Research Trends in Indonesian Sentiment Analysis Using Natural Language Processing: A Descriptive Review

Bima Satriatama[1*], Marhendra Putra[2],
[12]Department of System of Information, Bina Darma University, Palembang, Indonesia
*Correspondence: bima.satriatama16@gmail.com

ARTICLE HISTORY
Received: *[13/02/2026]* **Revised:** *[07/04/2026]* **Accepted:** *[12/05/2026]* **Published:** *[30/06/2026]*
Vol/No/Pages: *[Vol 1(No.2), 80-88]*

Abstract

Sentiment analysis is an application of Natural Language Processing (NLP) used to identify and interpret opinions, attitudes, and emotional tendencies in text. In the Indonesian context, the development of sentiment-analysis research is bound up with the growing use of social media, customer reviews, and digital platforms as data sources, and with advances in machine-learning methods and pretrained language models. This study aims to describe the tendencies in methodological approaches, the use of NLP models, application domains, and the direction of development of Indonesian-language sentiment analysis on the basis of relevant scholarly publications. The study is designed as a descriptive approach that groups the literature by year of publication, analytical method, language model, data source, research domain, and the development of interpretability approaches. A synthesis of the available references reveals several conceptual tendencies: the use of classical methods based on feature representation, growing attention to Indonesian transformer models such as IndoBERT and IndoBERTweet, the application of sentiment analysis to social and political issues and to digital services, and the expansion of research toward emotion classification, aspect-based sentiment analysis, and explainable artificial intelligence. In addition, data quality, class imbalance, variation in informal language, and limited interpretability remain relevant methodological concerns. The study provides a descriptive framework for understanding the direction of Indonesian sentiment-analysis research. Because the number of publications and the results of bibliographic coding have not been counted systematically, the tendencies presented are not interpreted as a quantitative measure of the overall research landscape.

Keywords: sentiment analysis; Indonesian language; Natural Language Processing; IndoBERT; transformer; descriptive analysis

1. Introduction

The development of digital technology has increased the volume of text produced through social media, e-commerce platforms, public services, mobile applications, and various online communication channels. This text contains information about users' opinions, experiences, satisfaction, dissatisfaction, and emotional responses to a product, policy, event, or service. Yet the volume and diversity of digital text make manual analysis difficult, especially when information must be processed at scale and updated continuously. In this context, Natural Language Processing (NLP) offers a range of methods for processing human language computationally, including sentiment classification.

Sentiment analysis is an NLP task that aims to identify the polarity or tendency of opinion in text—for example, positive, negative, and neutral. Over time, sentiment analysis has also come to involve the identification of emotion, the classification of opinion by particular aspects, and the interpretation of the factors that influence a prediction. Applying such methods allows organizations to obtain a picture of public responses, evaluate the user experience, and identify issues that require further attention.

Indonesian-language sentiment analysis has characteristics that require particular consideration. The use of informal language, abbreviations, non-standard words, code-mixing, spelling variation, irony, sarcasm, and local expressions can affect the interpretation of text. On social media, a single short message can contain several forms of expression that cannot always be interpreted through positive or negative words alone. The success of sentiment analysis is therefore determined not only by the algorithm but also by data quality, preprocessing strategy, language representation, and the accuracy of labeling.

In the early stages of sentiment-analysis development, classical machine-learning methods such as Naive Bayes, Support Vector Machine, and feature-based approaches using TF-IDF or Bag of Words were widely used because they are relatively simple and can serve as a basis for comparison. These approaches represent text through lexical

features and learn the relationship between features and sentiment labels. Although effective for a number of classification tasks, word-frequency-based representation is limited in capturing context, relationships among words, and changes in meaning that depend on sentence structure.

The development of pretrained language models has brought change to Indonesian-language processing. Through their work on IndoLEM and IndoBERT, Koto et al. provide an important basis for developing evaluation resources and pretrained language models for Indonesian NLP [1]. Transformer-based models allow a word's representation to be influenced by its context of use, potentially improving a system's ability to handle complex text. Subsequent developments include the use of IndoBERT and IndoBERTweet in various sentiment- and emotion-classification tasks [2], [3], [4].

The application of Indonesian sentiment analysis has also grown more diverse in domain. Research has used social-media data to examine issues of climate change, presidential elections, government performance, and responses to public programs. Mubaraq and Maharani, for example, study sentiment on climate change using IndoBERT [5], while Putri et al. examine sentiment in the context of Indonesia's 2024 presidential election [6]. In the e-commerce domain, Sumendap and Mahendra compare sentiment analysis of Shopee and Tokopedia customer reviews using the Google NLP API [7]. This variety of domains shows that sentiment analysis is used not only to test models but also to understand opinion in particular social and service contexts.

Beyond polarity classification, research has begun to broaden its attention to emotion classification, aspect-based sentiment analysis, and model interpretability. Ahmadian et al. discuss hybrid models for emotion classification and sentiment analysis in Indonesian [2]. The study by Anthera Mumtaz and Abdul Majid addresses aspect-based sentiment analysis with fine-tuning of IndoBERT [8], while Widiyatmoko et al. combine IndoBERT-LSTM with an explainable-AI approach using SHAP [9]. Delvian et al. also discuss the use of LIME to interpret sentiment classification on a number of public issues [10]. This direction shows a broadening of focus from merely producing labels to a more detailed understanding of the aspects and factors that influence a prediction.

Although research on Indonesian sentiment analysis continues to grow, a descriptive mapping is needed to see how approaches, models, data sources, and application domains relate to one another. Setiawan reviews sentiment-analysis applications in Indonesia over the 2023–2024 period [11], while Daviansyah et al. discuss the effectiveness of NLP models for Indonesian-text sentiment analysis in the public-service domain through a systematic review [12]. Zara et al. also present a meta-analysis of the performance of NLP methods for Indonesian-text sentiment analysis [13]. The presence of these studies shows the importance of distinguishing research focused on model development, application to particular domains, and the mapping of the research landscape.

Against this background, this study aims to describe the tendencies in Indonesian sentiment-analysis research on the basis of a collection of relevant publications. The focus of the analysis covers the development of computational approaches, the use of language models, data sources and application domains, the expansion of analytical tasks, and the methodological issues that arise in developing systems. The study addresses three questions: (1) what NLP approaches and models are discussed in Indonesian sentiment-analysis publications; (2) what data sources and application domains are reflected in the literature; and (3) what development directions and methodological challenges can be identified from these publications?

The contribution of this study is to provide a descriptive framework for organizing the Indonesian sentiment-analysis literature by the dimensions of method, model, domain, and development direction. This framework is intended to help researchers identify opportunities for further research while encouraging more transparent methodological reporting and more comparable results.

2. Method

Research Design

This study uses a descriptive approach, with scholarly publications on Indonesian sentiment analysis using NLP as its object of study. The descriptive approach was chosen to depict the characteristics and tendencies of the literature without conducting experimental testing of algorithms or claiming causal relationships among variables. The analysis is directed at the bibliographic and substantive characteristics of publications, including year of publication, computational method, language model, data source, application domain, type of analytical task, and aspects of evaluation. Through this approach, the development of research can be understood by grouping themes and comparing the characteristics of studies.

The initial material for preparing this article is the reference list provided for this study. Because that list is not yet accompanied by database-search results, selection records, or coding sheets, this manuscript is a draft descriptive analysis based on the available references, not a completed bibliometric study or quantitative mapping. The number of publications per year, the proportion of model use, and the frequency of themes are therefore not stated as measured results.

Scope of Literature

The literature of interest covers research on Indonesian-text sentiment analysis, the use of NLP models for sentiment or emotion classification, the development of pretrained language models, and the application of sentiment analysis to social media, customer reviews, digital applications, and public issues.

Publications relevant to the foundations of Indonesian NLP are also considered in order to explain the development of models and language resources. Koto et al., for example, are used to contextualize the development of IndoBERT and Indonesian evaluation resources [1]. Conversely, research on other languages or NLP research unrelated to sentiment analysis needs to be separated from the main corpus unless it has clear methodological relevance.

The range of publication years follows the available references, which span the period before and after 2020 and publications up to 2026. This range describes the initial scope of the research material, not the result of a comprehensive search of all publications issued in the same period.

Data Extraction and Classification

Each publication included in the final corpus should be recorded by the following variables:

Bibliographic information: author names, year of publication, title, journal or proceedings name, and DOI or publication link.

Computational approach: feature-based methods, classical machine learning, deep learning, transformers, or hybrid models.

Language model and representation: TF-IDF, Bag of Words, Naive Bayes, IndoBERT, IndoBERTweet, other language models, or API-based NLP services.

Data source: social media, customer reviews, application reviews, public documents, or other text sources.

Application domain: politics, public policy, environment, e-commerce, digital services, and other relevant domains.

Task type: polarity classification, emotion classification, aspect-based sentiment analysis, sentiment-related topic modeling, or explainable classification.

Evaluation and limitations: evaluation metrics, handling of class imbalance, data validation, interpretability, and the limitations reported by researchers.

The consistent use of categories is important for avoiding double counting. For example, one publication may use IndoBERT and also discuss a political issue. In the analysis, that publication can be recorded once under the model category and once under the domain category, but it must not be treated as two different publications.

Descriptive Analysis Procedure

The descriptive analysis is designed in four stages. First, bibliographic verification, checking the consistency of metadata, year of publication, title, authors, and publication identity. This check is also needed to identify possible mismatches between bibliographic information and the DOI.

Second, grouping publications—categorizing sources by year, method, language model, domain, and task type. Grouping is carried out with attention to information in the abstract and full text, not only to words appearing in the title.

Third, descriptive comparison—comparing the characteristics of publications to recognize shifts in methodological focus and the expansion of applications. If all data have been verified, this stage can include tables of the number of publications by year or category. However, counts should be performed only after the corpus and coding rules have been established.

Fourth, interpretation of tendencies—explaining the meaning of changes in approach, recurring challenges, and opportunities for further research. The interpretation needs to distinguish tendencies that are genuinely measured in the corpus from indications of development drawn from example studies.

Quality and Transparency

The quality of the analysis is maintained through consistency of categories, traceability of bibliographic information, and the separation of facts reported by sources from the author's interpretation. Publications that use different evaluation metrics must not be compared as though they used identical testing conditions. Likewise, accuracy scores reported from different datasets cannot be used directly to conclude that one model is always superior to another.

To make this study replicable, the final stage needs to record the databases used, search dates, keywords, inclusion and exclusion criteria, the number of initial results, the number of publications after deduplication, and the number of publications included in the analysis. If the analysis remains descriptive, the procedure need not be claimed as a systematic review when the systematic stages have not actually been carried out.

3. Results

General Characteristics of the Literature

The available references show that research on Indonesian sentiment analysis has several complementary foci. Some research develops or compares computational methods, some tests pretrained language models on particular data, and some directs analysis toward social issues or practical needs. In addition, there are studies that seek to summarize the development of applications and evaluate NLP methods within a broader scope [11], [12], [13].

Conceptually, the literature can be grouped into four main clusters. First, foundational research related to Indonesian language resources and models, including IndoLEM and IndoBERT [1]. Second, sentiment-classification research using classical methods, hybrid models, or transformers. Third, applied research that uses sentiment to understand public issues, politics, customer reviews, and digital services. Fourth, developmental research that extends analysis to emotion classification, aspect-based analysis, the handling of class imbalance, and model interpretability.

This grouping is a descriptive framework drawn from the available references. There is as yet no basis for stating the proportion of each cluster or concluding that one cluster dominates the whole of Indonesian scholarly publication.

Development of Computational Approaches

One important dimension in the development of sentiment analysis is the change in text representation and learning approach. Feature-based methods such as TF-IDF and Bag of Words map text into numerical representations based on word occurrence. These representations are then used by classification algorithms to predict sentiment categories. The strength of this approach lies in its simplicity of implementation, relatively affordable computational needs, and ease of use as a baseline.

The study by Ansori et al. specifically addresses a baseline for Indonesian Twitter sentiment analysis using TF-IDF and classical machine-learning methods with data-quality control [14]. Wibawa and Pramartha, meanwhile, study sentiment analysis of Balinese-language text using Multinomial Naive Bayes with TF-IDF and Bag of Words [15]. That Balinese-language study is not direct evidence of method performance on Indonesian, but it can provide methodological context for analyzing text in a regional language and the challenges of processing a language with limited resources.

Subsequent development is evident in the use of contextual representations and transformer-based models. IndoBERT provides a pretrained language model designed for Indonesian that supports a variety of NLP tasks [1]. Lin and Nuha discuss a hybrid deep-learning strategy for sentiment analysis of Indonesian datasets [3], while Ahmadian et al. study hybrid models for emotion classification and sentiment analysis in Indonesian [2].

In applied research, IndoBERT has been used on social-media data and public issues. Mubaraq and Maharani examine sentiment on climate change [5], while Putri et al. apply IndoBERT to sentiment analysis related to Indonesia's 2024 presidential election [6]. These studies show that pretrained language models have become an important approach in the study of Indonesian sentiment analysis.

However, the change in approach does not mean that classical methods have lost relevance. Classical baselines are still needed to assess whether an increase in model complexity yields a meaningful benefit. A fair comparison must use comparable data splits, label definitions, preprocessing procedures, and evaluation metrics.

The Use of Indonesian Pretrained Language Models

IndoBERT and IndoBERTweet are important examples in the development of models adapted to the characteristics of Indonesian and of social-media text. IndoBERT was developed in the context of Indonesian NLP resources, while IndoBERTweet is used in research focused on text from social platforms. These models can help represent word context better than representations that merely count word occurrence, although the actual advantage still depends on the dataset and the evaluation design.

Shaw et al. study emotion classification of Indonesian tweets using large-scale transfer learning via IndoBERT [4]. Aqiilah et al. address sentence-level sentiment classification of Indonesian app reviews with IndoBERTweet [16]. Both studies reflect the expanded use of pretrained models on different tasks and text types.

Transformer models are also used in research oriented to social and policy issues. Ihalauw et al. examine sentiment toward the early performance of a government using IndoBERT and SMOTE on platform X [17]. Muhabbab et al. study sentiment toward the Free Nutritious Meal program on platform X using IndoBERT [18], while Mursyid et al. discuss the use of IndoBERT for the same program topic [19]. Other research addresses political discourse, the "Indonesia Gelap" issue, and the alleged-fake-diploma issue in digital conversation [20], [21], [22].

The presence of these various studies shows that pretrained language models are used in diverse contexts. Even so, the large number of application examples does not, in itself, prove that one model is most dominant or most accurate. Such a conclusion requires a verified corpus count and a performance comparison under appropriate testing conditions.

Diversity of Data Sources and Application Domains

Data sources are another important dimension in the sentiment-analysis literature. The social-media platform X (formerly Twitter) provides relatively spontaneous text related to current issues. Customer reviews and app reviews, on the other hand, more often reflect user experience with a particular product or service. The two sources have different language characteristics, sentiment distributions, and analytical purposes.

In the social and environmental domain, Mubaraq and Maharani use Twitter data to study sentiment on climate change in Indonesia [5]. In the political domain, Putri et al. examine sentiment related to the 2024 presidential election [6], while Aljabar et al. study the presidential election using the DistilBERT-base-uncased transformer approach [23]. Research related to government, public programs, and political discourse is also evident in Ihalauw et al. [17], Muhabbab et al. [18], Mursyid et al. [19], Sidauruk and Herowati [20], and Prakosta et al. [21].

In the e-commerce domain, Sumendap and Mahendra compare sentiment analysis of Shopee and Tokopedia customer reviews using the Google NLP API [7]. Aqiilah et al., meanwhile, focus on app reviews [16]. These studies show the relevance of sentiment analysis in evaluating the user experience and digital services.

Differences in data source have methodological consequences. Social-media text can contain informal language, sarcasm, abbreviations, and strong emotional responses, while product reviews can be more focused on features, quality, price, or service. A model that performs well in one domain does not necessarily yield similar results in another. Cross-domain generalization therefore needs to be tested explicitly.

Expansion toward Emotion Classification and Aspect-Based Sentiment Analysis

The development of sentiment analysis is not limited to the classification of positive, negative, and neutral polarity. A number of studies extend the scope of analysis to the recognition of emotion and the understanding of particular aspects in text.

Ahmadian et al. discuss hybrid models for emotion classification and sentiment analysis in Indonesian [2]. Shaw et al. also study emotion classification of Indonesian tweets with transfer learning [4]. This expansion is important because sentiment polarity and emotion are not fully identical. Negatively toned text, for example, can represent anger, sadness, fear, or disappointment, which require more detailed interpretation.

Aspect-based sentiment analysis (ABSA) seeks to identify sentiment directed at particular aspects. In service reviews, a user may express satisfaction with price but dissatisfaction with delivery. An overall sentiment label risks erasing such distinctions. Anthera Mumtaz and Abdul Majid address the use of web scraping and the fine-tuning of IndoBERT for ABSA on the Kurikulum Merdeka topic [8].

An aspect-based approach has the potential to give more informative results but requires more detailed annotation. Researchers need to establish the definition of aspects, the relationship between aspects and opinions, and rules for handling sentences that discuss several aspects at once. Annotation quality becomes an important factor in determining a model's validity.

Data Quality, Class Imbalance, and Model Evaluation

Data quality is an important foundation for the validity of sentiment analysis. A dataset containing duplication, irrelevant text, inconsistent labels, or an imbalanced class distribution can produce misleading evaluation. High accuracy, for example, does not necessarily indicate good classification ability when one class is far more dominant than the others.

Ihalauw et al. discuss the use of SMOTE in sentiment analysis of a government's early performance [17]. Ansori et al. emphasize a Twitter sentiment-analysis baseline with a classical approach and quality control [14]. Both examples underscore the importance of attending to dataset characteristics in addition to choosing an algorithm.

SMOTE or other class-balancing methods need to be applied carefully. In predictive evaluation, data balancing should be performed only on the training portion, after the data split, to avoid leaking information from the test data. In addition, the method of generating synthetic samples must fit the data representation used.

Model evaluation also needs to report relevant metrics, such as precision, recall, F1-score, and the confusion matrix, in addition to accuracy. For multi-class classification with an imbalanced distribution, macro-F1 can help depict average performance across classes. Comparisons among models also need to consider dataset size, annotation quality, the validation scheme, and the possibility of overlap between training and test data.

Explainable AI and Emerging Research Directions

The development of increasingly complex language models creates a need to understand the basis of the predictions produced. A model with high predictive performance is not always easy to explain, especially when it is used to interpret public opinion on sensitive issues. In this context, explainable artificial intelligence (XAI) becomes a relevant direction of development.

Widiyatmoko et al. combine IndoBERT-LSTM with SHAP for an explainable sentiment-analysis approach [9]. Delvian et al. use LIME to interpret sentiment classification on the IKN, PPN, and MBG issues [10]. Both approaches aim to help users understand the features or parts of a text that contribute to a model's prediction.

Interpretability is important for improving transparency, but the explanations produced by XAI methods are not automatically causal explanations of why a person holds a particular sentiment. Such explanations mainly describe model behavior according to the interpretation method used. The evaluation of XAI therefore needs to consider the stability of explanations, their consistency with model behavior, and their comprehensibility to users.

Beyond XAI, the available references show attention to topic modeling, emotion detection, and response generation. Walikhsani et al. combine NLP and LDA topic modeling to identify the dominant factors related to negative sentiment toward the weakening of the rupiah [24]. Yushendri and Musa discuss emotion detection and the generation of motivational responses in Indonesian text [25]. These approaches show the possibility of developing from sentiment classification toward richer analysis, although each task has different aims and evaluation needs.

Overall, the development directions reflected in the references include improving language representation, expanding analytical tasks, controlling data quality, and model interpretability. The degree of dominance of each direction still needs to be demonstrated through bibliographic coding and measured descriptive analysis.

4. Discussion

From Feature-Based Methods to Contextual Language Models

The shift from feature-based methods to contextual language models is one of the main conceptual developments in Indonesian sentiment analysis. Classical methods offer a baseline that is relatively easy to apply and can help researchers understand the contribution of lexical features to classification. However, representation based on word frequency is limited when meaning depends on sentence context, negation, or relationships among words.

Transformer-based models provide an alternative approach by learning contextual representations. IndoBERT, for example, was built in the context of developing Indonesian NLP models and resources [1]. Applied research using IndoBERT on environmental and political issues shows that pretrained models can be used in various Indonesian-text contexts [5], [6].

Even so, an increase in model complexity does not guarantee improved performance in every condition. Small datasets, noisy labels, domain change, and differing language distributions can limit the benefit of pretrained models. Transformer models can also require larger computational resources and be harder to explain than classical baselines.

Future research should therefore not only report the use of the latest model but also compare it with an appropriate baseline. Comparisons need to be designed with consistent data-split procedures and relevant evaluation metrics. This approach allows researchers to distinguish improvements that genuinely come from the model from improvements that arise because of differences in dataset or testing procedure.

Domain-Specific Sentiment Analysis and Generalization

The diversity of domains in the literature shows that Indonesian sentiment analysis is used to meet different needs. Political research focuses on public opinion and government discourse, customer-review research is oriented to the user experience, and environmental research examines responses to social issues or policy. Each domain has different vocabulary, language style, sentiment distribution, and annotation characteristics.

These differences create a challenge of generalization. A model trained on political tweets may not be directly suitable for app reviews, because the expression of opinion and lexical context in the two domains are not identical. Likewise, sentiment labels used to measure customer satisfaction do not necessarily capture the nuances of opinion in public-policy discussion.

Research therefore needs to distinguish in-domain evaluation from cross-domain evaluation. In-domain evaluation measures a model's ability on data relatively comparable to the training data, while cross-domain evaluation tests whether a model still works in a different context. Reporting the type of evaluation will help readers judge the limits of applying research results.

The Importance of Data and Annotation Quality

Data quality determines how well a model can learn the relationship between text and sentiment labels. Labeling errors, inconsistent category definitions, and a lack of context can limit the validity of results even when the model used has high capacity. This challenge grows with informal language, sarcasm, irony, code-mixing, and local expressions.

The use of class-balancing techniques, such as those discussed in the study by Ihalauw et al. [17], can help address imbalanced label distributions in certain conditions. However, class balancing is not a substitute for annotation quality. If the definition of sentiment is inconsistent, adding samples does not automatically improve label quality.

Future research needs to report annotation guidelines, the number of annotators, the mechanism for resolving differences, the class distribution, and the quality-control strategy. If a dataset is built from social media, the data-collection process also needs to consider representativeness, changes in platform-access policy, and the limits of generalizing from the users whose data could be collected.

Beyond Polarity: Emotion, Aspects, and Explanations

The expansion from polarity classification toward emotion and ABSA shows that the needs of sentiment analysis are increasingly oriented to more specific information. Polarity gives a general picture of the tendency of opinion, while emotion classification can help distinguish types of affective response. ABSA provides information about the aspect targeted by an opinion, so that results are more easily linked to corrective action.

However, this expansion of tasks also increases the complexity of evaluation. Emotion classification requires consistent category definitions, while ABSA requires connected annotation of aspects and sentiment. Researchers need to clarify whether a task is performed at the document, sentence, or aspect level so that results can be interpreted correctly.

XAI adds a dimension of transparency by helping explain the parts of a text that influence a model's prediction [10], [9]. However, model explanations must be interpreted according to the limitations of the technique used. Markers of words that contribute to a prediction do not always reveal the substantive reason behind a user's opinion. An XAI approach will therefore be stronger if complemented by an evaluation of explanation stability and a check of its consistency with human interpretation.

Implications for Research and Practice

For researchers, the diversity of approaches in the literature shows the importance of building comparable experimental designs. The use of open datasets, documented baselines, consistent data splits, and appropriate evaluation metrics can improve the replication and comparability of research. Transparent reporting also helps prevent overstated claims about a model's superiority.

For organizations, sentiment analysis can support opinion monitoring, service evaluation, and the identification of emerging issues. However, the results of analysis should not be treated as a perfect representation of the opinion of the entire population. Data from social media, for example, reflect only the activity of participating users and can be affected by platform mechanisms, account characteristics, and the dynamics of an issue.

Using the results of analysis for decision-making also requires contextual interpretation. A proportion of negative sentiment does not always indicate rejection of every aspect of a policy, just as positive sentiment does not always indicate strong support. Model results need to be combined with aspect analysis, contextual reading, or qualitative examination when the decisions to be made have important consequences.

Future Research Agenda

Based on the conceptual synthesis, several research agendas are relevant. First, research needs to broaden cross-domain and cross-time evaluation. A model trained on particular data needs to be tested in different periods and contexts to determine its resilience to changes in vocabulary, topic, and patterns of expression.

Second, research needs to improve the quality and representativeness of Indonesian datasets, including variations of informal language, code-mixing, regional languages, and user groups that are not yet well represented. The study of Balinese-language text by Wibawa and Pramartha can serve as a methodological context for discussing the development of analysis for languages with local characteristics, although that domain needs to be separated from Indonesian sentiment analysis specifically [15].

Third, comparison between classical and transformer models needs to be carried out through a fair evaluation design. In addition to accuracy, research should report macro-F1, computational needs, inference time, and generalization ability where such metrics are relevant.

Fourth, the development of ABSA, emotion classification, and XAI needs to be accompanied by clear task definitions and evaluation. Research needs to distinguish classification performance from the quality of explanation and its practical benefit.

Fifth, descriptive and bibliometric analyses of the literature need to use a verified corpus. Further research can measure the development of publications per year, the distribution of methods, relationships among topics, and author collaboration, while still distinguishing bibliometric indicators from evidence of scientific quality.

Limitations

This article has several limitations. First, the synthesis is compiled on the basis of the provided reference list and is not yet supported by a corpus from a comprehensive search of scholarly databases. Second, the number of publications per year and the distribution of methods have not been counted through bibliographic coding, so this article does not state percentages or category frequencies as empirical findings.

Third, the available references include studies with diverse aims, data sources, and evaluation methods. The discussion of development directions therefore cannot be used directly to conclude the relative superiority of a model. Fourth, the metadata and completeness of publications need to be verified before the reference list is used in a final manuscript.

To improve the validity of the research, the next stage needs to examine the original publications, establish inclusion and exclusion criteria, code each publication with consistent categories, and compute descriptive statistics from verified data.

5. Conclusion

This study presents a descriptive synthesis of the development of Indonesian sentiment analysis using Natural Language Processing on the basis of the available collection of references. The synthesis shows several directions of conceptual development: the use of feature-based methods and classical machine learning as a basic approach, the use of pretrained language models such as IndoBERT and IndoBERTweet, the expansion of application to various social domains and digital services, and the development of tasks for emotion classification, aspect-based sentiment analysis, and explainable AI.

The literature discussed also shows that the success of sentiment analysis is not determined by the choice of model alone. Data quality, consistency of annotation, class imbalance, language variation, evaluation strategy, and generalization ability all play an important role in determining the validity of results. Research using more complex models therefore still needs to provide baselines, transparent testing procedures, and appropriate evaluation metrics.

The main contribution of this article is a descriptive framework that organizes the development of research by computational approach, language model, data source, application domain, and the expansion of analytical tasks. This framework can be used as a basis for building a more measured mapping of the literature and for formulating an agenda for further research.

Even so, the tendencies discussed remain conceptual and do not yet represent a systematic count of all relevant scholarly publications. Subsequent research needs to complete the database search, metadata verification, publication coding, and computation of descriptive statistics so that conclusions about annual trends and the dominance of methods can be supported by traceable quantitative evidence.

Declarations

Acknowledgements. The authors have no additional individuals or institutions to acknowledge for their contributions to this research.

Author Contributions. Bima Satriatama contributed to the conceptualization, methodology, literature review, descriptive synthesis, analysis of research trends, and preparation of the original manuscript. Marhendra Putra contributed to the literature analysis, interpretation of the reviewed publications, development of the descriptive framework, and review and editing of the manuscript. Both authors contributed to the development of the article and approved the final manuscript.

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest. The authors declare that they have no conflict of interest regarding the publication of this article.

Data Availability. This study is a literature-based descriptive review. The publications discussed in the analysis are listed in the reference section, together with their available digital object identifiers (DOIs) or publication links. No new datasets were generated for this study.

6. References

[1] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP,” in Proc. 28th Int. Conf. Computational Linguistics (COLING), 2020, pp. 757–770. https://doi.org/10.18653/v1/2020.coling-main.66

[2] H. Ahmadian, T. F. Abidin, H. Riza, and K. Muchtar, “Hybrid models for emotion classification and sentiment analysis in Indonesian language,” Applied Computational Intelligence and Soft Computing, vol. 2024, art. 2826773, 2024. https://doi.org/10.1155/2024/2826773

[3] C.-H. Lin and U. Nuha, “Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy,” Journal of Big Data, vol. 10, art. 88, 2023. https://doi.org/10.1186/s40537-023-00782-9

[4] C. Shaw, P. LaCasse, and L. Champagne, “Exploring emotion classification of Indonesian tweets using large scale transfer learning via IndoBERT,” Social Network Analysis and Mining, vol. 15, art. 22, 2025. https://doi.org/10.1007/s13278-025-01439-6

[5] M. F. Mubaraq and W. Maharani, “Sentiment analysis on Twitter social media towards climate change in Indonesia using the IndoBERT model,” Jurnal Media Informatika Budidarma, vol. 6, no. 4, pp. 2426–2431, 2022. https://doi.org/10.30865/mib.v6i4.4570

[6] D. I. Putri, A. N. Alfian, M. Y. Putra, and P. D. Mulyo, “IndoBERT model analysis: Twitter sentiments on Indonesia’s 2024 presidential election,” Journal of Applied Informatics and Computing, vol. 8, no. 1, pp. 7–12, 2024. https://doi.org/10.30871/jaic.v8i1.7440

[7] R. V. Sumendap and I. B. M. Mahendra, “Comparing sentiment analysis of Shopee and Tokopedia customer reviews using Google’s NLP API,” JELIKU (Jurnal Elektronik Ilmu Komputer Udayana), vol. 11, no. 4, pp. 655–662, 2023. https://doi.org/10.24843/JLK.2023.v11.i04.p02

[8] D. S. Anthera Mumtaz and N. W. Abdul Majid, “Web scraping and fine-tuning IndoBERT for aspect-based sentiment analysis (ABSA) on Twitter/X data: A case study of the Kurikulum Merdeka topic,” Rabit: Jurnal Teknologi dan Sistem Informasi Univrab, vol. 11, no. 1, pp. 949–963, 2026. https://doi.org/10.36341/rabit.v11i1.7155

[9] A. T. Widiyatmoko, A. Nugroho, and M. N. Firdaus, "Implementation of the IndoBERT-LSTM model for Indonesian sentiment analysis with an explainable AI approach using SHAP," Journal of Electrical Engineering and Informatics, vol. 4, no. 1, pp. 48–57, 2026. https://doi.org/10.59562/jeeni.v4i1.14030
[10] L. Delvian, E. Firmansyah, and B. Sutara, "Application of Explainable AI LIME to sentiment classification of the IKN, PPN, and MBG issues using IndoBERTweet," Journal Software, Hardware and Information Technology, vol. 6, no. 1, 2026. https://doi.org/10.24252/shift.v6i1.236
[11] B. Setiawan, "A review of sentiment analysis applications in Indonesia between 2023–2024," Journal of Information Engineering and Educational Technology, vol. 8, no. 2, pp. 71–83, 2025. https://doi.org/10.26740/jieet.v8n2.p71-83
[12] R. F. P. Daviansyah, G. B. Hertantyo, and P. Assiroj, "The effectiveness of natural language processing models for Indonesian-text sentiment analysis in the public-service domain: A systematic literature review 2021–2026," Didaktik: Jurnal Ilmiah PGSD STKIP Subang, vol. 12, no. 3, 2026. https://doi.org/10.36989/didaktik.v12i03.19124
[13] D. M. Zara et al., "A meta-analysis of the performance of natural language processing methods in Indonesian-text sentiment analysis," IKRA-ITH Teknologi: Jurnal Sains dan Teknologi, vol. 10, no. 2, 2026. [Online]. Available: https://journals.upi-yai.ac.id/index.php/ikraith-teknologi/article/view/7608
[14] A. Ansori, R. Kania, T. I. Solihati, and F. Damyati, "Quality-controlled Indonesian Twitter sentiment analysis baseline using TF-IDF and classical machine learning approaches," Transactions on Informatics and Data Science, vol. 3, no. 2, pp. 107–122, 2026. https://doi.org/10.24090/tids.v3i2.16900
[15] P. W. A. Wibawa and C. R. A. Pramartha, "Sentiment analysis of Balinese-language text using the Multinomial Naive Bayes method with TF-IDF and BoW," Jurnal Nasional Teknologi Informasi dan Aplikasinya, vol. 2, no. 1, pp. 37–46, 2023. https://doi.org/10.24843/JNATIA.2023.v02.i01.p05
[16] I. N. Aqiilah, R. Saptono, and A. Syaifuddin, "Sentence-level sentiment analysis of Indonesian app reviews using IndoBERTweet," Journal of Computing Theories and Applications, vol. 4, no. 1, pp. 74–91, 2026. https://doi.org/10.62411/jcta.16240
[17] S. A. Ihalauw, N. T. Lapatta, D. W. Nugraha, Wirdayanti, and C. A. Lamasitudju, "Sentiment analysis of a government's early performance using IndoBERT and SMOTE on social media X," Jurnal Algoritma, vol. 22, no. 2, pp. 803–815, 2025. https://doi.org/10.33364/algoritma/v.22-2.2957
[18] A. Z. Muhabbab, Bunyamin, and Hasmawati, "Sentiment analysis of Indonesia's Free Nutritious Meal program on platform X (formerly Twitter) using IndoBERT," ZERO: Jurnal Sains, Matematika dan Terapan, vol. 9, no. 3, 2025. https://doi.org/10.30829/zero.v9i3.27629
[19] M. Mursyid, A. Setiawan, and M. Arifin, "Implementation of IndoBERT in sentiment analysis of Free Nutritious Meal programs on the X social media platform," Jurnal Teknologi Informasi dan Pendidikan, vol. 19, no. 1, 2026. https://doi.org/10.24036/jtip.v19i1.1104
[20] V. E. Sidauruk and W. Herowati, "IndoBERT-based sentiment analysis of political discourse on platform X: The case of the Prabowo-Gibran administration," Journal of Applied Informatics and Computing, vol. 10, no. 1, 2026. https://doi.org/10.30871/jaic.v10i1.11586
[21] R. Prakosta, A. M. Widodo, R. Widayanti, and R. T. Iswahyudi, "Sentiment analysis of the 'Indonesia Gelap' issue using IndoBERT," Emerging Information Science and Technology, vol. 7, no. 1, 2026. https://doi.org/10.18196/eist.v7i1.32902
[22] N. R. Ramadhani and S. F. C. Haviana, "Implementation of the IndoBERT model for public sentiment analysis of the alleged fake-diploma issue of President Joko Widodo in conversations on platform X (Twitter)," Jurnal Ilmiah Sultan Agung, vol. 5, no. 1, 2026. [Online]. Available: https://jurnal.unissula.ac.id/index.php/JIMU/article/view/52032
[23] A. Aljabar, B. M. Karomah, N. Tarisafitri, and Jeffry, "Sentiment analysis in Indonesia's presidential election 2024 using a transformer (DistilBERT-base-uncased)," Journal of System and Computer Engineering, vol. 6, no. 2, 2025. https://doi.org/10.61628/jsce.v6i2.1867
[24] Z. Walikhsani, W. Anggaraeni, and R. D. Handayani, "Identifying the dominant factors in negative public sentiment toward the weakening of the rupiah using NLP and LDA topic modeling," Jurnal SINTA: Sistem Informasi dan Teknologi Komputasi, vol. 3, no. 4, pp. 239–247, 2026. https://doi.org/10.61124/sinta.v3i3.303
[25] J. Yushendri and P. Musa, "Application of natural language processing for emotion detection and motivational response generation in Indonesian text using the CRISP-DM method," SISTEMASI: Jurnal Sistem Informasi, vol. 15, no. 5, 2026. [Online]. Available: https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/6334

© Journal of Sinar Data and Informatics (JSDI) — Sinar Asia Institute. This article is distributed under the Creative Commons Attribution (CC BY 4.0) license.