factor do not only reflect fatalistic attitudes. They also often function as coping mechanisms, moral framing devices, or discursive strategies that depoliticize responsibility. Conversely, human factor narratives operate as bottom-up articulations of accountability, through which citizens negotiate blame, trust, and institutional legitimacy through language. These patterns suggest that causal attribution in disaster discourse constitutes a socially embedded process of meaning construction rather than a simple evaluative response.

At the same time, the findings reveal epistemic ambiguity within public discourse. Many comments combine religious expression, empathy, and structural critique, occupying a liminal space between theological acceptance and political accountability. Therefore, the adoption of a binary attribution framework represents an analytical abstraction rather than a claim of discursive exclusivity. Acknowledging this boundary strengthens the interpretive validity of the study and situates its findings within the inherent complexity of public discourse practices.

Qualitative inspection of LSTM misclassification patterns reveals interpretable error typologies that reflect the inherent ambiguity of causal discourse rather than arbitrary model failure. The dominant error direction involved human factor comments misclassified as nature/prayer factor (n=30). These comments mainly embedded structural critique within religious framing, for instance, expressions that combined theological resignation with implicit blame directed at governance actors or institutional negligence. Such comments contain lexical signals from both causal categories at the same time, creating decision boundary ambiguity that even sequential modeling cannot fully resolve. Conversely, nature/prayer factor comments misclassified as human factor (n=16) frequently contained disaster-related institutional terms, such as references to government agencies, infrastructure, or policy actors, used in empathetic or intercessory contexts rather than accusatory ones. The relatively low and asymmetric error counts in both directions suggest that the remaining misclassifications are linguistically motivated rather than systematic. They also reflect genuine discourse ambiguity at the boundary between theological and anthropogenic attribution. This pattern supports the interpretation that LSTM generalizes beyond pure keyword matching, while acknowledging that the influence of lexicon-derived training labels on model behavior cannot be entirely discounted.

The high classification performance achieved by the LSTM model warrants critical reflection. Rather than indicating that the model captures objective causal truth, the results demonstrate its capacity to learn dominant discursive patterns shaped by media framing, lexical repetition, and narrative reinforcement. Consequently, the model captures how causality is articulated in public discourse, not how causality objectively exists. This distinction is essential to prevent the overinterpretation of predictive accuracy as epistemic understanding.

From a policy and risk communication perspective, these findings carry important strategic implications. Monitoring shifts in causal attribution patterns may support the early detection of delegitimization narratives, erosion of institutional trust, and emerging blame dynamics following disasters. At the same

time, such systems entail ethical risks, including potential misuse for discourse control or the suppression of public critique. Recognizing this dual-use character is essential to ensure that computational discourse analysis contributes to accountability rather than undermining democratic deliberation.

Beyond the Indonesian context, the findings offer broader insights into disaster discourse in digitally mediated societies characterized by high religiosity and moral framing. While much of the global literature prioritizes the detection of emotional polarity or urgency, this study demonstrates that causal reasoning is a critical but underexplored dimension of public response. Accordingly, causal attribution analysis provides a transferable analytical lens for examining disaster discourse across socio-cultural settings where faith, governance, and environmental risk intersect. These implications should be interpreted within the scope of digitally mediated public discourse rather than as universal patterns of human causal reasoning.

IV. Conclusion

This study demonstrates that analyzing public causal attribution toward disasters requires an approach that goes beyond polarity-based sentiment classification. The structure of discourse in YouTube comments is inherently contextual and sequential, making sequence-based modeling more suitable for capturing the implicit cause-and-effect reasoning embedded in disaster-related discussions. The causal attribution approach employed in this study enables a more substantive mapping of public perception than analyses that focus only on emotional expression.

Comparative evaluations reveal that LSTM achieves superior performance relative to Naïve Bayes and SVM in every experimental configuration. The superiority of LSTM is reflected not only in higher accuracy but also in greater performance stability and a more balanced error distribution. These findings indicate that LSTM's capability to model long-term dependencies and linguistic context provides a significant advantage in classifying causal attribution within Indonesian-language social media texts.

The study contributes to the field of disaster text analysis through the development of a rigorously structured and empirically validated causal attribution approach. The results reinforce the role of sequential approaches in understanding public perceptions of disasters and highlight their potential to support more responsive public opinion monitoring systems for disaster risk communication and mitigation based on social data.

Declaration of AI and AI assisted technologies in the writing process

During the preparation of this work, the authors used Claude (Anthropic) in order to assist with cross-language proofreading and to verify consistency in terminology and phrasing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT Authorship Contribution Statement

Kunti Najma Jalia: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, and Writing – review & editing. **Adi Suwondo**: Conceptualization, Methodology, Supervision, Validation, and Writing – review & editing. **Hidayatus Sibyan**: Methodology, Investigation, Validation, Supervision, and Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Maruli, "TRAGEDI SUMATRA 2025 — 'Banjir Neraka Sumatra Pecah Rekor: Korban Lewat 700 Jiwa, 1 Juta Warga Mengungsi, Negara Dikepung Krisis!,'" Gakorpan News, 2025. [Online]. Available: https://www.gakorpan.com/tragedi-sumatra-2025-banjir-neraka-sumatra-pecah-rekor-korban-lewat-700-jiwa-1-juta-warga-mengungsi-negara-dikepung-krisis

[2] Matius Alfons Hutajulu, "Korban Tewas Bencana Sumatera Capai 1.006 Orang, 217 Masih Hilang'" DetikNews, 2025. [Online]. Available: https://news.detik.com/berita/d-8258300/korban-tewas-bencana-sumatera-capai-1-006-orang-217-masih-hilang

[3] Agungnoe, "Bencana Banjir Bandang Sumatra, Pakar UGM Sebut Akibat Kerusakan Ekosistem Hutan di Hulu DAS," Universitas Gadjah Mada, 2025. [Online]. Available: https://ugm.ac.id/id/berita/bencana-banjir-bandang-sumatra-pakar-ugm-sebut-akibat-kerusakan-ekosistem-hutan-di-hulu-das/

[4] J. Hladík, L. Herman, D. Snopková, and M. Konečný, "Spatio-temporal patterns of disaster impact and recovery in YouTube content," *Int. J. Digit. Earth*, vol. 17, no. 1, pp. 1–23, 2024.

[5] J. Tian and R. Zhang, "Moral judgments influence emotional responses and comment lengths through the moderating role of linguistic style matching," *Sci. Rep.*, vol. 15, no. 1, p. 20972, 2025.

[6] S. Kwon and A. Park, "Examining thematic and emotional differences across Twitter, Reddit, and YouTube: The case of COVID-19 vaccine side effects," *Comput. Hum. Behav.*, vol. 144, p. 107734, 2023.

[7] K. L. Po, A. J. R. Sy, and R. D. G. Jamora, "Evaluating YouTube as a source of information on hemifacial spasm," *Clin. Park. Relat. Disord.*, vol. 12, p. 100311, 2025.

[8] Kompas.com, "Pakar ITB Ungkap Penyebab Banjir Sumatera, Termasuk Hilangnya Resapan," Youtube, 2025. [Online]. Available: https://www.youtube.com/watch?v=tfJC0bn9gHs

[9] Narasi, "Meliput Banjir Sumatera: Air Masih Tinggi, Jalan Terputus | Mata Najwa," *Mata Najwa*, 2025. [Online]. Available: https://www.youtube.com/watch?v=ouR11bOMzVM

[10] B. Ilyas and A. Sharifi, "A systematic review of social media-based sentiment analysis in disaster risk management," *Int. J. Disaster Risk Reduct.*, vol. 123, p. 105487, 2025.

[11] W. Maharani, H. Daud, N. Muhammad, and E. A. Kadir, "Leveraging Social Media Data for Forest Fires Sentiment Classification: A Data-Driven Method," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 10, no. 3, pp. 392–407, 2024.

[12] P. M. Lavanya and E. Sasikala, "Auto capture on drug text detection in social media through NLP from the heterogeneous data," *Meas. Sens.*, vol. 24, p. 100550, 2022.

[13] N. Arlim *et al.*, "Dictionary-based extraction of hyperbole and swear words for sarcasm detection in Indonesian Tweets," *Int. J. Inf. Technol.*, vol. 17, no. 5, pp. 2671–2678, 2024.

[14] A. F. Hidayatullah, R. A. Apong, D. T. C. Lai, and A. Qazi, "Pre-trained language model for code-mixed text in Indonesian, Javanese, and English using transformer," *Soc. Netw. Anal. Min.*, vol. 15, no. 1, p. 30, 2025.

[15] N. Khamidah, K. A. Notodiputro, and S. D. Oktarina, "Sentiment Analysis of Imbalanced Sarcastic Flood Disaster Texts Using Deep Learning Models," *Int. Res. J. Innov. Eng. Technol.*, vol. 08, no. 11, pp. 150–158, 2024.

[16] C. Mallikarjuna and S. Sivanesan, "Tweet question classification for enhancing Tweet Question Answering System," *Nat. Lang. Process. J.*, vol. 10, p. 100130, 2025.

[17] S. Christina and D. Ronaldo, "Studi Literatur Sistematis Terhadap Pengembangan Leksikon Sentiment," *J. ELTIKOM*, vol. 4, no. 2, pp. 121–131, 2020.

[18] B. Ghojogh and A. Ghodsi, "Recurrent Neural Networks and Long Short-Term Memory Networks: Tutorial and Survey," Apr. 22, 2023, *arXiv*: arXiv:2304.11461.

[19] Y. Yu, X. Si, C. Hu, and J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, 2019.

[20] J. García Cabello and S. Carbó-García, "LSTM new gate for computing the efficiency on inputdata," *Knowl.-Based Syst.*, vol. 322, p. 113622, 2025.

[21] E. Kim, W. Luo, and H. Jho, "Perception and argumentation in the LK-99 superconductivity controversy: a sentiment and argument mining analysis," *Sci. Rep.*, vol. 15, no. 1, p. 13254, 2025.

[22] T. Abdurahmonov, M. N. T. Abiyyu, D. N. Khayat, and M. A. Heryanto, "Aspects-Based Sentiment Analysis of Extreme Weather on Twitter Using Long Short-Term Memory," *J. Inform. Web Eng.*, vol. 4, no. 2, pp. 430–443, 2025.

[23] Y. Ji, W. Tao, and C. Wan, "A Systematic Review of Attribution Theory Applied to Crisis Events in Communication Journals: Integration and Advancing Insights," Journal. Mass Commun. Q., 2025.

[24] W. Kassa and R. Lavin, "Assessment of Blame and Responsibility Through Social Media in Disaster Recovery in the Case of #FlintWaterCrisis," Front. Commun., vol. 3, p. 45, 2018.

[25] A. Yusima, "Digital Traces of Collective Trauma in the 2025 Aceh Climate Disaster," Renai, vol. 12, no. 1, pp. 1–8, 2026.

[26] M. Siino, I. Tinnirello, and M. La Cascia, "Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers," *Inf. Syst.*, vol. 121, p. 102342, 2024.

[27] C. P. Chai, "Comparison of text preprocessing methods," Nat. Lang. Eng., vol. 29, pp. 509–553, 2022.

[28] L. Zhu and D. Luo, "A Novel Efficient and Effective Preprocessing Algorithm for Text Classification," *J. Comput. Commun.*, vol. 11, no. 03, pp. 1–14, 2023.

[29] R. Patil, S. Boit, V. Gudivada, and J. Nandigam, "A Survey of Text Representation and Embedding Techniques in NLP," *IEEE Access*, vol. 11, pp. 36120–36146, 2023.

[30] G. Erasmo Ndomba, M. Edmund Mswahili, and Y.-S. Jeong, "Tokenizers for African Languages," *IEEE Access*, vol. 13, pp. 1046–1054, 2025.

[31] K. Madatov, S. Bekchanov, and J. Vičič, "Dataset of stopwords extracted from Uzbek texts," *Data Brief*, vol. 43, p. 108351, 2022.

[32] H. T. Y. Achsan, H. Suhartanto, W. C. Wibowo, D. A. Dewi, and K. Ismed, "Automatic Extraction of Indonesian Stopwords," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 2, 2023.

[33] Rianto, A. B. Mutiara, E. P. Wibowo, and P. I. Santosa, "Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation," *J. Big Data*, vol. 8, no. 1, p. 26, 2021.

[34] T. W. G. Cuizon and H. S. Alar, "Lexicon-based Sentence Emotion Detection Utilizing Polarity-Intensity Unit Circle Mapping and Scoring Algorithm," *Procedia Comput. Sci.*, vol. 212, pp. 161–170, 2022.

[35] F. E. L. Otto and E. Raju, "Harbingers of decades of unnatural disasters," *Commun. Earth Environ.*, vol. 4, no. 1, p. 280, 2023.

[36] D. Marks and I. G. Baird, "The urban political ecology of worsening flooding in Phnom Penh, Cambodia: Neopatrimonialism, displacement, and uneven harm," *Int. J. Disaster Risk Reduct.*, vol. 118, p. 105229, 2025.

[37] I. Cero, J. Luo, and J. M. Falligant, "Lexicon-Based Sentiment Analysis in Behavioral Research," *Perspect. Behav. Sci.*, vol. 47, no. 1, pp. 283–310, 2024.

[38] B. Drury, H. Gonçalo Oliveira, and A. De Andrade Lopes, "A survey of the extraction and applications of causal relations," *Nat. Lang. Eng.*, vol. 28, no. 3, pp. 361–400, 2022.

[39] K. S. Tan, Y.-C. Yeh, P. S. Adusumilli, and W. D. Travis, "Quantifying Interrater Agreement and Reliability Between Thoracic Pathologists: Paradoxical Behavior of Cohen's Kappa in the Presence of a High Prevalence of the Histopathologic Feature in Lung Cancer," *JTO Clin. Res. Rep.*, vol. 5, no. 1, p. 100618, 2024.

[40] Y. Ma, "Construction and Data Analysis of a New Media Content Popularity Prediction Model Based on Naive Bayes Algorithm," *Procedia Comput. Sci.*, vol. 261, pp. 294–302, 2025.

[41] B. A. Chandio, A. S. Imran, M. Bakhtyar, S. M. Daudpota, and J. Baber, "Attention-Based RU-BiLSTM Sentiment Analysis Model for Roman Urdu," *Appl. Sci.*, vol. 12, no. 7, p. 3641, 2022.

[42] J. Duan, P.-F. Zhang, R. Qiu, and Z. Huang, "Long short-term enhanced memory for sequential recommendation," *World Wide Web*, vol. 26, no. 2, pp. 561–583, 2023.

[43] A. Iqbal, A. Shahid, M. Roman, M. T. Afzal, and U. U. Hassan, "Optimising window size of semantic of classification model for identification of in-text citations based on context and intent," *PLOS ONE*, vol. 20, no. 3, p. e0309862, 2025.

[44] S. F. Taskiran, B. Turkoglu, E. Kaya, and T. Asuroglu, "A comprehensive evaluation of oversampling techniques for enhancing text classification performance," *Sci. Rep.*, vol. 15, no. 1, p. 21631, 2025.