Indonesian Journal of Data and Science

Volume 6 Issue 3
ISSN 2715-9930
https://doi.org/10.56705/ijodas.v6i3.303

Zero-Shot Sentiment Analysis Of DeepSeek AI App Reviews Using DeepSeek-R1

Restu Sri Pamungkas [1*]; Adhitia Erfina [2]; Cecep Warman [3]

[1]Universitas Nusa Putra, Sukabumi, 43152, Indonesia, restu.sri_si21@nusaputra.ac.id
[2]Universitas Nusa Putra, Sukabumi, 43152, Indonesia, adhitia.erfina@nusaputra.ac.id
[3]Universitas Nusa Putra, Sukabumi, 43152, Indonesia, cecep.warman@nusaputra.ac.id

Correspondence should be addressed to Restu Sri Pamungkas; restu.sri_si21@nusaputra.ac.id

Received 17 October 2025; Accepted 22 December 2025; Published 31 December 2025

© Authors 2025. CC BY-NC 4.0 (non-commercial use with attribution, indicate changes).
License: https://creativecommons.org/licenses/by-nc/4.0/ — Published by Indonesian Journal of Data and Science.

Abstract:
This study aims to evaluate the effectiveness of the Zero-Shot Learning (ZSL) approach using the DeepSeek-R1-Distill-Qwen-1.5B model in performing sentiment classification on Indonesian-language reviews of the DeepSeek AI application from the Google Play Store. A total of 2,000 unlabeled user reviews were collected and processed through instructional prompts to guide the model in classifying sentiments into three categories: positive, negative, and neutral. The model operates without fine-tuning and relies entirely on Zero-Shot Learning using Indonesian-language prompts. Out of 2,000 reviews, 1,348 were successfully classified with valid sentiment labels. Of these, 1,131 reviews (83.9%) were labeled as positive, 211 reviews (15.7%) as negative, and only 6 reviews (0.4%) as neutral. Evaluation results indicated an overall accuracy of 77.67%. The F1-Score for the positive class reached 86.66%, while the negative and neutral classes scored 33.56% and 16.66%, respectively, highlighting the performance disparity between dominant and underrepresented sentiment categories. These findings demonstrate that the DeepSeek-R1 model has strong potential in detecting positive sentiment in Indonesian without requiring additional training. However, its performance on negative and neutral sentiments remains limited, revealing the challenge of handling low-resource and imbalanced data in Zero-Shot settings. Future research should explore improved prompt engineering or multilingual adaptation to address the current limitations and enhance classification consistency across all sentiment categories.

Keywords: Zero-Shot Learning, DeepSeek-R1, Sentiment Analysis, Indonesian Language, Google Play Store.

1. Introduction

Background and Context of the Research

The advancement of Artificial Intelligence (AI) technologies in recent years has led to significant developments, particularly in the field of Natural Language Processing (NLP). NLP has become a central component in many real-world applications, including chatbots, recommendation systems, and public opinion analysis. One of the core tasks in NLP is sentiment analysis, which refers to identifying and classifying emotions or opinions expressed in text data [1]. In the era of digital communication dominated by online reviews and social media, sentiment analysis offers a powerful tool to understand public perception in an automatic and scalable way.

Despite the growing need for sentiment analysis, most available approaches still rely heavily on labeled data, which is often unavailable or limited, especially for non-English languages such as Indonesian. The lack of annotated datasets in Indonesian presents a considerable challenge, particularly in extracting sentiment from user-generated content like app reviews. To address this, Zero-Shot Learning (ZSL) emerges as a promising method, as it enables classification into categories not explicitly seen during training by leveraging only task instructions or label descriptions [3].

In mid-2023, a new open-source AI project from China called DeepSeek AI gained traction, offering competitive performance among large language models (LLMs). Its flagship model, DeepSeek-R1, demonstrated high reasoning capabilities and instruction-following performance comparable to GPT-4 in complex tasks without the need for fine-tuning [2]. Unlike commercial models such as ChatGPT or Claude AI, DeepSeek-R1 is open-source and has shown promising results in multilingual tasks, including low-resource languages.

Although DeepSeek AI is relatively new, its Android app has garnered a substantial number of reviews on the Google Play Store. As of the data collected for this study, it had received approximately 185,000 user reviews, with a significant portion written in Indonesian. In contrast, more established models such as ChatGPT and Gemini AI had larger review counts (22.5 million and 9.05 million, respectively), while Claude AI had about 96,300 reviews. Despite the smaller volume, DeepSeek AI's fast-growing user base and open accessibility make it a compelling case for sentiment analysis in the Indonesian context.

However, most of the reviews lack explicit sentiment labels, making them unsuitable for conventional supervised classification. Therefore, this study explores the use of DeepSeek-R1 in a Zero-Shot setup to classify sentiment in Indonesian-language reviews of the DeepSeek AI app, using only prompts and label descriptions in natural language.

This research aims to evaluate the effectiveness of the DeepSeek-R1 model in performing Zero-Shot sentiment classification without additional fine-tuning. It contributes to the growing body of work in multilingual NLP, particularly in Indonesian, by demonstrating the capabilities of instruction-following LLMs to generalize across tasks and languages. Moreover, it provides insight into public perception of DeepSeek AI, which can benefit developers and researchers exploring low-resource applications.

Problem to be Solved

Despite its capabilities, DeepSeek-R1's performance on sentiment classification in non-English, low-resource languages like Indonesian has not been thoroughly evaluated. Most sentiment analysis models rely on labeled datasets and fine-tuning, which are resource-intensive. Therefore, this study addresses the problem of evaluating whether DeepSeek-R1 can effectively classify sentiment in Indonesian user reviews using Zero-Shot Learning without any additional training.

Research Objectives

- Evaluate the effectiveness of Zero-Shot Learning with DeepSeek-R1 model in sentiment classification of Indonesian language data.
- Analyze the sentiment classification accuracy of user reviews against the DeepSeek AI application.
- Identify the challenges of applying Zero-Shot Learning to unlabeled Indonesian data.

Research Questions or Hypotheses

- Is the Zero-Shot Learning approach using the DeepSeek-R1 model able to classify the sentiment of Indonesian reviews without retraining?
- What is the accuracy of sentiment classification (positive, negative, neutral) performed by the DeepSeek-R1 model?
- What are the obstacles in applying the Zero-Shot Learning model to public review data on the Google Play Store in Indonesian?

Scope and Limitations of the Research

- The data used was 2,000 DeepSeek AI app reviews from the Google Play Store, in Indonesian.
- Sentiment is divided into three categories: positive, neutral, and negative.
- The model used is the open-source DeepSeek-R1 version of the DeepSeek AI ecosystem.
- The approach used is prompt-based Zero-Shot Learning and no retraining (Fine-Tuning).
- Evaluation of model performance is done using classification metrics such as accuracy, precision, recall, and F1-score.

Research Contributions

This research contributes to the growing field of multilingual NLP by demonstrating the application of Zero-Shot Learning for sentiment classification in the Indonesian language using DeepSeek-R1. It offers empirical insights into the model's performance without additional training, contributing to both academic literature and practical application. It also provides actionable findings for developers of DeepSeek AI to understand public perception of their application [4], [5].

Article Structure

This article is organized into five main sections: Section 1 provides an introduction to the research background, objectives, and scope. Section 2 presents related studies and the theoretical foundation, including discussions on NLP, LLMs, ZSL, and DeepSeek-R1. Section 3 describes the research methodology, including data collection, preprocessing, and model evaluation. Section 4 contains the results and discussion. Finally, Section 5 presents the conclusions and suggestions for future research.

2. Method

This study is an applied research in the field of Natural Language Processing (NLP) using a quantitative experimental approach. The research aims to evaluate the performance of the DeepSeek-R1-Distill-Qwen-1.5B model for Zero-Shot sentiment classification on 2,000 Indonesian-language user reviews of the DeepSeek AI application from the Google Play Store.

The research was conducted using Google Colab, a cloud-based environment that supports machine learning libraries such as transformers, pandas, scikit-learn, nltk, and google-play-scraper. Data scraping was done using google-play-scraper, with reviews collected using parameters for Indonesian language (lang='id') and region (country='id'). The collected reviews were stored in CSV format.

Preprocessing involved cleaning (removing emojis, URLs, mentions, hashtags, numbers, symbols, and whitespace) and case folding (converting text to lowercase), which are sufficient for prompt-based classification. Sentiment classification was carried out using the instruction-based Zero-Shot Learning approach, where prompts were crafted to assign each review into one of three sentiment classes: positive, negative, or neutral—without any fine-tuning.

The DeepSeek-R1 model was accessed from Hugging Face and applied directly. Python libraries such as re and nltk were used for preprocessing, while scikit-learn was used to evaluate performance through metrics such as accuracy, precision, recall, and F1-score.

This method was selected for its efficiency, relevance to unlabeled local data, and support in prior literature. Studies by Goncalves et al. (2022) and Munir et al. (2023) showed that instruction-based Zero-Shot Learning is viable for multilingual sentiment analysis, including in low-resource languages like Indonesian [14], [15].



```
[ ]  def zero_shot_label(text, label_set=["positif", "negatif", "netral"]):
        prompt = f"""
    Berikut ini adalah sebuah ulasan aplikasi:

    "{text}"

    Pertanyaannya: Termasuk kategori apa ulasan di atas?
    Pilih salah satu dari: {', '.join(label_set)}.
    Jawaban:
    """.strip()

        inputs = tokenizer(prompt, return_tensors="pt").to(model.device)
        output = model.generate(**inputs, max_new_tokens=10)
        decoded = tokenizer.decode(output[0], skip_special_tokens=True)

        if "Jawaban:" in decoded:
            label = decoded.split("Jawaban:")[-1].strip().split("\n")[0]
        else:
            label = decoded.strip().split("\n")[0]

        return label.lower().strip()
```

Figure 2. Classification Prompt Example

Data Collection (Scrapping)

Data collection was done by scrapping using the Python google-play-scrapper library, which allows automatic extraction of reviews from Google Play Store app pages. The target is the DeepSeek AI application, with the amount of data taken as many as 2,000 Indonesian-language reviews. The data retrieved is only in the form of username and comment content, then stored in CSV format.

Scrapping process using parameters:

- Lang='id', the reviews used are in Indonesian.
- Country='id', the review comes from the Indonesian region.
- sort=Sort.MOST_RELEVANT, displays the most relevant reviews.
- Count=2000, took 2,000 comments.

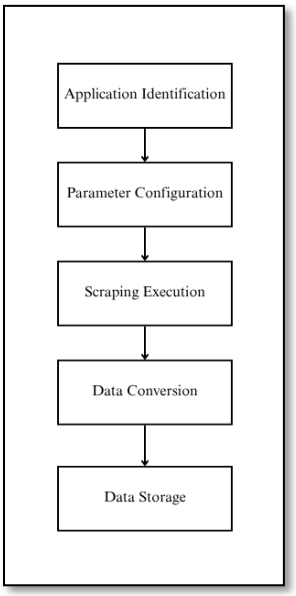


Figure 3. Data Scrapping Flow

The flow in scrapping data starts by determining the target application, namely DeepSeek AI which is available on the Google Play Store with the application ID com.deepseek.chat. Next, parameter configurations such as language (Lang='id'), region (country='id'), number of reviews (count='2000'), and sorting (Sort.MOST_RELEVANT) are adjusted to ensure the data matches the research needs. After that, scrapping was executed using the Python google-play-scrapper library, which automatically retrieves the contents of reviews and usernames. Next, data conversion is carried out, which is initially raw data and then converted into DataFrame format using the pandas library, so that it is easy to analyze further. And finally, the data is stored in a deepseek.csv file with UTF-8 format to be compatible with the next preprocessing stage.

This scrapping technique was chosen because of several advantages:

- Big data efficiency, data capable of retrieving thousands of comments in a short period of time.
- Real-time relevance, the data stored is the latest public opinion.
- Legality and ethics, only accessing public data without login/credentials.

The use of Google Play Store scrapping has also been proven in previous research conducted by Zhang et al (2021) who successfully extracted and analyzed thousands of health application reviews using google-play-scrapper for user

satisfaction research [16]. Arifin and Rimawan (2022) scrapped e-commerce reviews to evaluate user feedback using a similar method [17].

Preprocessing Data

This study applied two main preprocessing steps: cleaning and case folding, to prepare clean and consistent input for the DeepSeek-R1 Zero-Shot Learning model. The cleaning process removed non-ASCII characters, URLs, mentions, hashtags, numbers, symbols, and excessive whitespace—reducing noise and enhancing data quality for sentiment analysis, particularly in social media and app reviews [18]. Following that, case folding converted all text to lowercase to maintain token consistency. Although tokenization, stemming, or stopword removal were not used, this minimal preprocessing was sufficient, as the Zero-Shot model requires only clean text inputs for effective prompt-based inference.

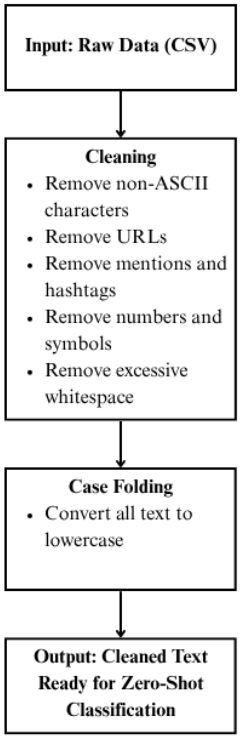


Figure 4. Preprocessing Data Flow

Sentiment Classification Using Zero-Shot Learning

After preprocessing, sentiment classification was conducted using the DeepSeek-R1-Distill-Qwen-1.5B model from Hugging Face, applying a Zero-Shot Learning approach without fine-tuning. The model received prompts in the form of label descriptions (positive, negative, neutral) and was able to classify Indonesian-language reviews directly. According to Fajri et al. (2024), a multilingual lexicon-based Zero-Shot method performs well across 34 languages, including Indonesian [5]. Mu et al. (2023) noted that prompt complexity can affect performance by ±10%, though simpler formats yield more stable results for tri-label tasks [19]. This supports the use of concise prompts, enabling the model to understand social and linguistic context effectively. Overall, this approach is both efficient and adaptable for under-resourced languages.

Model Performance Evaluation

Performance evaluation is an important stage in measuring the effectiveness of the DeepSeek-R1 model in sentiment classification using the Zero-Shot Learning approach. Since the model is not re-trained (Fine-Tuning),

testing is done directly based on the model's response to classification instructions or prompts. This process is done to determine how well the model understands Indonesian text and correctly categorizes it into three sentiment classes: positive, negative and neutral.

The evaluation was conducted using standard classification evaluation metrics, namely:

Accuracy

Accuracy measures the proportion of total correct predictions compared to the entire data.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

which means:

- TP (True Positive) : positive correct prediction
- TN (True Negative) : correct negative prediction
- FP (False Positive) : positive correct prediction
- FN (False Negative) : correct negative prediction

Precision

Precision measures the accuracy of positive predictions, i.e. how many of the outcomes classified as positive are actually positive.

$$\text{Precision} = \frac{TP}{(TP + FP} \quad (2)$$

Recall

Recall measures how much positive data the model has found out of all the data that should be positive.

$$\text{Recall} = \frac{TP}{(TP + FN} \quad (3)$$

F1-Score

F1-Score is the harmonic mean between precision and recall, used when the balance between the two is important.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

This metric is used because it can provide a comprehensive overview of model performance not only on accuracy, but also on sensitivity and balance between classes [19].

The performance evaluation will also be visualized in the form of:

- Confusion Matrix table, to see the number of errors between categories.
- Bar chart of accuracy, precision, recall, and F1-score for each sentiment class.

The evaluation approach using such metrics is supported by Borst et al (2023) who evaluated Zero-Shot Sentiment Classification on German stock market reports with a prompt-based approach. Their research shows that even without retraining, the Zero-Shot Learning model can provide reasonably accurate results when given proper instructions and systematic evaluation based on classification metrics. They also emphasized the importance of visualizing the metrics to clarify the distribution of model errors on each sentiment label, especially on domain-specific data [20].

3. Result and Discussion

Data Collection

A total of 2,000 public reviews in Indonesian were collected from the Google Play Store using the google-play-scraper library. These reviews, which spanned across all star ratings, were stored in CSV format for further analysis.

	userName	content
0	Abdul Aziz	aplikasi nya bagus hanya dari segi mengetik ba...
1	Muhammad Affan	Hasilnya bagus, seringku pakai untuk membandin...
2	Rochyiat Ramdani	Saya suka deepseek Tampilan UI gampang dimenge...
3	Shira	Its a good app but minus terbesarnya itu serin...
4	Balyan Ali	Halo Deepseek, Tolong diperbaiki dan tingkatka...

Figure 6. Data Scrapping Results

Preprocessing Data

Preprocessing was conducted in two main steps:

(1) cleaning — removing emojis, URLs, hashtags, symbols, and excessive whitespace, and
(2) case folding — converting text to lowercase. This preprocessing approach aligns with the requirements of Zero-Shot Learning (ZSL), where clean and simple text improves model performance without additional linguistic features.

```
[ ] import re
    import pandas as pd

    def bersihkan_teks(text):
        text = str(text)
        text = re.sub(r'[^\x00-\x7F]+', ' ', text)  # non-ASCII
        text = re.sub(r'http\S+|www\S+|https\S+', '', text)  # URL
        text = re.sub(r'\@\w+|\#','', text)  # mention/hashtag
        text = re.sub(r"[^a-zA-Z\s]", ' ', text)  # angka & simbol
        text = re.sub(r'\s+', ' ', text).strip()  # spasi berlebih
        return text.lower()
```

Figure 7. Preprocessing Data Prompt

```
# Hapus data kosong setelah preprocessing
df = df[df['content'].str.strip() != '']
df = df.dropna(subset=['content'])

print(f"Jumlah data setelah hapus konten kosong: {len(df)}")
```

```
Jumlah data setelah hapus konten kosong: 2000
```

Figure 8. Total Data After Preprocessing

Based on the preprocessing results, the amount of data before and after cleaning remains the same, which is 2,000 review data. This shows that all scraped data has met the completeness requirement, there are no empty entries, and all review content is suitable enough to be processed further in sentiment classification.

Sentiment Classification Using Zero-Shot Learning

The cleaned reviews were passed to the DeepSeek-R1-Distill-Qwen-1.5B model through an instruction-based prompt. The model, without any fine-tuning, returned one of three sentiment labels: positive, negative, or neutral. Out of 2,000 reviews, only 1,348 (67.4%) were returned with recognizable labels: **1,131 positive (83.9%)**, **211 negative (15.7%)**, and **6 neutral (0.4%)**.

This distribution is visualized in the form of bar charts and pie charts as follows:

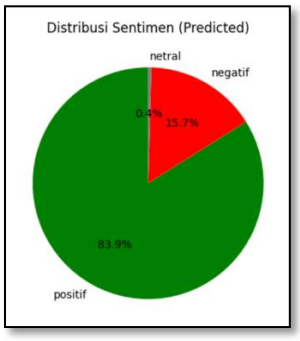


Figure 9. Pie Chart Result Distribution

This result indicates a heavy skew toward positive sentiment, consistent with the word cloud dominated by "*bagus*", "*fitur*", and "*cepat*".



Figure 10. Positive Text WordCloud

Performance Evaluation

The model's classification performance showed:

- **Accuracy**: 77.67%
- **F1-Score (positive)**: 86.66%
- **F1-Score (negative)**: 33.56%
- **F1-Score (neutral)**: 16.66%

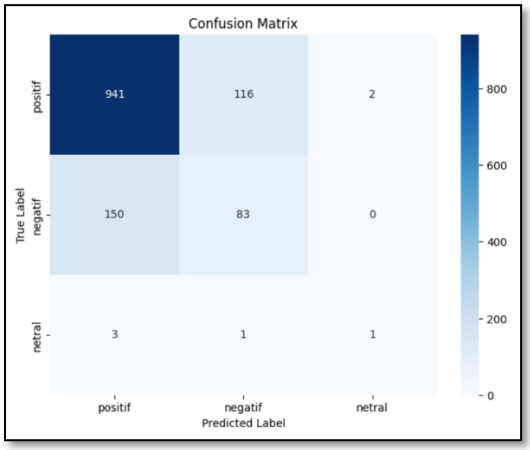


Figure 11. Confusion Matrix of Classification Results

The confusion matrix reveals that 150 negative samples were wrongly labeled as positive, and only 1 of 5 neutral reviews was correctly identified. This confirms the model's bias toward positive sentiment, which can be attributed to several factors:

- **Prompt generality**: The instruction may lack nuanced cues to detect criticism.
- **Vocabulary mismatch**: Indonesian negative expressions are underrepresented due to training on predominantly English data.
- **Data imbalance**: The dataset naturally contains more positive than neutral or negative reviews.

These issues align with findings from Borst et al. [20], who observed similar performance drops in aspect-based sentiment classification using ZSL on domain-specific corpora. Compared to the study by Goncalves et al. [14], which applied GPT-3 on Portuguese reviews, the DeepSeek-R1 model performed reasonably well in detecting positive sentiment but **struggled more with ambiguous categories**, even in a less resource-constrained context.

Overall Analysis

Based on the classification results and performance evaluation, it can be concluded that the DeepSeek-R1 model has the initial ability to understand and classify Indonesian text by Zero-Shot Learning, even though the model was initially trained using English data.

This can be seen from the high accuracy of the "positive" label (F1-Score 86.88%) and the overall accuracy of 77.67%, which shows that the model can capture common sentiment patterns in Indonesian without fine-tuning.

However, the model still has difficulty in distinguishing between negative and neutral sentiments, which may be due to limited understanding of the typical linguistic context of Indonesian or due to the more complex sentence structure compared to English. This is reasonable considering the model was not fine-tuned on Indonesian data and only relied on instructional understanding through prompts.

This research thus provides preliminary evidence that Zero-Shot Learning models based on Large Language Models (LLMs) such as DeepSeek-R1 can be used for Natural Language Processing (NLP) tasks in Indonesian, but needs further development to improve accuracy on more ambiguous labels such as negative and neutral.

This finding is in line with previous literature by Huang & Wang (2025) and Fajri et al (2024) who mentioned that Zero-Shot Learning in local languages remains effective but sensitive to prompt design and label distribution.

4. Conclusion

This study demonstrates the effectiveness of Zero-Shot Learning (ZSL) using the DeepSeek-R1-Distill-Qwen-1.5B model in classifying sentiment from 2,000 Indonesian-language reviews of the DeepSeek AI app. Without any fine-tuning, the model classified 1,348 reviews with valid sentiment labels, dominated by positive sentiments (83.9%), while negative and neutral made up 15.7% and 0.4% respectively. The model achieved 77.67% overall accuracy, with F1-Scores of 86.66% (positive), 33.56% (negative), and 16.66% (neutral), indicating strong performance for positive sentiment but notable challenges in detecting more subtle or negative expressions.

Given these findings, DeepSeek-R1 holds promise for low-resource sentiment analysis using ZSL. However, to enhance performance across all sentiment classes, future research could investigate optimized prompt engineering, integration of few-shot learning strategies, and evaluation with other LLMs such as IndoBERT-ZSL or LLaMA-3. Expanding the study to multilingual or code-switched datasets may also offer insights for broader application. These efforts can support real-world use cases in public opinion monitoring and user feedback analysis in local languages.

Acknowledgments:

The author would like to express sincere gratitude to all individuals and institutions who contributed to the completion of this research. Special thanks are extended to Mr. Adhitia Erfina, S.T., M.Kom and Mr. Cecep Warman, M.Kom as thesis supervisors at Universitas Nusa Putra for their invaluable guidance, feedback, and motivation throughout the writing process.

Appreciation is also extended to the academic staff and lecturers of the Information Systems Department who provided the academic environmen...

The author would also like to thank the Rector of Universitas Nusa Putra Sukabumi, Dr. Kurniawan, S.T., M.Si., MM., the Dean of the Faculty of Computer Engineering and Design, Ir. Paikun, S.T., M.T., IPM., and the Head of the Information Systems Study Program, Mr. Falentino Sembiring, M.Kom., for their support and academic leadership.

Deepest gratitude is extended to the author's beloved parents, Enung Siti Nurhayati and Asep Sudrajat, for their endless prayers, encouragement, and moral support. Special thanks to the author's husband, Yogi Iskandar, S.Kom., for his motivation and unwavering support during the research process.

To fellow students of the Information Systems Program, especially the BION group—Mutia, Nabila, Kaysha, and Sindy—thank you for the companionship and collaboration throughout the academic journey. Appreciation is also given to Halana Hijab and SBCSkin for the support given during this final stage of study.`

Lastly, the author would like to thank all parties who contributed directly or indirectly to this research. May all their kindness and support be rewarded with blessings and success.

References:

[1] N. S. Suryawanshi, “Sentiment Analysis with Machine Learning and Deep Learning: A Survey of Techniques and Applications,” Int. J. Sci. Res. Arch., vol. 12, no. 2, pp. 5–15, 2024.

[2] J. Chen, X. Xie, J. Liu, et al., “Knowledge-aware Zero-Shot Learning: Survey and Perspective,” in Proc. IJCAI, 2021.

[3] D. Guo et al., “DeepSeek R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning,” arXiv preprint arXiv:2501.12948, 2025.

[4] D. Huang and Z. Wang, “Explainable Sentiment Analysis with DeepSeek-R1: Performance, Efficiency, and Few-Shot Learning,” arXiv preprint arXiv:2503.11655, 2025.

[5] F. Fajri, P. Eiswirth, and S. Ramon, “Zero-Shot Sentiment Analysis in Low-Resource Languages Using a Multilingual Lexicon,” in Proc. EACL, Dubrovnik, 2024.

[6] F. Koto, T. Beck, Z. Talat, I. Gurevych, dan T. Baldwin, “Zero-shot Sentiment Analysis in Low-Resource Languages Using a Multilingual Sentiment Lexicon,” Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Long Papers), hal. 298–320, Mar. 2024. DOI: 10.18653/v1/2024.

[7] DeepSeek-AI, “DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning,” Technical Report, DeepSeek, Jan. 2025.

[8] C. Shaw, P. LaCasse, and L. Champagne, “Exploring Emotion Classification of Indonesian Tweets Using Large Scale Transfer Learning via IndoBERT,” Soc. Netw. Anal. Min., vol. 15, Art. no. 22, 2025.

[9] J. Eisenstein et al., “Natural Language Processing and Its Applications,” in Proc. EMNLP Tutorial, 2023.

[10] M. A. Khan et al., “A Comprehensive Review on Sentiment Analysis and Opinion Mining: Approaches and Challenges,” Appl. Sci., vol. 12, no. 5, p. 2345, 2022.

[11] Y. Yin et al., “Zero-Shot Learning with Knowledge Graphs for Text Classification,” in Proc. ACL, 2021.

[12] K. L. Tan, C. P. Lee, and K. M. Lim, “A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research,” Appl. Sci., vol. 13, no. 7, p. 4550, 2023.

[13] A. Gupta and D. Kamthania, “Study of Sentiment on Google Play Store Applications,” in Proc. ICICC, Apr. 2021.

[14] R. Goncalves, L. Silva, and M. Ferreira, “Zero-Shot Sentiment Analysis in Portuguese Using Prompted LLMs,” arXiv preprint arXiv:2203.12345, 2022.

[15] S. Munir, A. Rizvi, and K. Raj, “Evaluating Instruction-Based Zero-Shot Learning on Low Resource Languages,” arXiv preprint arXiv:2306.09876, 2023.

[16] M. Zhang, J. Li, and S. Wang, “Large-Scale Analysis of Health App Reviews from Google Play Store Using Scraping Techniques,” in Proc. 2021 IEEE Int. Conf. Data Mining Workshops, pp. 18–25, 2021.

[17] R. A. Putri, “Adapting Language Models to Indonesian Local Languages: An Empirical Study of Language Transferability on Zero-Shot Settings,” arXiv preprint, Jul. 2025.

[18] K. S. Nugroho, A. Y. Sukmadewa, H. Wuswilahaken, F. A. Bachtiar, dan N. Yudistira, “BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews,” arXiv preprint, Jul. 2021.

[19] Y. Mu et al., “Navigating Prompt Complexity for Zero Shot Classification,” in Proc. LREC-COLING, 2024.

[20] J. Borst, L. Wehrheim, A. Niekler, and M. Burghardt, “An Evaluation of a Zero-Shot Approach to Aspect-Based Sentiment Classification in Historic German Stock Market Reports,” in Proc. 18th Conf. Comput. Sci. Intell. Syst. – AI in Digital Humanities Workshop, ACSIS, vol. 37, pp. 51–60, 2023.